



SAS[®] Visual Analytics: 機械学習オブジェクトの操作

2026.04

このドキュメントは、ソフトウェアの追加バージョンに適用される場合があります。このドキュメントを [SAS Help Center](#) で開き、バナーのバージョンをクリックすると、使用できるすべてのバージョンが表示されます。

機械学習オブジェクトについて	3
機械学習オブジェクトとは?	3
機械学習オブジェクトを使用する利点	3
モデリング情報	3
使用可能なモデル	3
共通情報の概要	4
Model Studio との統合	4
非決定論的行動	5
自動調整	6
一般的なモデリングタスク	10
モデルのエクスポートとスコアリング	10
機械学習オブジェクトの入門ガイド	12
概要	12
サンプルデータをダウンロードする	12
レポートの作成	12
フォレストの作成	13
勾配ブースティングモデルの作成	14
ニューラルネットワークの作成	16
モデルの比較の実行	17
ベイジアンネットワークの操作	20
ベイジアンネットワークの概要	20
ベイジアンネットワークの作成	20

ベイジアンネットワークオプション	21
ベイジアンネットワークモデル表示オプション	23
ベイジアンネットワークの結果	25
詳細テーブル	26
ファクタライゼーションマシンについて	27
ファクタライゼーションマシンの概要	27
ファクタライゼーションマシンの作成	27
ファクタライゼーションマシンのオプション	28
ファクタライゼーションマシンモデルの表示オプション	28
ファクタライゼーションマシンの結果	30
詳細テーブル	30
フォレストについて	31
フォレストの概要	31
フォレストの作成	32
フォレストのオプション	32
フォレストモデル表示オプション	34
フォレストの結果	36
詳細テーブル	37
勾配ブースティングモデルについて	38
勾配ブースティングモデルの概要	38
勾配ブースティングモデルの作成	39
勾配ブースティングオプション	39
勾配ブースティングモデル表示オプション	41
勾配ブースティング結果	43
詳細テーブル	45
ニューラルネットワークについて	46
ニューラルネットワークの概要	46
ニューラルネットワークの作成	46
ニューラルネットワークオプション	46
ニューラルネットワークモデルの表示プロパティ	49
ニューラルネットワークの結果	51
詳細テーブル	53
サポートベクターマシンについて	54
サポートベクターマシンの概要	54
サポートベクターマシンの作成	54
サポートベクターマシンのオプション	55
サポートベクターマシンモデル表示オプション	57
サポートベクターマシンの結果	59
詳細テーブル	60

機械学習オブジェクトについて

機械学習オブジェクトとは?

SAS Viya は機械学習オブジェクトを提供し、SAS Visual Analytics および SAS Visual Statistics を使用すると、SAS サーバーのインメモリ機能を使用してモデルを開発し、テストできます。SAS Visual Analytics は、関連するパターンを明らかにするためにデータソースを探索、調査、ビジュアル化できます。SAS Viya は、SAS Visual Analytics で発見されたパターンに基づいてモデルを作成、テスト比較することにより、これらの機能を拡張します。SAS Viya は、作成したモデルを他の SAS 製品で使用したり、本番環境に移行したりするために、モデルの比較を実行する前または後にそのスコアコードまたは分析ストアをエクスポートできます。

機械学習オブジェクトを使用する利点

SAS Viya は、使いやすい Web ベースのインターフェイスで強力なデータマイニングおよび機械学習モデルを素早く作成できます。SAS Visual Statistics は、SAS Viya 機械学習オブジェクトで動作するモデル比較ツールを提供します。モデル比較ツールを使用すると、2 つ以上のモデルの相対的なパフォーマンスを相互に評価し、チャンピオンモデルを選択できます。さまざまなモデル選択基準が使用できます。モデルを比較するかどうかにかかわらず、モデルをエクスポートできます。エクスポートされたモデルを使用すると、モデルを新規データに簡単に適用できます。

モデリング情報

使用可能なモデル

次の機械学習モデルは SAS Viya で使用できます。

- [ベイジアンネットワーク \(20 ページ\)](#)は、ノードがランダム変数を表し、ノード間のリンクが 2 つのランダム変数間の条件付き依存関係を表す有向非巡回プロットモデルです。
- [ファクタライゼーションマシン \(27 ページ\)](#)は、大規模で疎なデータセットを効果的に処理するための因子分解モデルを実装しています。
- [フォレスト \(31 ページ\)](#)は、指定された数のディシジョンツリーを含むアンサンブルモデルです。各ツリーは、利用可能なデータの一部のみで学習されています。

- [勾配ブースティング \(38 ページ\)](#)は、複数のツリーを作成する反復的なアプローチです。各ツリーは、データの独立したサンプルに基づいています。
- [ニューラルネットワーク \(46 ページ\)](#)は、人間の脳の生物学的構造を模倣するように設計された一連の接続を作成します。
- [サポートベクターマシン \(SVM\) \(54 ページ\)](#)は、2 つのクラス間の距離を最大化する 1 組の超平面を作成します。

共通情報の概要

SAS Viya 機械学習モデルは、SAS Visual Statistics と同じ基本構造を使用します。[変数](#)、[検証データ](#)、[欠損値](#)、[フィルター変数](#)、[当てはめの統計量](#)、[スコアコード](#)、[モデルの登録](#)、[Model Studio との統合](#)、および[予測値](#)については、SAS Visual Statistics ドキュメントを参照してください。

Model Studio との統合

SAS Visual Analytics を使用すると、特定の分析モデルを SAS Visual Analytics から Model Studio に転送できます。

モデルを SAS Visual Analytics から Model Studio へ移動するには、右上隅の[パイプラインの作成](#)をクリックします。このアクションは、次の要素を含む Model Studio に新しいプロジェクトを作成します。

- アクティブなデータセット
- すべてのデータ処理、フィルタリング、変換を適用するスコアコード
- エクスポートされたモデルを実行するためのスコアコード

モデリングノードのプロパティを編集し、その後モデルを Model Studio で再学習させることができます。モデリングノードを右クリックし、[プロパティを有効にする](#)を選択してノードを編集します。モデルの解釈可能性プロパティは常に自動的に有効になります。このパイプラインでは、他の Model Studio パイプラインと同様にノードを追加および削除できます。Model Studio のモデルの比較およびパイプラインの比較ツールを使用して、転送されたモデルを新しいモデルに対して評価できます。

現時点でサポートされているモデルは、ベイジアンネットワーク、ファクタライゼーションマシン、フォレスト、勾配ブースティング、ニューラルネットワーク、サポートベクターマシンです。これらのモデルには例外もあります。

- 勾配ブースティングまたはフォレストモデルを転送する場合、Model Studio にマッピングされる **最大レベル数**プロパティの値は、SAS Visual Analytics で指定した値よりも 1 つ少なくなります。

モデルを SAS Visual Analytics から Model Studio に転送する際には、いくつかの注意点があります。

- Model Studio 既存のプロジェクトに SAS Visual Analytics モデルを追加するには、ターゲット変数名と種類が一致し、Model Studio のターゲット変数水準が SAS Visual Analytics の応答変数分類に適合している必要があります。また、既存の Model Studio プロジェクトにあるデータノードも事前に実行されている必要があります。
- 元のデータセットに存在する変数名を使用するかわりに、SAS Visual Analytics は変数ラベルを使用します。しかし、Model Studio は元のデータセットに存在する変数名を使用します。したがっ

て、入力データの変数名と変数ラベルが異なる場合、モデルの転送時に予期しない名前付けの問題が発生することがあります。Model Studio は**データペインの変数テーブル**レイアウトに変数名と変数ラベルの両方を表示します。

- SAS Visual Analytics は特定の状況下では、新しいターゲット変数が作成されることがあります (たとえば、サポートベクターマシンを転送する場合など)。
- 名義尺度のターゲット変数の場合、書式設定されたターゲット値には、空白、セミコロン、カンマ、ピリオド、アスタリスクのうち少なくとも 1 つの特殊文字が含まれていてはなりません。ターゲット値に 5 つの特殊文字がすべて含まれる場合、エラーが発生します。
- モデルの比較オブジェクトからエクスポートしている場合、チャンピオンモデルのみがエクスポートされます。
- SAS Visual Analytics からのパーティション情報が Model Studio と異なる場合、Model Studio プロジェクトのパーティションが使用されます。
- 既存の Model Studio プロジェクトに追加する場合、度数変数と重み変数は SAS Visual Analytics から転送されません。このような変数は、既存の Model Studio プロジェクトに対してすでに定義されています。
- 負の値を持つ頻度変数と重み変数は、Model Studio に転送されるときにゼロに置き換えられます。
- Model Studio から SAS Visual Analytics にモデルを転送することはできません。ただし、SAS Visual Analytics に入力データをコピーして、探索とビジュアル化を行うことができます。
- SAS Visual Analytics のデータアイテムを作成する特定のアクションは、Model Studio の**ビジュアルデータ準備**ノードで実行されます。たとえば、SAS Visual Analytics でクラスター ID アイテムを派生させると、データアイテムが**データペイン**に作成されます。Model Studio に転送される SAS Visual Analytics で作成されたこのデータアイテムを指定すると、プロジェクトの**データペイン**には表示されません。代わりに、**ビジュアルデータ準備**ノードが実行されるときに再作成されます。
- **Model Studio** で転送された **SAS Visual Analytics** モデリングノードのプロパティを有効にすると、モデルが再学習され、新しい結果が得られる場合があります。プロパティを有効にすると、モデルを元の **SAS Visual Analytics** スコアコードを使用するように戻すことはできません。モデルの解釈可能性プロパティは、モデルの再学習を必要としません。したがって、これらは自動的に有効になり、元の **SAS Visual Analytics** スコアコードには影響しません。
- パイプラインをテンプレートとして **Exchange** に保存することはできません。

非決定論的行動

一部の機械学習モデルは非決定論的なプロセスで作成されます。そのため、レポート内にモデルを作成し、そのレポートを保存し、レポートを閉じた後、後からレポートを開いたり印刷したりすると、異なる結果が表示されることがあります。この動作は、SAS Visual Analytics でモデルを作成し、後で SAS Visual Analytics で表示した場合にも発生する可能性があります。この動作は、決定論的パーティションが指定されている場合でも発生する可能性があります。

Model Studio で使用するパイプラインを作成する場合、安定したバージョンのモデルが作成されます。

自動調整

優れた統計量モデルを作成するには、使用するモデルとハイパーパラメーターを決定する必要があります。試行錯誤のアプローチを試すか、経験や個人的な選択に頼ることができます。ただし、いずれの場合でも、最適なデータモデルを検索できることは保証されません。SAS Viya は、機械学習オブジェクトを提供し、有用なモデルハイパーパラメーターを決定するプロセスを自動化できます。

自動調整は、モデルのハイパーパラメーターを自動的かつアルゴリズム的に調整して、ある特定のモデルの一連の競合バージョンを作成するプロセスです。これら競合するバージョンを比較して、どのセットのハイパーパラメーターが最適モデルを生成するかを決定します。ハイパーパラメーターとは、調整されるモデルのオプションのことです。

SAS Visual Analytics の SAS Viya 機械学習で使用可能なすべてのモデルを自動調整することができます。すべてのモデルについて、モデルの学習を実行する最大時間と、作成されるモデルの数を制限することができます。これらの制限を設定することにより、サーバーリソースが単一のユーザーまたはタスクによって独占されないようにします。パーティションが指定されていない場合オーバーフィッティングを避けるために、自動調整によって一時的に検証用のパーティションが作成されます。固定カットオフ値を使用すると、自動調整は実行後にそれを最適なカットオフ値に設定します。

自動調整セッションの結果を表示するには、**オプションペイン**で**以前の結果を表示**をクリックします。**自動調整の結果**テーブルでは、最適なモデルが強調表示されます。

ヒント オブジェクトを右クリックして、**自動調整の結果**を表示することもできます。**自動調整**を選択し、**以前の結果の表示**を選択します。

自動調整の結果テーブルの行を選択し、**モデルに適用**をクリックして、その行にあるハイパーパラメーターを適用します。既存の自動調整の結果はまだ使用可能です。

不完全な自動調整セッションをキャンセルすると、予期しない動作が発生する可能性があります。たとえば、自動調整クエリをキャンセルし、自動調整クエリを再起動すると、再起動されたクエリは終了しません。また、モデルで元の値が使用されていても、**オプションペイン**で自動調整されたオプションの値が更新されることがあります。

オプションペインで**自動調整の実行**をクリックして、**自動調整設定**ウィンドウで次のオプションを指定できます。

検索法

自動調整プロセスの各反復で使用するハイパーパラメーター値を決定するアルゴリズム。

ベイジアン

ベイジアン探索法は、目的値を近似するためにクリギングの代理モデルを構築し、この代理モデルを使用して、各反復で新しい代替構成を生成します。クリギングモデルは、検索プロセスの間、継続的に更新されます。

遺伝的アルゴリズム

遺伝的アルゴリズム(GA)は、最初にラテンハイパーキューブサンプリングを使用します。次に、LHS からの最良のサンプルが GA のシードに使用され、各イテレーションで代替構成の新しい母集団が生成されます。GA はデフォルトの検索オプションです。

グリッド

グリッド検索では、対象の各ハイパーパラメーターは、検討されるべき希望の値セットに離散化されます。モデルは、すべてのハイパーパラメーター全体に対してすべての値の組み合わせについて学習およびアセスメントされます(したがって、多次元の"グリッド"を形成します)。実行は簡単で分かりやすいのですが、グリッド検索は計算コストが高く、費用はそれぞれのハイパーパラメーターの数と離散値の数に応じて指数関数的に増加します。

ラテンハイパーキューブサンプル

ラテンハイパーキューブサンプル法(LHS)は、グリッド検索に似ていますが、グリッド検索よりも構造化されています。ラテンハイパーキューブサンプルは、各ハイパーパラメーター全体で一樣な方法で値を選択するが、組み合わせはランダムな組み合わせオブジェクトです。この基準では、スペースを効率的に埋めるため、ポイントは互いにほぼ等距離にあります。このサンプリングにより、ハイパーパラメーターの全範囲にわたるカバレッジが可能になり、ハイパーパラメーターごとに適切な値を見つける可能性が高くなります。ハイパーパラメーターごとに適切な値を使用して、適切な組み合わせを識別できます。

ランダム

ランダム検索法では、ハイパーパラメーター値のランダムな組み合わせが作成されます。一部のハイパーパラメーターは実際には特定のデータセットのモデルにほとんど影響を与えない可能性があるため、すべての組み合わせを評価するための努力を無駄にしないようにするのが賢明です。これは、高次元のハイパーパラメーター空間にとって特に重要です。ランダムな組み合わせを使用すると、同じ計算コストで各ハイパーパラメーターのより多くの値を探索できます。

目的メトリック

自動調整されたモデルを相互に比較するために使用される統計量です。使用される目的メトリックは検証パーティションからのものであり、それらの計算の結果が**自動調整の結果**テーブルに表示されます。分類モデルの場合、カットオフベースの目的メトリックは最適なカットオフ値を使用します。チャンピオンモデルには、この最適なカットオフが含まれています。

カテゴリ応答モデルでは、次のメトリックが使用可能です。

- 曲線下面積
- F0.5
- F1
- 偽陰性率
- 偽陽性率
- ガンマ分布
- ジニ
- Kolmogorov-Smirnov
- 誤分類の誤差率
- 誤分類率(デフォルト)
- 多群対数損失
- Tau
- 真陰性率
- 真陽性率

メジャー応答モデルでは、次のメトリックが使用可能です。

- 平均平方誤差(デフォルト)

- 平均絶対誤差
- 平均平方誤差
- 平均平方対数誤差
- 平均平方誤差の平方根
- 平均絶対誤差の平方根
- 平均平方対数誤差の平方根
- スコア時間
- 学習時間

最大分数

モデルの学習プロセスが実行される最大時間(分単位)。1 から 35,791,394 までの値を指定する必要があります。自動調整のアルゴリズムは、常に最低 1 分間実行されます。

モデルの最大数

自動調整プロセスの間に作成される異なるモデルの最大数。3 から 2,147,483,647 までの整数を指定する必要があります。

次のテーブルに、調整可能なハイパーパラメーターを示します。

ベイジアンネットワーク				
パラメーター	最小値	最大値	メモ	
ビンの数 ¹	2	100	値は整数であること	
事前スクリーン予測子	該当しません	該当しません	真と偽が評価される	
変数選択	該当しません	該当しません	真と偽が評価される	
独立性検定統計	該当しません	該当しません	デフォルトでは、すべての可能なテスト統計量が評価されます	
有意水準 ¹	0	1	実数値、0 を除く	
ネットワーク構造	該当しません	該当しません	デフォルトでは、可能なすべてのネットワーク構造が評価されます	
親の最大数 ¹	1	16	値は整数であること	
ペアレンティング法	該当しません	該当しません	デフォルトでは、すべての可能なペアレンティング法が評価されます	
ファクタライゼーションマシン				
パラメーター	最小値	最大値	メモ	
要因数 ¹	1	100	値は整数であること	
最大反復回数 ¹	1	10000	値は整数であること	
学習ステップ ¹	0	1.5	実数値、0 を除く	

フォレスト				
	パラメーター	最小値	最大値	メモ
	ツリー数 ¹	1	1000	値は整数であること
	ブートストラップ ¹	0.000001	1	実数値
	ノードを分岐する予測子の数 ¹	1	4	4 または最大入力数のうち小さい方を選択します。値は整数であること
	最大レベル数 ¹	2	51	値は整数であること
	リーフのサイズ ¹	1	2,147,483,647	値は整数であること
	予測子ビン ¹	2	500	値は整数であること
勾配ブースティング				
	パラメーター	最小値	最大値	メモ
	学習率 ¹	0.000001	1	実数値
	サブサンプル率 ¹	0.000001	1	実数値
	Lasso ¹	0	2,147,483,647	実数値
	リッジ ¹	0	2,147,483,647	実数値
	ノードを分岐する予測子の数 ¹	1	4	値は整数であること
	最大レベル数 ¹	2	51	値は整数であること
	リーフのサイズ ¹	1	2,147,483,647	値は整数であること
	予測子ビン ¹	2	500	値は整数であること
	ビンメソッド	該当しません	該当しません	デフォルトでは、すべての可能なメソッドが評価されます
ニューラルネットワーク				
	パラメーター	最小値	最大値	メモ
	学習率 ¹	1.0e-12	100	実数値。 最適化メソッド が SGD に設定されている場合にのみ使用できます。
	アニーリング率 ¹	0	100	実数値。 最適化メソッド が SGD に設定されている場合にのみ使用できます。
	L1 ¹	0	10	実数値
	L2 ¹	0	10	実数値

隠れ層の数 ¹	0	5	最適化メソッドが LBFGS に設定されている場合、最大値は 2 です。
隠れ層 N ニューロン ¹	1	1000	値は整数であること可能な隠れ層ごとに、ハイパーパラメーターが表示されます。このハイパーパラメーターは常に有効で、 隠れ層の数 ハイパーパラメーターに基づいています。ハイパーパラメーターの値を調整することはできますが、このプロパティを無効にすることはできません。
サポートベクターマシン			
パラメーター	最小値	最大値	メモ
カーネル関数	該当しません	該当しません	デフォルトでは、すべての可能なカーネル関数が評価されます
ペナルティ 値 ¹	1.0e-10	100,000	実数値
ディシジョンツリー			
パラメーター	最小値	最大値	メモ
最大レベル数 ¹	2	20	値は整数であること
リーフのサイズ ¹	1	2,147,483,647	値は整数であること
予測子ビン ¹	2	1500	値は整数であること

¹ この特性は、有効な値の範囲または特定の値の離散的なセットで自動調整することができます。

一般的なモデリングタスク

モデルのエクスポートとスコアリング

モデルスコアリングとは、対象応答変数を含んでいない可能性のあるデータセットの予測値を生成するプロセスを示します。エクスポートされたスコアコードは、任意の SAS 環境の新しいデータセットで実行できます。インタラクション項やグループ化変数など、モデルによって使用されるすべての変数がスコアコードに含まれます。

モデルに応じて、DATA ステップコードまたは分析ストアテーブルを生成することを選択できます。

DATA ステップをエクスポートしてスコアリングするには、次のようにします。

1 モデルキャンバスを右クリックし、**スコアコードを生成** ⇨ **DATA ステップ**を選択します。

- 2 **コードのエクスポート**をクリックして、SAS DATA ステップコードをブラウザーの**ダウンロード**フォルダーにダウンロードします。
- 3 コードを実行して、新規データをスコアリングします。

DATA ステップスコアコードは.sas ファイルとして保存され、任意のワープロプログラムで表示できます。

注: エクスポートされた DATA ステップスコアコードには、最大行長 32768 を超えるコード行が含まれている可能性があります。この問題には 2 つの解決策があります。第 1 の解決策では、長い行それぞれに改行を挿入し、/lrecl=1000000 を %INCLUDE ステートメントに Insert するようエクスポートされたテキストファイルを編集する必要があります。第 2 の解決策では、エクスポートされたテキストファイルをコードエディターで開き、長い行それぞれに改行を挿入する必要があります。コードエディターでは、1 行に 6000 文字の制限があります。

分析ストアテーブルをエクスポートしてスコアリングするには、次のようにします。

- 1 モデルキャンバスを右クリックし、**スコアコードを生成** ⇨ **ASTORE** を選択します。
- 2 **モデルの保存** ウィンドウで、分析ストアテーブルの基本名を指定します。**保存**をクリックします。分析ストアテーブルの名前では大文字と小文字が区別されます。
分析ストアテーブルが作成され、SAS Cloud Analytic Services(CAS)サーバーのモデルライブラリに保存されます。
- 3 **コードのエクスポート**をクリックして、SAS DATA ステップコードをブラウザーの**ダウンロード**フォルダーにダウンロードします。
- 4 ダウンロードしたコードを開きます。コードのコメントに示されているように、マクロ変数を変更します。
 - **SOURCE_HOST:** 分析ストアテーブルが保存される CAS サーバーのホスト名。
 - **SOURCE_PORT:** 分析ストアテーブルが保存される CAS サーバーのポート番号。
 - **SOURCE_LIB:** スコアリングするソースデータが配置されている CAS ライブラリ。
 - **SOURCE_DATA:** スコアリングするソースデータの CAS テーブル名。
 - **DEST_LIB:** スコアリングされたデータを保存する CAS ライブラリ。
 - **DEST_DATA:** スコアリングされたデータの CAS テーブル名。

注: CAS サーバーに関する情報を確認するには、次のようにします。

- 1 ウィンドウの左上隅にあるをクリックし、**環境の管理**を選択します。
 - 2 左ペインでをクリックし、**サーバー**ウィンドウにアクセスします。サーバー名、ホスト、およびポートはこのテーブルにリストされています。
-

- 5 コードを実行して、新規データをスコアリングします。

機械学習オブジェクトの入門ガイド

概要

この概要では、機械学習オブジェクトを使用して新しい変数を派生させ、3つの異なるモデルを作成し、それらのモデルを比較しています。この例では、<http://support.sas.com/documentation/onlinedoc/viya/examples.htm>にある Framingham Heart Study データセットを使用し、フォレスト、勾配ブースティングおよびニューラルネットワークのパフォーマンスを比較しています。目標は、収集された健康因子に基づいて死亡年齢を予測することです。これらの因子には、性別、体重、身長、喫煙者であるかどうか、血圧などが含まれます。この例では、最適なモデルの構築方法ではなく、SAS Viya 機械学習オブジェクトの使用方法に注目しています。

サンプルデータをダウンロードする

- 1 Web ブラウザーで、<http://support.sas.com/documentation/onlinedoc/viya/examples.htm> に移動します。
- 2 heart.csv ファイルをローカルマシンにダウンロードします。

レポートの作成

- 1 SAS Visual Analytics にサインインします。
- 2 ホームページで、**レポートの新規作成**をクリックします。**データペイン**と**レポートキャンバス**が表示されます。
- 3 **データペイン**を選択して、**データのインポート**をクリックします。**データのインポートウィンドウ**が表示されます。
- 4 **インポートペイン**で、**ローカルファイル**をクリックします。
- 5 heart.csv ファイルを保存した場所に移動します。**heart.csv**を選択し、**開く**をクリックします。**データのインポートウィンドウ**が表示されます。
- 6 **アイテムのインポート**をクリックします。
- 7 テーブルが正常にインポートされたら、**レポートに追加**をクリックします。インポートされたデータが**データペイン**に表示されます。

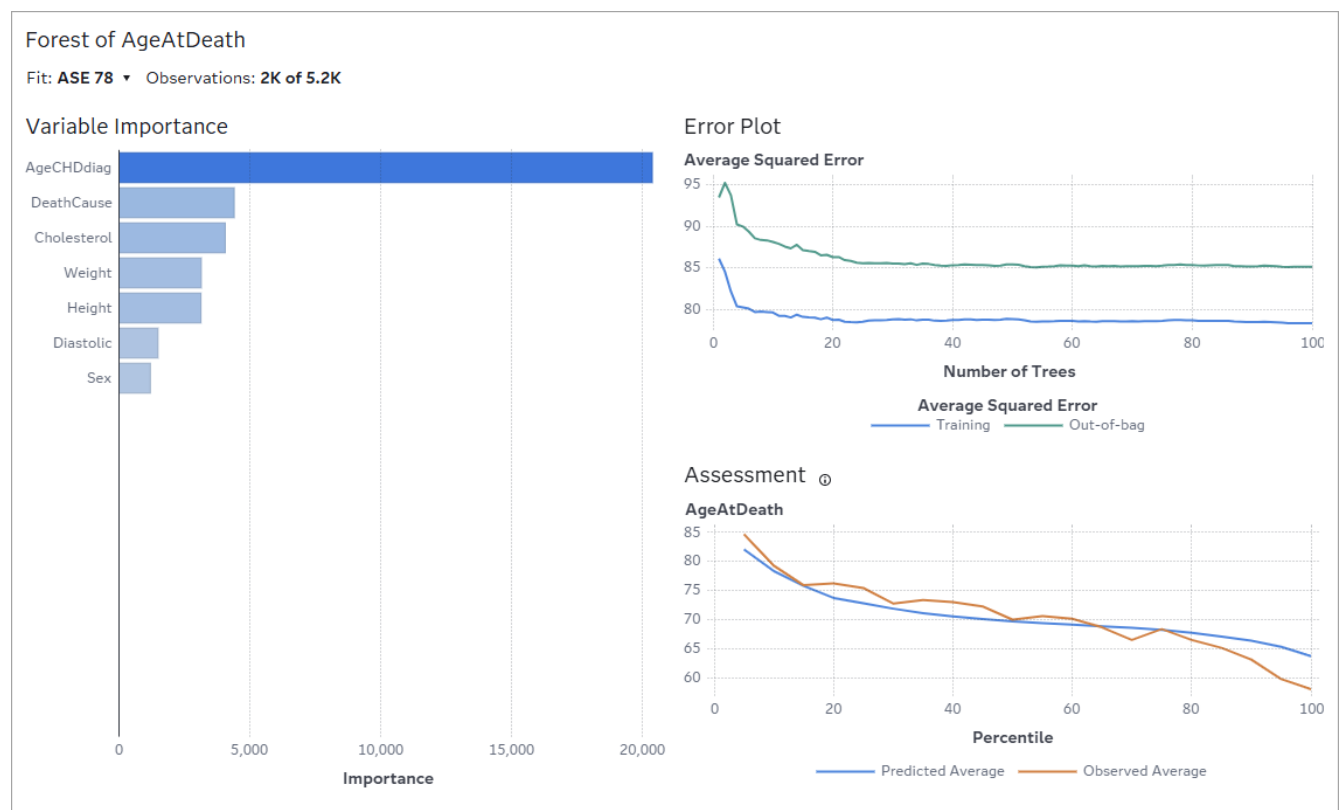
- レポートツールバーで、**名前を付けて保存**を選択します。名前を付けて保存ウィンドウで、書き込み権限のある場所に移動します。名前フィールドに **Heart Study** と入力し、保存をクリックします。

通常、**マイフォルダー**に作業内容を保存できます。

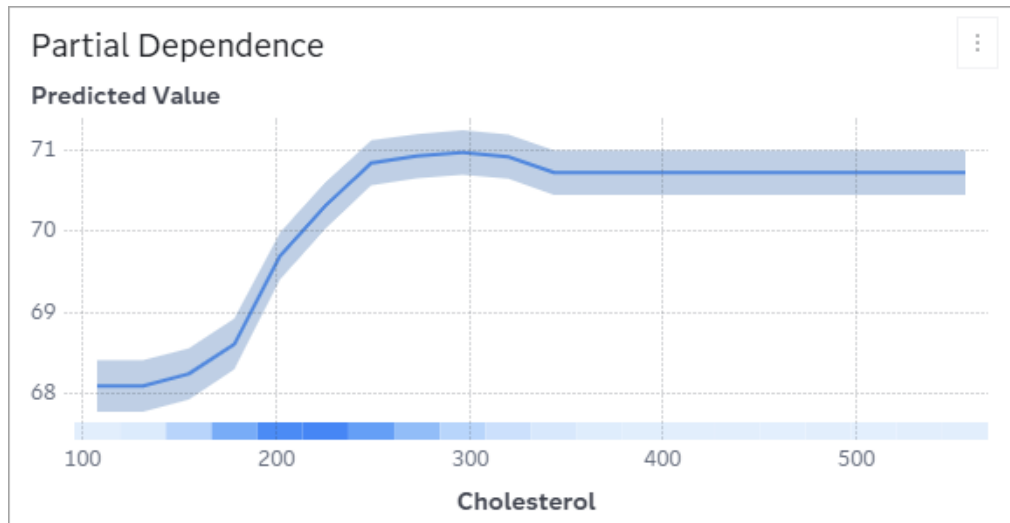
レポートの詳細情報については、[SAS Visual Analytics: Designing Reports](#) を参照してください。

フォレストの作成

- 横のペインで、**変数**をクリックしてオブジェクトを選択します。フォレストを作成するには、**変数**をキャンバスにドラッグします。
- フォレストモデルをの中央にある**データの割り当て**ボタンをクリックします。応答フィールドで、**追加**をクリックし、**AgeAtDeath** を選択します。
- 予測子**フィールドで、**追加**をクリックし、**DeathCause**、**Sex**、**AgeCHDdiag**、**Cholesterol**、**Diastolic**、**Height** および **Weight** を選択します。**適用**をクリックし、**閉じる**をクリックします。フォレストモデルが自動的に更新されます。



- 誤差プロットで右クリックし、**部分従属**を選択します。
- 部分従属プロットを右クリックし、**Cholesterol** を選択します。



コレステロール値が増加すると、死亡年齢の予測値も増加することに注意してください。コレステロールの値が 300 に達すると、コレステロールの増加はモデルの予測にほとんど影響を与えません。

プロットの下部にあるヒートマップは、プロットを作成するためにサンプリングされたオブザベーションの大多数が 175~250 の範囲のコレステロール値を持つことを示しています。

- 6 をクリックして、最大化モードに入ります。

詳細テーブルでは、モデルに関する詳細情報を表示できます。これには、変数の重要度の値とエラー基準の結果が含まれます。

- 7 をクリックして、最大化モードを終了します。
- 8 レポートツールバーで、 をクリックして、レポートを保存します。

勾配ブースティングモデルの作成

- 1 をクリックし、レポートに新しいページを追加します。
- 2 横のペインで、 をクリックしてオブジェクトを選択します。 をキャンバスにドラッグして勾配ブースティングモデルを作成します。
- 3 この例では、関心のある変数は **AgeAtDeath** です。
勾配ブースティングモデルの中央にある **データの割り当て** ボタンをクリックします。応答フィールドで、**追加** をクリックし、**AgeAtDeath** を選択します。
- 4 予測子フィールドで、**追加** をクリックし、**BP_Status**、**DeathCause**、**Sex**、**Smoking_Status**、**Cholesterol**、**Height**、**Smoking**、および **Weight** を選択します。**適用** をクリックし、**閉じる** をクリックします。勾配ブースティングモデルが自動的に更新されます。
- 5 横のペインで、 をクリックし、次に **自動調整の実行** をクリックします。

Autotuning Setup

Search method:

Genetic algorithm

Objective metric:

Average square error

Setting items below will determine when autotuning is complete.

10

Maximum minutes

50

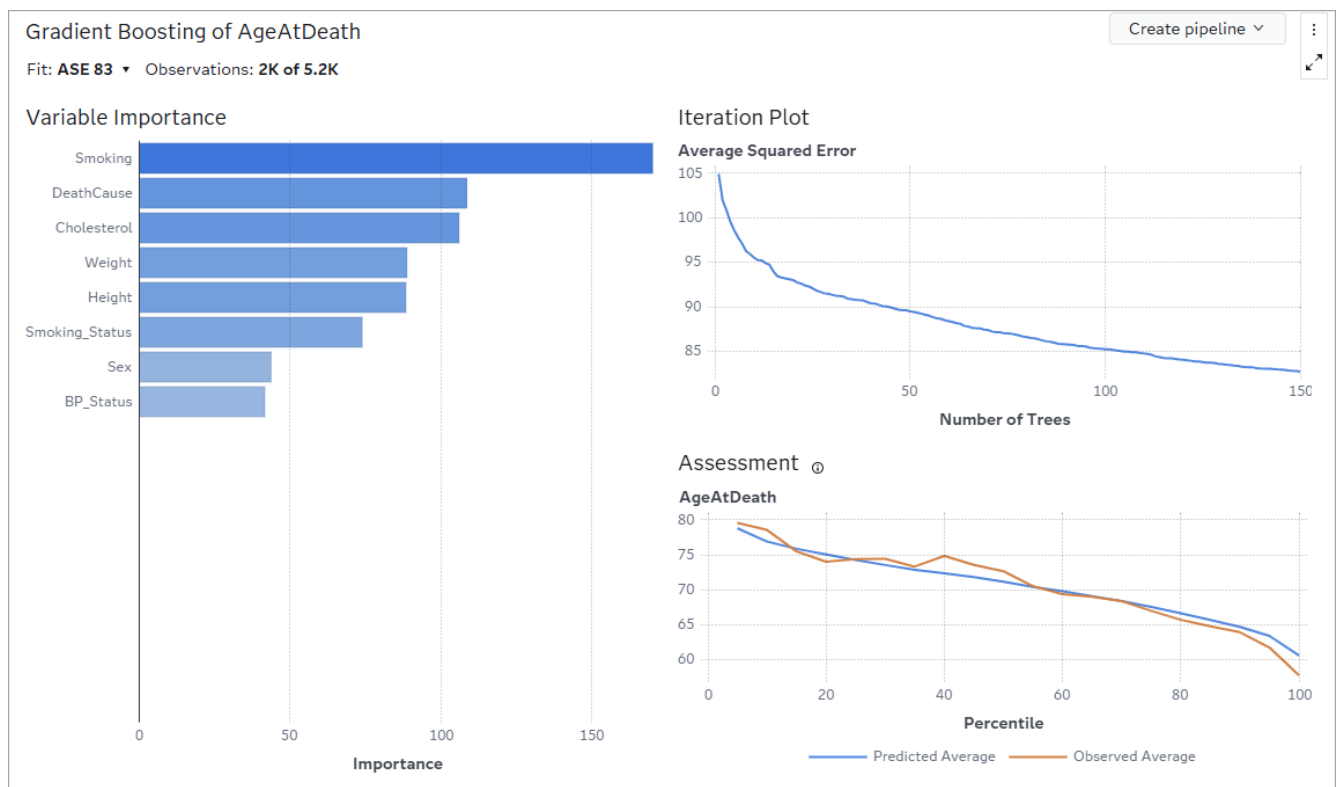
Maximum number of models

> Hyperparameters

OK

Cancel

モデルを自動調整するときは、SAS Viya 機械学習は指定されたプロパティの最適値を機械学習自動的かつアルゴリズム的に決定します。

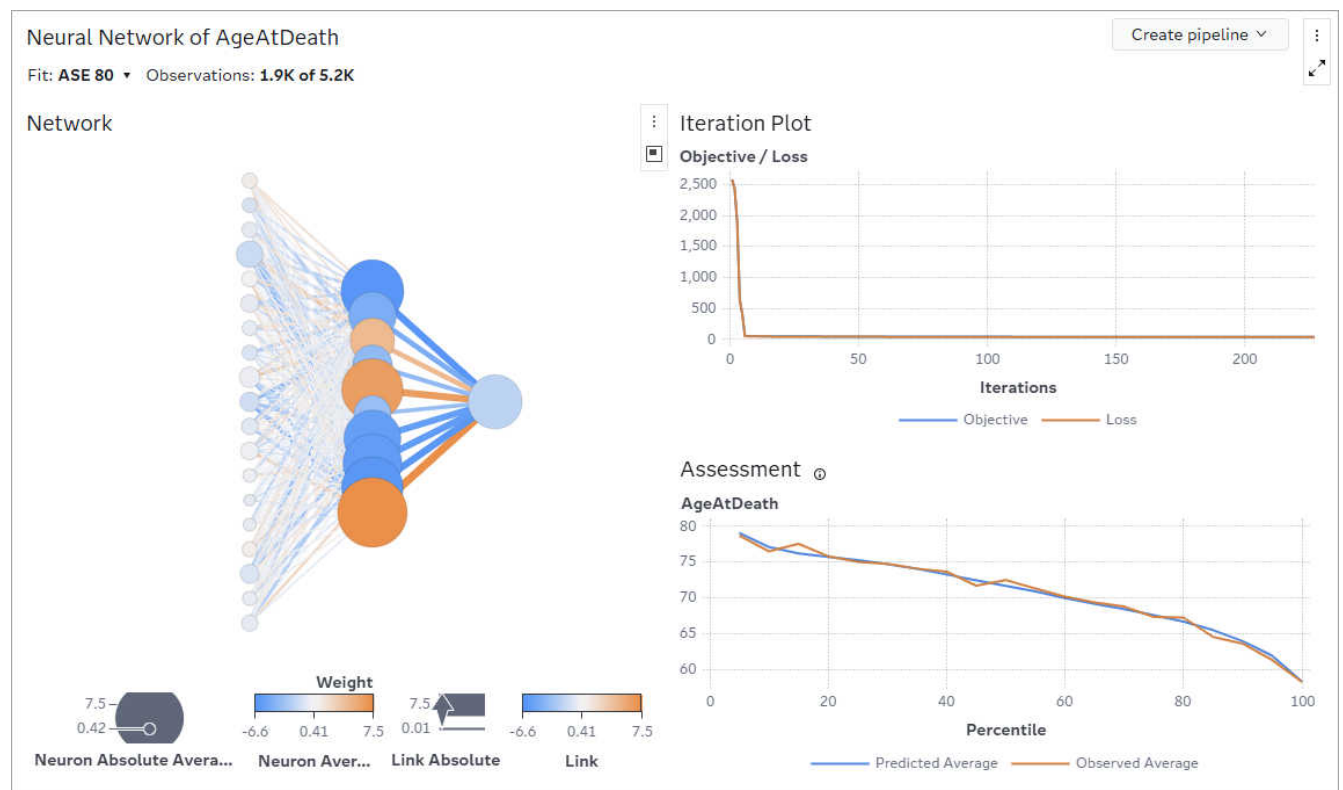


評価プロットは、観測平均と予測平均がビン間でわずかに異なることを示しています。

6 レポートを保存します。

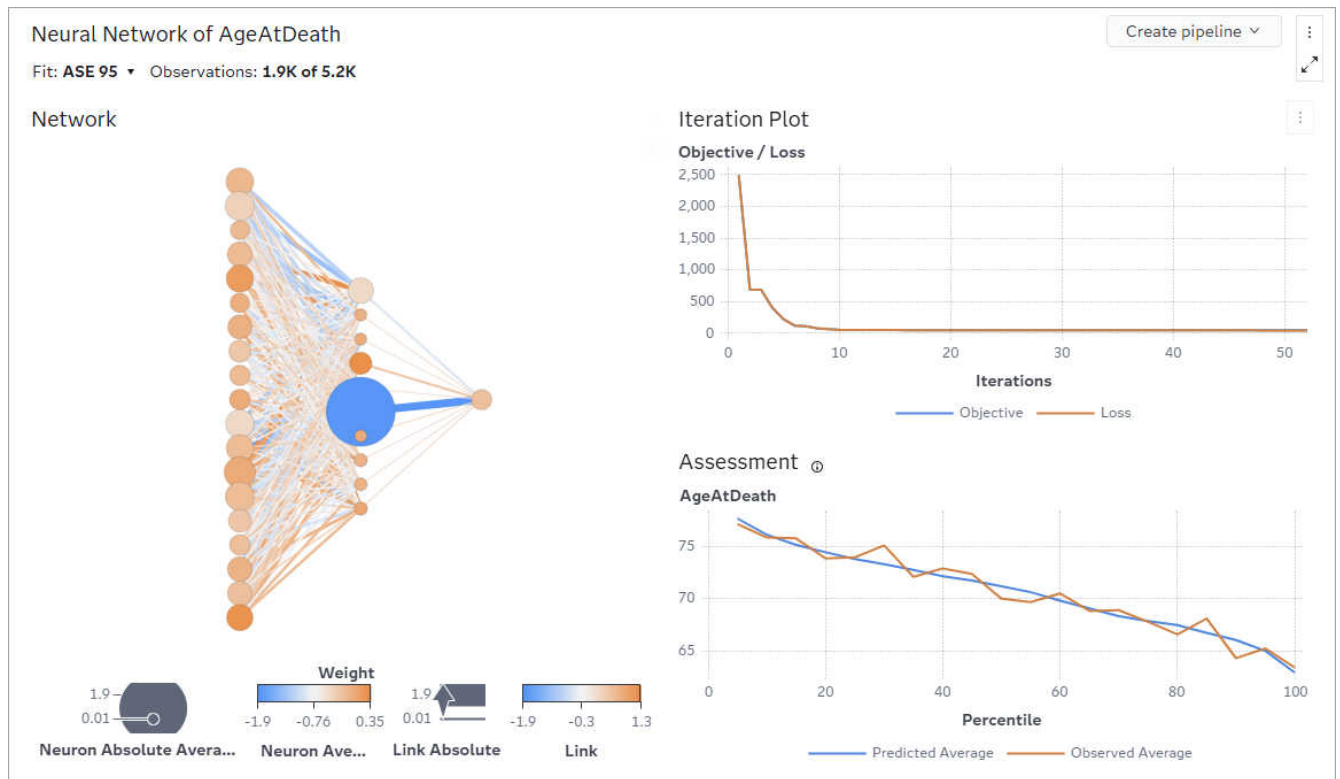
ニューラルネットワークの作成

- 1 **+**をクリックし、レポートに新しいページを追加します。
- 2 横のペインで、**📄**をクリックしてオブジェクトを選択します。**🔗**をキャンバスにドラッグしてニューラルネットワークを作成します。
- 3 ニューラルネットワークの中央にある**データの割り当て**ボタンをクリックします。応答フィールドで、**追加**をクリックし、**AgeAtDeath**を選択します。
- 4 **予測子**フィールドで、**追加**をクリックし、**BP_Status**、**DeathCause**、**Sex**、**Smoking_Status**、**Cholesterol**、**Height**、**Smoking**、および **Weight** を選択します。**適用**をクリックし、**閉じる**をクリックします。ニューラルネットワークは自動的に更新されます。



- 5 横のペインで、**📊**をクリックします。**分布オプション**を使用すると、応答変数の分布を指定し、その分布に基づいてモデルを構築できます。デフォルトの分布は**正規分布**です。
 応答変数に正規分布が適用されるかどうかを判断するには、**+**をクリックして新しいページをレポートに追加します。**データ**ペインで、**AgeAtDeath**を選択し、キャンバスにドラッグします。ヒストグラムが自動的に作成されます。
- 6 **AgeAtDeath** が正規分布しておらず、少し左に歪んでいることに注目してください。ニューラルネットワークに戻るには、**3 ページ**をクリックしてください。
- 7 分布は正確には Poisson ではありませんが、この例では Poisson 分布を使用します。**分布オプション**の場合は、**Poisson**を選択します。

注: この例をさまざまな分布と標準化手法で繰り返し、そのパフォーマンスを比較して理解を深め、SAS Viya 機械学習モデルに慣れてください。



8 レポートを保存します。

モデルの比較の実行

- 1 **+**をクリックし、レポートに新しいページを追加します。
- 2 横のペインで、**⌵**をクリックしてオブジェクトを選択します。**🖱**をキャンバスにドラッグし、モデルの比較を作成します。

×

Add Model Comparison

Data:

HEART (1)

Response:

AgeAtDeath

Available models:

[Select all](#)

☐

Forest - AgeAtDeath 1

☐

Gradient boosting - AgeAtDeath 1

☐

Neural network - AgeAtDeath 1

OK

Cancel

応答変数は **AgeAtDeath** にすでに設定されており、イベントレベルとグループ化する基準は使用できません。

これらの設定では、使用可能なモデルは、**フォレスト - AgeAtDeath 1**、**勾配ブースティング - AgeAtDeath 1**、および**ニューラルネットワーク - AgeAtDeath 1**です。

- 3 **すべて選択**を選択し、**OK** をクリックします。



デフォルトでは、当てはめの統計量平均平方誤差(**ASE**)がモデルの比較に使用されます。**当てはめの統計量** Option を使用すると、モデルの比較に使用される当てはめの統計量を変更できます。使用可能なその他の当てはめの統計量は、**SSE** と **観測平均** です。値が小さいほど好ましいため、**ASE** または **SSE** が基準である場合、フォレストモデルがチャンピオンとして選択されます。モデルは非常に似ています。

当てはめの統計量が**観測平均**の場合、**百分位**スライダーが使用可能です。このスライダーは、観測平均と予測平均が比較される百分位を指定します。チャンピオンモデルは百分位によって異なります。

評価プロットを確認すると、平均観測プロットと予測平均プロットの両方が、モデルが著しく異なっている場合があることを示しています。

- 4 チャンピオンモデルがあれば、そのモデルのモデルスコアコードをエクスポートして、新しいデータにスコアを付けることができます。詳細については、[モデルのエクスポートとスコアリング \(10 ページ\)](#)を参照してください。
- 5 レポートを保存します。

ベイジアンネットワークの操作

ベイジアンネットワークの概要

ベイジアンネットワークは、ノードがランダム変数を表し、ノード間のリンクが2つのランダム変数間の条件付き依存関係を表す有向非巡回グラフモデルです。ベイジアンネットワークの構造は、変数間の条件付き依存関係に基づいています。その結果、ネットワーク内の各ノードに対して条件付き確率テーブルが作成されます。これらの機能のおかげで、ベイジアンネットワークは教師ありデータマイニングの間、効果的な予測モデルになることができます。次のベイジアンネットワーク構造が使用可能です。

- Naive は、ターゲット変数を各入力変数に接続するベイジアンネットワークです。入力変数は条件的に互いに独立していると仮定されているため、変数間に他の接続はありません。
- Tree-augmented naive は、ターゲット変数を各入力変数に接続し、入力変数をツリー構造で接続するベイジアンネットワークです。入力変数を接続するツリーは、最大スパニングツリーアルゴリズムに基づいています。Tree-augmented naive ネットワークでは、各ノードは最大で2つの親を持つことができ、そのうちの1つはターゲット変数でなければなりません。
- 親子は、ターゲット変数を各入力変数に接続するベイジアンネットワークです。ただし、入力変数はターゲット変数の親またはターゲット変数の子のいずれかになります。ベイジアン情報基準(BIC)は、入力変数がターゲット変数の親かターゲット変数の子かを判断するために使用されます。
- Markov blanket は、ターゲット変数と入力変数との間に一連の接続を作成するベイジアンネットワークです。特定の入力変数間の接続も許可します。ターゲット変数の子のみが追加の親ノードを持つことができます。他のすべての変数は、条件付きでターゲット変数とは独立しています。したがって、それらの変数は分類モデルには影響しません。BIC は、入力変数がターゲット変数の親であるかターゲット変数の子であるかを判別するために使用されます。

ベイジアンネットワークの作成

ベイジアンネットワークを作成するには、次の手順を実行します。

- 1 左ペインで、をクリックしてオブジェクトを選択します。をキャンバスにドラッグしてベイジアンネットワークを作成します。
- 2 右ペインのをクリックします。1つのカテゴリ変数を**応答変数**として指定します。
- 3 **予測子**に少なくとも1つの変数を指定します。**予測子**には、メジャー変数またはカテゴリ変数のみを指定できます。
- 4 (オプション) **パーティション ID** を指定します。

注: パーティション ID を設定するには、パーティション変数が必要です。新しいパーティション変数の作成については、“[Create a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。既存のデータアイテムにパーティション変数を割り当てる方法については、“[Assign a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。

ベイジアンネットワークオプション

ベイジアンネットワークでは、次のオプションを指定できます。

全般

イベントレベル

関心のあるイベントレベルを選択できます。カテゴリ応答変数に 2 つ以上のレベルが含まれている場合、SAS Viya 機械学習は関心のあるレベルのすべてのオブザベーションをイベントとして、他のすべてのオブザベーションを非イベントとして扱います。

自動調整

自動調整アルゴリズムを制御する制約を指定できます。制約は、アルゴリズムの実行時間、アルゴリズムの実行回数、指定できるモデル評価の数を決定します。詳細については、“[自動調整](#)” (6 ページ)を参照してください。

自動調整アルゴリズムは、最適なモデルを生成する次のパラメーターの値を選択します。

- **ビンの数**
- **事前スクリーン予測子**
- **変数選択**
- **独立性検定統計**
- **有意水準**
- **ネットワーク構造**
- **親の最大数**
- **ペアレンティング法**

欠損値を含める

欠損値がモデルに含まれているかどうかを指定します。欠損しているカテゴリ値は、自身の測定レベルとして扱われます。欠損しているメジャー変数は、その変数の平均値に転換されます。

変数選択を実行し、欠損値を除外するベイジアンネットワークでは、使用されるオブザベーション数は、パーティションデータ項目の有無により異なる場合があります。分割されたモデルの場合、すべての予測子の非欠損値がモデルの学習に使用されます。ただし、CAS サーバーによって実行される検証アクションでは、選択した予測子の欠損値のみが拒否されます。

ビンの数

メジャー変数のビンの数を指定します。指定可能な値の範囲は 2～100 です。

事前スクリーン予測子

ターゲット変数と各入力変数間の独立性検定を使用して変数を事前スクリーンするかどうかを指定します。デフォルトでは、このオプションが選択されています。

変数選択

ネットワークに基づき、ターゲット変数と各入力変数間の条件付き独立性検定を使用して変数を選択するかどうかを指定します。デフォルトでは、このオプションは選択されていません。

独立性検定統計

独立性検定に使用される統計を指定します。指定可能な値は、**カイ 2 乗**、**G 2 乗値**、**カイおよび G 2 乗値**です。両方の独立性検定の統計を使用するには、**カイ 2 乗および G 2 乗**を選択します。独立性検定では、両方の統計を満たす必要があります。デフォルト値は **G 2 乗**です。

有意水準

カイ 2 乗と G 2 乗を使用して独立性テストの有意水準(p 値)を指定します。指定する値が小さいほど、選択される変数は少なくなります。デフォルト値は 0.2 です。

ネットワーク構造

ネットワーク構造を指定します。複数のネットワーク構造を選択できます。次が使用可能です。

- **Naive**
- **Tree-augmented naive**
- **Parent-child**
- **Markov blanket**

これらの構造の中で最適な構造を選択する場合は、任意の組み合わせで複数の値を選択できます。これらの構造の詳細については、この章の概要を参照してください。デフォルトでは、すべての構造が選択されています。

親の最大数

ネットワーク構造内の各ノードに許可される親の最大数を指定します。可能な値の範囲は 1～16 です。デフォルト値は 5 です。

最適な構造と親の数を選択する

1 から始まって **親の最大数** の値で終わる最適な親の数を選択するかどうかを指定します。デフォルトでは、このオプションが選択されています。複数のネットワーク構造が選択されている場合、このオプションは常に有効です。

ペアレンティング法

各ノードのペアレンティング方を指定します。指定できる値は、**1 つの親と親のセット**です。最適な候補をノードの親として追加するには、**1 つの親**を選択します。各ノードの親である可能性のある候補の中から変数のセットをテストするには、**親のセット**を選択します。次に、最適な変数のセットがノードの親として追加されます。デフォルト値は **親のセット**です。

部分従属

オブザベーションの数

入力データからサンプリングするオブザベーションの最大数を指定します。これらのオブザベーションは、部分従属プロットに表示されるモデル予測を生成するために使用されます。

ビンの数

部分従属プロットでメジャー変数を分類するために使用するビンの数を指定します。

評価

ビンの数

評価で使用するビンの数を指定します。5 から 100 の範囲の整数値を指定します。

カットオフ

計算された確率がイベントと見なされる値を指定します。計算されたカットオフ値を使用するには、**最適カットオフ**を選択します。独自のカットオフ値を設定するには、**固定カットオフ**を選択します。

統計パーセント点

観測平均、リフト、累積リフト、累積 % 捕捉、累積 % イベント、およびゲインの計算に使用される百分位のビンの深さを指定します。

許容値

百分位を推定する反復アルゴリズムの収束を決定するのに使用される許容値を指定します。アルゴリズムの精度を高めるには、小さい値を指定します。

ベイジアンネットワークモデル表示オプション

ベイジアンネットワークでは、次の表示オプションを指定できます。

全般

表示されたビジュアル

表示されるオブジェクトの要素を選択します。

グラフのレイアウト

オブジェクトをキャンバスに表示する方法を指定します。**自動調整**は、キャンバス上のすべてのオブジェクトを自動的に揃えます。**積み上げ**は、オブジェクトをスライドデッキ形式のように表示します。一度に表示できるオブジェクトは1つのみです。**積み上げ**が指定されている場合、オブジェクト間の移動にスクロールバーではなくコントロールバーが用意されています。

表示する統計量

モデルに表示する評価統計量を指定します。パーティション変数を使用している場合、パーティションごとに次の当てはめの統計量を使用できます。オブジェクトツールバーには、使用可能な各パーティションタイプ(学習、検証、およびテスト)のサブメニューが含まれています。

- C 統計量
- 累積 % 捕捉
- 累積 % イベント
- 累積リフト
- F1 スコア
- FDR
- FPR
- ゲイン
- ガンマ分布
- ジニ
- KS (Youden)
- リフト
- 誤分類率
- 誤分類率 (イベント)
- Tau

使用可能な当てはめの統計量の詳細情報については[当てはめの統計量](#)を参照してください。

数値を省略表記

大きなメジャー値が省略表記の数値を使用することを指定します。たとえば、**1,100,000,000** は **1.1B** のように表示されます。**スケール**と**精度の桁**を指定できます。

ヘッダー**表示するフィールド**

表示されるヘッダーの要素を指定します。

オブザベーションの形式

モデルで使用されたオブザベーションに関する情報の表示方法を指定します。2つの異なる値が指定された精度の桁で表示される場合、明確にするために追加の桁が追加されます。

ネットワーク/部分従属プロットの変数**表示するプロット**

ネットワークプロットまたは部分従属プロットの変数を表示するかどうかを指定します。

X 軸

部分従属プロットに表示される変数を指定します。

評価プロット**テストパーティションの表示**

テストパーティションの評価プロットを表示するかどうかを指定します。このオプションは、テストパーティションでパーティション ID を指定する場合にのみ使用できます。

表示するプロット

どの評価プロットを表示するかを指定します。**混同行列**、**リフト**、**ROC**、**カットオフプロット**、または**誤分類**を選択します。

カットオフプロット統計量

表示される評価統計量を指定します。**カットオフ**オプションの値は、常に垂直線として表示されます。次の評価統計量が使用可能です。

- **精度**
- **F0.5 スコア**
- **F1 スコア**
- **偽発見率**
- **偽陰性率**
- **偽陽性率**
- **Kolmogorov-Smirnov (KS)**
- **感度**
- **特異度**

Y 軸

標準リフトプロットまたは累積リフトプロットを表示するかどうかを指定します。

凡例の表示

評価プロットに凡例を表示するかどうかを指定します。

ベジアンネットワークの結果

ネットワークプロット

ネットワークプロットには、誤分類率が最も高いベジアンネットワークが表示されます。選択されたベジアンネットワークは、モデル選択プロットに★アイコンで示されます。

Variables in Network プロット

Variables in Network プロットは、BIC スコアでソートされたモデルの作成に使用される変数を示します。

部分従属プロット

部分従属(PD)プロットは、モデルの予測が特定の予測変数の値にどのように依存するかを理解するのに役立ちます。このプロットは、特定の予測変数の値が変更したときに応答の値がどのように変更するかを判断するのに役立ちます。Y 軸では、カテゴリ応答モデルはイベントの予測確率を表示し、メジャー応答モデルは予測応答値を表示します。どちらの場合も、値はそのビンのオブザベーションの平均です。多項モデルの場合、イベントレベルを変更して、他の関係を調べることができます。X 軸には、指定された予測変数の値が表示されます。デフォルトで表示される予測変数は、重要度が最大の予測変数です。プロットには、点推定と 95%信頼区間を示す線が表示されます。プロットの下部にあるヒートマップは、サンプリングされたオブザベーションの密度を示しています。一部のビンのオブザベーションがゼロになる場合があります。部分従属プロットは、パーティション変数を使用している場合でも、すべてのオブザベーションのサンプルを使用して生成されます。

モデル選択プロット

モデル選択プロットは、**親の最大数値**が変化すると、ベジアンネットワークの誤分類率がどのように変化するかを表示します。誤分類率はネットワーク構造によってグループ分けされ、最良のモデルは★アイコンで示されます。最良のモデルはネットワークプロットに表示されています。

定義上、**Tree-augmented naive** ネットワークには厳密に 2 つの親が含まれています。モデル選択プロットのこのセクションには常に、プロットされた値が 1 つだけ含まれます。

評価プロット

混同行列には、カテゴリの応答モデルの分類結果が表示されます。モデルの作成後、各オブザベーションには観測値と予測値があります。各観測/予測ペアの総数が計算されます。混同行列は、各ペアに含まれるオブザベーションの数を表示します。完全なモデルは常に観測値を予測し、すべての値は行列の対角線上にあります。非対角値は誤分類を表します。セルは、各レベルの観測値の数に対する各セルの値の割合に基づいて陰影が付けられます。暗い影付きのセルは、観測水準の予測の集中度を示しています。バイナリ応答の場合、混同行列の情報は誤分類プロットと同じです。

リフトは、各百分位のビン内の捕捉応答のパーセントとモデルの応答の平均パーセントの比率です。同様に、累積リフトは、現在の百分位のビンまでのすべてのデータを使用して計算されます。

受信者操作特性(ROC)チャートには、偽陽性分類および偽陰性分類を回避するモデルの能力が表示されます。偽陽性分類とは、オブザベーションが実際には非イベントである場合(タイプ I エラーとも呼ばれる)、イベントとして特定されたことを意味します。偽陰性分類とは、オブザベーションが実際にイベントである場合(タイプ II エラーとも呼ばれる)、非イベントとして特定されたことを意味します。

モデルの**特異度**は真陰性率です。偽陽性率を派生させるには、特異度を 1 から引きます。**1 - 特異度**とラベル付けされた偽陽性率は、ROC チャートの X 軸です。モデルの**感度**は真陽性率です。これは ROC チャートの Y 軸です。したがって、ROC チャートは、偽陽性率が変化すると真陽性率がどのように変化するかをプロットしています。

優れた ROC チャートには、非常に急峻な初期スロープがあり、すぐに横ばいになります。つまり、オブザベーションの誤分類ごとに、著しく多くのオブザベーションが正しく分類されています。偽陽性も偽陰性もない完璧なモデルでは、ROC チャートは(0,0)から始まり、垂直に(0,1)に上がり、水平に(1,1)になります。この例では、誤分類が 1 回発生する前に、モデルはあらゆるオブザベーションを正しく分類します。

ROC チャートには、その解釈に役立つ 2 つの線があります。最初の線は、傾きが 1 のベースラインモデルです。この線は、誤って分類するのと同率でオブザベーションを正しく分類するモデルを模倣しています。理想的な ROC チャートは、ベースラインモデルと ROC チャート間の距離が最大になっています。オブザベーションを正しくではなく誤って分類するモデルは、ベースラインモデルよりも低くなります。2 番目の線は、ROC チャートとベースラインモデルの Kolmogorov-Smirnov 値の差が最大になっている偽陽性率の垂直線です。

カットオフプロットでは、カットオフ値がモデルのパフォーマンスにどのように影響するかを視覚化できます。カットオフ値は**カットオフ**オブションで指定され、プロットでは垂直線で表されます。カットオフ値は、オブザベーションをイベントとしてモデル化するか、非イベントとしてモデル化するかを決定します。垂直線をドラッグしてカットオフ値を調整することで、モデルの再評価が行われます。

誤分類プロットには、応答変数の各値に対していくつのオブザベーションが正しくまた誤って分類されたかが表示されます。応答変数がバイナリでない場合、ベイジアンネットワークモデルは、イベントではないすべてのレベルを等しいとみなします。著しい数の誤分類は、モデルがデータに適合していないことを示している可能性があります。

詳細テーブル

オブジェクトツールバーのをクリックすると、詳細テーブルがキャンバスの下部に表示されます。詳細テーブルには、次の情報が含まれています。

モデル情報

モデルの概要を示します。

オブザベーションの数

さまざまなモデル作成タスクに使用されたオブザベーションの数の要約です。

当てはめ統計量

当てはめの統計量すべてを一覧表示します。

モデル選択

さまざまなモデルパラメーターに基づいてモデル選択結果の概要を示します。

変数選択

選択プロセスの各手順での変数選択結果の要約を示します。

変数レベル

カテゴリ変数レベルとモデルで使用されるインデックス値とのマッピングを提供します。

変数の順序

モデルで使用されている変数を BIC スコアでソートして一覧表示します。

混同行列

混同行列の生成に使用されるモデルの正しい分類と誤った分類の要約を提供します。

リフト

リフトプロットの生成に使用されるビニングされた評価結果を一覧表示します。

ROC

ROC プロットの生成に使用される結果を一覧表示します。

カットオフ統計

カットオフ値の変更に伴って変化するいくつかの統計の値を一覧表示します。

誤分類

モデルの正しい分類と誤った分類の要約を表示します。

評価統計量

モデルに対して計算された評価統計量の値を表示します。

評価要約

モデルに対して計算された評価統計量の説明を表示します。応答の種類とパーティション変数の割り当ての有無によって、表示される出力が決まります。

ファクタライゼーションマシンについて

ファクタライゼーションマシンの概要

ファクタライゼーションマシンは、因数分解モデルを作成する予測モデルです。因子分解されたパラメーターとのすべての変数の相互作用をモデル化することにより、ファクタライゼーションマシンは大規模で非常に疎なデータを扱うことができ、線形時間で学習できます。

ファクタライゼーションマシンの一般的な用途は、推奨エンジンです。ファクタライゼーションマシンは、ユーザーが評価したすべてのアイテムを考慮し、他のアイテムの評価を予測できます。

ファクタライゼーションマシンの作成

- 1 左ペインで、をクリックしてオブジェクトを選択します。をキャンバスにドラッグしてファクタライゼーションマシンを作成します。

- 2 右ペインのをクリックします。単一のメジャー変数を**応答変数**として指定します。
- 3 **予測子変数**には、少なくとも 2 つのカテゴリ変数と任意の数のメジャー変数を指定します。

ファクタライゼーションマシンのオプション

ファクタライゼーションマシンでは、次のオプションを指定できます。

全般

自動調整

自動調整アルゴリズムを制御する制約を指定できます。制約は、アルゴリズムの実行時間、アルゴリズムの実行回数、指定できるモデル評価の数を決定します。

自動調整アルゴリズムは、最適なモデルを生成する**要因数**、**最大反復回数**、**学習ステップ**の値を選択します。詳細は、“[自動調整](#)” (6 ページ) を参照してください。

要因数

フィッティングプロセス中に各予測子レベルに使用する因子の数を指定します。1 から 100 の範囲の整数値を指定します。

非負分解

バイアス項を持つ負でない因子を強制的にゼロに設定します。これにより、学習データが非常に疎な場合にパフォーマンスが向上する可能性があります。

最大反復回数

モデルの学習が停止する前に使用される最適化の最大反復回数を指定します。

学習ステップ

モデル推定を実施するための損失関数の最小変化を指定します。

学習ステップ値は、応答範囲上限、要因数、および**学習ステップ**値の積が 5 を超える場合に調整されます。応答範囲上限の値は、応答の最大値から応答の平均値を引いた値です。**学習ステップ**の値が調整されると、その値は応答範囲上限、要因数、および 5 の積の逆になります。

評価

ビンの数

評価で使用するビンの数を指定します。5 から 100 の範囲の整数値を指定します。

統計パーセント点

観測平均の計算に使用される百分位のビンの深さを指定します。

許容値

最適化を実施するための最適化関数値の最小変化を指定します。

ファクタライゼーションマシンモデルの表示オプション

ファクタライゼーションマシンでは、次の表示プロパティを指定できます。

全般

表示されたビジュアル

表示されるオブジェクトの要素を選択します。

グラフのレイアウト

オブジェクト内のサブプロットをキャンバスに表示する方法を指定します。**自動調整**は、キャンバス上のすべてのサブプロットを自動的に揃えます。**積み上げ**は、一度に1つのプロットのみが表示されるスライドデッキのように、サブプロットを表示します。**積み上げ**が指定されている場合、サブプロット間の移動にスクロールバーではなくコントロールバーが用意されています。

表示する統計量

モデルに表示する評価統計量を指定します。次の統計量が使用できます。

- ASE
- 観測平均
- SSE

使用可能な当てはめの統計量の詳細情報については[当てはめの統計量](#)を参照してください。

数値を省略表記

大きなメジャー値が省略表記の数値を使用することを指定します。たとえば、**1,100,000,000** は **1.1B** のように表示されます。**スケール**と**精度の桁**を指定できます。

ヘッダー

表示するフィールド

表示されるヘッダーの要素を指定します。

オブザベーションの形式

モデルで使用されたオブザベーションに関する情報の表示方法を指定します。2つの異なる値が指定された精度の桁で表示される場合、明確にするために追加の桁が追加されます。

おすすめプロット

予測子

各カテゴリ変数について、ファクタライゼーションマシンは、最も推奨するまたは最も推奨されないカテゴリを示します。**非負数の因子分解**が有効になっている場合、推奨値はすべての因子の合計でソートされます。

ランク付けの方法

ファクタライゼーションマシンが最も推奨されるカテゴリを表示するか、最も推奨されないカテゴリを表示するかを指定します。

カウント

推奨カテゴリの表示数を指定します。この値は、詳細テーブルの**因子タブ**のカテゴリごとの最大行数も決定します。

スコアリングされた応答 / 相対重要度

表示するプロット

どの評価プロットを表示するかを指定します。

ヒストグラムの使用

スコアリングされた応答プロットがヒストグラムかどうか指定します。

凡例の表示

スコアリングされた応答プロットと相対重要度プロットのどちらに凡例を表示するかを指定します。

評価プロット

凡例の表示

評価プロットに凡例を表示するかどうかを指定します。

ファクタライゼーションマシンの結果

反復プロット

反復プロットは、モデル作成プロセスの各手順での目的関数値の変化を表示します。プロットの縦線は、各パフォーマンス反復の最初の内部反復を表します。目的関数値は各垂直線で増加する可能性があります、パフォーマンスの反復内では常に減少するはずです。

スコアリングされた応答

スコアリングされた応答プロットは計算された応答を真の観測値に対してプロットします。この情報はヒストグラムとして表示できます。

相対重要度

相対重要度プロットは、各入力変数の重要度値をプロットします。変数は、採点された学習データに適用されるとき、最初の分割対数値値を使用してランク付けされます。このプロットは、変数が重要でないと判断される場合は空の場合があります。

評価プロット

ファクタライゼーションマシンの場合、評価プロットは、ビンニングされたデータに対する平均予測応答値および平均観測応答値をプロットします。このプロットを使用して、モデルに強いバイアスがあることが明らかになります。平均予測値および平均観測値の大きな差はバイアスを示す可能性があります。

おすすめ

最上位のおすすめプロットと最下位のおすすめプロットは、カテゴリ変数の上位または下位のイベントレベルを表示します。

詳細テーブル

オブジェクトツールバーのをクリックすると、詳細テーブルがキャンバスの下部に表示されます。詳細テーブルには、次の情報が含まれています。

モデル情報

モデルの概要を示します。

反復履歴

損失関数の反復結果を表示します。このタブには、各反復時の損失関数の値が表示されます。

メジャー

入力メジャー変数の要約統計情報を表示します。

カテゴリ

入力カテゴリ変数に関する情報を表示します。

最終確定損失

モデルのエラー統計情報を表示します。

因子

入力データに表示されるオブザベーションに対するバイアスと係数の重みを表示します。**カウント**フィールドの値によって、カテゴリごとの最大行数が決まります。

評価

評価プロットの生成に使用されるビンニングされた評価結果を一覧表示します。

評価統計量

モデルに対して計算された評価統計量の値を表示します。

評価要約

モデルに対して計算された評価統計量の説明を表示します。応答の種類とパーティション変数の割り当ての有無によって、表示される出力が決まります。

フォレストについて

フォレストの概要

フォレストは、指定された数のディジションツリーを含むアンサンブルモデルです。フォレストによるデータのオーバーフィットが発生しないようにするために、2つの重要な手順を実施します。まず、フォレスト内の各ツリーは、学習データの異なるサンプル上に構築されます。次に、各ノードを分割するとき、分割の候補入力のセットがランダムに選択され、最適な分割がそこから選択されます。これらの2つの手順以外に、フォレストのツリーは標準ツリーのように学習されます。

フォレスト内の各ツリーの学習データには、使用可能なデータの一部が除外されています。このデータはバグ外サンプルと呼ばれます。ツリー内の各リーフについて、間隔ターゲットの予測はターゲット値の平均です。名義ターゲットのターゲットカテゴリに対する事後確率は、リーフ内のオブザベーション中のターゲットカテゴリの割合です。これらの値は、いくつかのバグ外サンプルの学習に基づく場合があります。新しいオブザベーションのターゲット値を予測するには、そのオブザベーションが各ツリーに供給され、予測値は**投票**オプションによって決まります。

デフォルトでは、フォレストを実装する SAS 手順では、学習サンプルが使用されます。

フォレストの作成

- 1 左ペインで、をクリックしてオブジェクトを選択します。フォレストを作成するには、をキャンバスにドラッグします。
- 2 右ペインのをクリックします。単一の変数を**応答変数**として指定します。
- 3 **予測子**にメジャー変数またはカテゴリ変数を 1 つ以上指定します。
- 4 (オプション)**パーティション ID**、**度数**、または**重み**を指定します。

注: パーティション ID を設定するには、パーティション変数が必要です。新しいパーティション変数の作成については、“[Create a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。既存のデータアイテムにパーティション変数を割り当てる方法については、“[Assign a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。

フォレストのオプション

フォレストでは、次のオプションを指定できます。

全般

イベントレベル

関心のあるイベントレベルを選択できます。カテゴリ応答変数に 2 つ以上のレベルが含まれている場合、SAS Viya 機械学習は評価統計量を計算する際に関心のあるレベルのすべてのオブザベーションをイベントとして、他のすべてのオブザベーションを非イベントとして扱います。

自動調整

自動調整アルゴリズムを制御する制約を指定できます。制約は、アルゴリズムの実行時間、アルゴリズムの実行回数、指定できるモデル評価の数を決定します。

自動調整アルゴリズムは、最適なモデルを生成する次のパラメーターの値を選択します。**ツリー数**、**ブートストラップ**、**ノードを分岐する予測子の数**、**最大レベル**、**リーフのサイズ**、**予測子ビン**。詳細は、“[自動調整](#)” (6 ページ)を参照してください。

ツリー数

フォレストのツリー数を指定します。1 から 10,000 の範囲の整数値を指定します。

ブートストラップ

ブートストラップ値を指定します。これは、フォレスト内の各ツリーの成長に使用されるデータの割合です。

投票

分類応答を指定すると、各オブザベーションの予測値を決定するのに使用される方法を指定します。**過半数投票**の場合、最も頻繁に予測されるイベントレベルが選択されます。**確率投票**の場合、各ツリーによって予測された確率の平均が計算され、**カットオフ値**と比較されます。

メジャー応答では、すべてのツリーの平均値が常に予測値として使用されます。

分割基準

枝を作成するために使用される分割基準を指定します。カテゴリ応答の場合、使用可能な基準は、**エントロピー**、**情報のゲインレシオ**、**カイ 2 乗**、**ジニー指数**、および **CHAID** です。メジャー応答の場合、使用可能な基準は **CHAID**、**FTest**、および **RSS/分散** です。

ノードを分岐する予測子の固定番号を設定

各ディシジョンツリーでノードを分割するのにデフォルトの予測子の数を使用するかどうかを指定します。

ノードを分岐する予測子の数

各ディシジョンツリーで使用される予測子の数を指定します。

ディシジョンツリー

欠損値の割り当て

欠損値のあるオブザベーションがどのようにモデルに含まれるかを指定します:

- **なし**: 欠損値のあるオブザベーションはモデルから除外されます。
- **検索に使用**: 欠損値のあるオブザベーションの数が**最小値**以上である場合、欠損値は固有の測定レベルと見なされ、モデルに含まれます。
- **マシン最小値として**: 欠損間隔値は、オブザベーションが常に低い変数値で分岐されるように、可能な限り最小のマシン値に設定されます。欠損しているカテゴリ値は、一意の測定レベルとして扱われます。

最小値

欠損値が個別のカテゴリレベルとして扱われる前に、欠損値を持つことができるオブザベーションの最小数を指定します。このオプションは、**欠損値の割り当て**が**検索に使用**に設定されている場合にのみ使用されます。

最大ブランチ数

ノードを分岐するときに許可されるブランチの最大数を指定します。

最大レベル数

ディシジョンツリーの最大深さを指定します。

リーフのサイズ

リーフノードで許可されるオブザベーションの最小数を指定します。

予測子ビン

メジャー変数である予測子の分類に使用されるビンの数を指定します。

ビンメソッド

メジャー予測子のビン化に使用される方法を指定します。**バケット**を選択して、最大値と最小値の差に基づいてメジャー予測子を等間隔に分割します。**分位点**を選択して、メジャー予測子をおおよそ同じサイズのグループに分割します。デフォルト値は**分位点**です。

部分従属

オブザベーションの数

入力データからサンプリングするオブザベーションの最大数を指定します。これらのオブザベーションは、部分従属プロットに表示されるモデル予測を生成するために使用されます。

ビンの数

部分従属プロットでメジャー変数を分類するために使用するビンの数を指定します。

評価

ビンの数

評価で使用するビンの数を指定します。5 から 100 の範囲の整数値を指定します。

カットオフ

計算された確率がイベントと見なされる値を指定します。計算されたカットオフ値を使用するには、**最適カットオフ**を選択します。独自のカットオフ値を設定するには、**固定カットオフ**を選択します。

統計パーセント点

観測平均、リフト、累積リフト、累積 % 捕捉、累積 % イベント、およびゲインの計算に使用される百分位のビンの深さを指定します。

許容値

百分位を推定する反復アルゴリズムの収束を決定するのに使用される許容値を指定します。アルゴリズムの精度を高めるには、小さい値を指定します。

フォレストモデル表示オプション

フォレストでは、次の表示オプションを指定できます。

全般**表示されたビジュアル**

表示されるオブジェクトの要素を選択します。

グラフのレイアウト

オブジェクト内のサブプロットをキャンバスに表示する方法を指定します。**自動調整**は、キャンバス上のすべてのサブプロットを自動的に揃えます。**積み上げ**は、一度に1つのプロットのみが表示されるスライドデッキのように、サブプロットを表示します。**積み上げ**が指定されている場合、サブプロット間の移動にスクロールバーではなくコントロールバーが用意されています。

表示する統計量

モデルに表示する評価統計量を指定します。パーティション変数を使用している場合、パーティションごとに次の当てはめの統計量(SSEを除く)を使用できます。オブジェクトツールバーには、使用可能な各パーティションタイプ(学習、検証、およびテスト)のサブメニューが含まれています。

- ASE
- 観測平均
- SSE
- C 統計量
- 累積 % 捕捉
- 累積 % イベント
- 累積リフト
- F1 スコア
- FDR
- FPR
- ゲイン
- ガンマ分布

- ジニ
- KS (Youden)
- リフト
- 誤分類率
- 誤分類率 (イベント)
- Tau

使用可能な当てはめの統計量の詳細情報については[当てはめの統計量](#)を参照してください。

注: ASE、観測平均、および SSE は、メジャー応答でのみ使用できます。残りの当てはめ統計量は、カテゴリ応答でのみ使用できます。

数値を省略表記

大きなメジャー値が省略表記の数値を使用することを指定します。たとえば、**1,100,000,000** は **1.1B** のように表示されます。**スケール**と**精度の桁**を指定できます。

ヘッダー

表示するフィールド

表示されるヘッダーの要素を指定します。

オブザベーションの形式

モデルで使用されたオブザベーションに関する情報の表示方法を指定します。2つの異なる値が指定された精度の桁で表示される場合、明確にするために追加の桁が追加されます。

誤差プロット / 部分従属プロット

表示するプロット

誤差プロットまたは部分従属プロットを表示するかどうかを指定します。

X 軸

部分従属プロットに表示される変数を指定します。

凡例の表示

誤差プロットに凡例を表示するかどうかを指定します。

評価プロット

テストパーティションの表示

テストパーティションの評価プロットを表示するかどうかを指定します。このオプションは、テストパーティションでパーティション ID を指定する場合にのみ使用できます。

表示するプロット

どの評価プロットを表示するかを指定します。**混同行列**、**リフト**、**ROC**、**カットオフプロット**、または**誤分類**を選択します。

カットオフプロット統計量

表示される評価統計量を指定します。**カットオフ**オプションの値は、常に垂直線として表示されます。次の評価統計量が使用可能です。

- 精度
- F0.5 スコア
- F1 スコア
- 偽発見率
- 偽陰性率

- 偽陽性率
- Kolmogorov-Smirnov (KS)
- 感度
- 特異度

Y 軸

標準リフトプロットまたは累積リフトプロットを表示するかどうかを指定します。

凡例の表示

評価プロットに凡例を表示するかどうかを指定します。

フォレストの結果

変数の重要度

変数の重要度プロットには、損失削減変数の重要度値によって測定される各変数の重要度が表示されます。

誤差プロット

誤差プロットは、カテゴリ応答に対するモデルと Out-Of-Bag 誤分類率を比較します。Out-Of-Bag 誤分類率は検証測定値として使用して、モデルが入力データをオーバーフィットしないようにできます。誤差プロットは、モデル全体のツリー数の誤分類率を表示します。評価プロットには、単一のターゲットレベルの評価結果が表示されます。メジャー応答の場合、**誤差プロット**は平均平方誤差を比較します。

部分従属プロット

部分従属(PD)プロットは、モデルの予測が特定の予測変数の値にどのように依存するかを理解するのに役立ちます。このプロットは、特定の予測変数の値が変更したときに応答の値がどのように変更するかを判断するのに役立ちます。Y 軸では、カテゴリ応答モデルはイベントの予測確率を表示し、メジャー応答モデルは予測応答値を表示します。どちらの場合も、値はそのビンのオブザベーションの平均です。多項モデルの場合、イベントレベルを変更して、他の関係を調べることができます。X 軸には、指定された予測変数の値が表示されます。デフォルトで表示される予測変数は、重要度が最大の予測変数です。プロットには、点推定と 95%信頼区間を示す線が表示されます。プロットの下部にあるヒートマップは、サンプリングされたオブザベーションの密度を示しています。一部のビンのオブザベーションがゼロになる場合があります。部分従属プロットは、パーティション変数を使用している場合でも、すべてのオブザベーションのサンプルを使用して生成されます。

評価プロット

混同行列には、カテゴリの応答モデルの分類結果が表示されます。モデルの作成後、各オブザベーションには観測値と予測値があります。各観測/予測ペアの総数が計算されます。混同行列は、各ペアに含まれるオブザベーションの数を表示します。完全なモデルは常に観測値を予測し、すべての値は行

列の対角線上にあります。非対角値は誤分類を表します。セルは、各レベルの観測値の数に対する各セルの値の割合に基づいて陰影が付けられます。暗い影付きのセルは、観測水準の予測の集中度を示しています。バイナリ応答の場合、混同行列の情報は誤分類プロットと同じです。

リフトは、各百分位のビン内の捕捉応答のパーセントとモデルの応答の平均パーセントの比率です。同様に、累積リフトは、現在の百分位のビンまでのすべてのデータを使用して計算されます。

受信者操作特性(ROC)チャートには、偽陽性分類および偽陰性分類を回避するモデルの能力が表示されます。偽陽性分類とは、オブザベーションが実際には非イベントである場合(タイプ I エラーとも呼ばれる)、イベントとして特定されたことを意味します。偽陰性分類とは、オブザベーションが実際にイベントである場合(タイプ II エラーとも呼ばれる)、非イベントとして特定されたことを意味します。

モデルの**特異度**は真陰性率です。偽陽性率を派生させるには、特異度を 1 から引きます。**1 - 特異度**とラベル付けされた偽陽性率は、ROC チャートの X 軸です。モデルの**感度**は真陽性率です。これは ROC チャートの Y 軸です。したがって、ROC チャートは、偽陽性率が変化すると真陽性率がどのように変化するかをプロットしています。

優れた ROC チャートには、非常に急峻な初期スロープがあり、すぐに横ばいになります。つまり、オブザベーションの誤分類ごとに、著しく多くのオブザベーションが正しく分類されています。偽陽性も偽陰性もない完璧なモデルでは、ROC チャートは(0,0)から始まり、垂直に(0,1)に上がり、水平に(1,1)になります。この例では、誤分類が 1 回発生する前に、モデルはあらゆるオブザベーションを正しく分類します。

ROC チャートには、その解釈に役立つ 2 つの線があります。最初の線は、傾きが 1 のベースラインモデルです。この線は、誤って分類するのと同率でオブザベーションを正しく分類するモデルを模倣しています。理想的な ROC チャートは、ベースラインモデルと ROC チャート間の距離が最大になっています。オブザベーションを正しくではなく誤って分類するモデルは、ベースラインモデルよりも低くなります。2 番目の線は、ROC チャートとベースラインモデルの Kolmogorov-Smirnov 値の差が最大になっている偽陽性率の垂直線です。

カットオフプロットでは、カットオフ値がモデルのパフォーマンスにどのように影響するかを視覚化できます。カットオフ値は**カットオフ**オプションで指定され、プロットでは垂直線で表されます。カットオフ値は、オブザベーションをイベントとしてモデル化するか、非イベントとしてモデル化するかを決定します。垂直線をドラッグしてカットオフ値を調整することで、モデルの再評価が行われます。

誤分類プロットには、応答変数の各値に対していくつかのオブザベーションが正しくまた誤って分類されたかが表示されます。応答変数がバイナリでない場合、フォレストは、イベントではないすべてのレベルを等しいと見なします。著しい数の誤分類は、モデルがデータに適合していないことを示している可能性があります。

メジャー応答の場合、評価プロットは、ビニングされたデータに対する平均予測応答値および平均観測応答値をプロットします。このプロットを使用して、モデルに強いバイアスがあることが明らかになります。平均予測値および平均観測値の大きな差はバイアスを示す可能性があります。

詳細テーブル

オブジェクトツールバーの[↗]をクリックすると、詳細テーブルがキャンバスの下部に表示されます。詳細テーブルには、次の情報が含まれています。

変数の重要度

モデルの各変数の重要度情報および標準偏差を表示します。

エラー基準

フォレスト内のツリーの数でソートされたすべてのフォレストモデルのエラー率を計算します。

混同行列

混同行列の生成に使用されるモデルの正しい分類と誤った分類の要約を提供します。

リフト

リフトプロットの生成に使用されるビンングされた評価結果を一覧表示します。

ROC

ROC プロットの生成に使用される結果を一覧表示します。

カットオフ統計

カットオフ値の変更に伴って変化するいくつかの統計の値を一覧表示します。

誤分類

モデルの正しい分類と誤った分類の要約を表示します。

評価

評価プロットの生成に使用されるビンングされた評価結果を表示します。

評価統計量

モデルに対して計算された評価統計量の値を表示します。

評価要約

モデルに対して計算された評価統計量の説明を表示します。応答の種類とパーティション変数の割り当ての有無によって、表示される出力が決まります。

注: 評価詳細テーブルは、メジャー応答でのみ使用できます。リフト、ROC、および誤分類の詳細テーブルは、カテゴリ応答でのみ使用できます。

勾配ブースティングモデルについて

勾配ブースティングモデルの概要

勾配ブースティングは複数のツリーを作成する反復アプローチで、各ツリーは通常、データを置き換えることなく独立したサンプルに基づいています。勾配ブースティングモデルは、平均平方誤差などの指定の損失関数を最小化することによって予測に磨きをかけます。第1手順ではベースラインツリーが作成されます。その後続くツリーはそれぞれ、前のツリーの残差に自動調整し、ロス関数は最小化されます。このプロセスは特定の回数繰り返されます。最終的なモデルは単一の関数です。この関数は、新しいオブザベーションのターゲット値を予測するのに使用できる一連のツリーの集合です。

「確率的勾配ブースティング」という言葉は、データのサブサンプルに基づいて新しいツリーをそれぞれ学習させることを指します。これにより通常、より良いモデルが得られます。勾配ブースティングモデルでは、各新規オブザベーションは、各新規オブザベーションのターゲット値を予測するために作成された一連のツリーを通して供給されます。

勾配ブースティングモデルの作成

- 1 左ペインで、をクリックしてオブジェクトを選択します。をキャンバスにドラッグして勾配ブースティングモデルを作成します。
- 2 右ペインのをクリックします。単一のメジャー変数または単一のカテゴリ変数を**応答変数**として指定します。
- 3 **予測子**にメジャー変数またはカテゴリ変数を 1 つ以上指定します。
- 4 (オプション)**パーティション ID**、**度数**、または**重み**を指定します。

注: パーティション ID を設定するには、パーティション変数が必要です。新しいパーティション変数の作成については、“[Create a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。既存のデータアイテムにパーティション変数を割り当てる方法については、“[Assign a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。

勾配ブースティングオプション

勾配ブースティングモデルでは、次のオプションを指定できます。

全般

イベントレベル

関心のあるイベントレベルを選択できます。カテゴリ応答変数に 2 つ以上のレベルが含まれている場合、SAS Viya 機械学習は評価統計量を計算する際に関心のあるレベルのすべてのオブザベーションをイベントとして、他のすべてのオブザベーションを非イベントとして扱います。

自動調整

自動調整アルゴリズムを制御する制約を指定できます。制約は、アルゴリズムの実行時間、アルゴリズムの実行回数、指定できるモデル評価の数を決定します。

自動調整アルゴリズムは、最適なモデルを生成する次のパラメーターの値を選択します。**学習率、サブサンプル率、Lasso、リッジ、ノードを分岐する予測子の数、最大レベル、リーフのサイズ、予測子ビン、ビン法**。各候補モデルは、パフォーマンス向上に自動停止オプションを使用しています。その結果、停止条件が満たされると、**ツリー数**もチャンピオンモデルのツリー数に更新されます。詳細は、“[自動調整](#)” (6 ページ)を参照してください。

自動停止メソッド

モデル構築アルゴリズムの早期終了を制御するメソッドを指定します。**スタグネーション**の場合、指定された数の連続した検証エラー計算に改善がない場合、アルゴリズムは終了します。**許容値**の場合、検証エラーの相対的変化の大きさが指定された計算回数の指定値よりも小さい場合、アルゴリズムは終了します。このオプションは、パーティション ID を指定する場合にのみ使用できます。

自動停止メソッドオプションをなしに設定して自動調整を開始した場合、オプション値は**スタグネーション**に変更され、**自動停止反復値**は**4**になります。そうでない場合、自動調整では値は変わりません。

自動停止反復

スタグネーションまたは**許容値自動停止メソッド**をトリガーするのに必要な連続した検証エラー計算回数を指定します。1 から 10,000 の範囲の整数値を指定します。このオプションは、パーティション ID を指定する場合にのみ使用できます。

許容値

許容値自動停止メソッドをトリガーするのに必要な検証エラーで必要な最小変化を指定します。0 から 1 の範囲の値を指定します。このオプションは、パーティション ID を指定する場合にのみ使用できます。

ツリー数

勾配ブースティングモデル内のツリー数を指定します。

学習率

勾配ブースティングモデルの更新に使用される学習率を指定します。

サブサンプル率

勾配ブースティングモデルで各ツリーを作成するためのサブサンプル率を指定します。

Lasso

回帰係数を選択するのに使用される LASSO(最小絶対収縮および選択演算子)パラメーターを指定します。

リッジ

勾配ブースティングモデルの更新に使用されるリッジ率を指定します。

ノードを分岐する予測子の固定番号を設定

各ディビジョンツリーでノードを分割するのにデフォルトの予測子の数を使用するかどうかを指定します。

ノードを分岐する予測子の数

各ディビジョンツリーで使用される予測子の数を指定します。

ディビジョンツリー

欠損値の割り当て

欠損値のあるオブザベーションがどのようにモデルに含まれるかを指定します:

- **なし**: 欠損値のあるオブザベーションはモデルから除外されます。
- **検索に使用**: 欠損値のあるオブザベーションの数が**最小値**以上である場合、欠損値は固有の測定レベルと見なされ、モデルに含まれます。
- **マシン最小値として**: 欠損間隔値は、オブザベーションが常に低い変数値で分岐されるように、可能な限り最小のマシン値に設定されます。欠損しているカテゴリ値は、一意の測定レベルとして扱われます。

最小値

欠損値が個別のカテゴリレベルとして扱われる前に、欠損値を持つことができるオブザベーションの最小数を指定します。このオプションは、**欠損値の割り当て**が**検索に使用**に設定されている場合にのみ使用されます。

最大ブランチ数

ノードを分岐するときに許可されるブランチの最大数を指定します。

最大レベル数

ディビジョンツリーの最大深さを指定します。

リーフのサイズ

リーフノードで許可されるオブザベーションの最小数を指定します。

予測子ビン

メジャー変数である予測子の分類に使用されるビンの数を指定します。

ビンメソッド

メジャー予測子のビン化に使用される方法を指定します。**バケット**を選択して、最大値と最小値の差に基づいてメジャー予測子を等間隔に分割します。**分位点**を選択して、メジャー予測子をおおよそ同じサイズのグループに分割します。デフォルト値は**分位点**です。

部分従属**オブザベーションの数**

入力データからサンプリングするオブザベーションの最大数を指定します。これらのオブザベーションは、部分従属プロットに表示されるモデル予測を生成するために使用されます。

ビンの数

部分従属プロットでメジャー変数を分類するために使用するビンの数を指定します。

評価**ビンの数**

評価で使用するビンの数を指定します。5 から 100 の範囲の整数値を指定します。

カットオフ

計算された確率がイベントと見なされる値を指定します。計算されたカットオフ値を使用するには、**最適カットオフ**を選択します。独自のカットオフ値を設定するには、**固定カットオフ**を選択します。

統計パーセント点

観測平均、リフト、累積リフト、累積 % 捕捉、累積 % イベント、およびゲインの計算に使用される百分位のビンの深さを指定します。

許容値

百分位を推定する反復アルゴリズムの収束を決定するのに使用される許容値を指定します。アルゴリズムの精度を高めるには、小さい値を指定します。

勾配ブースティングモデル表示オプション

勾配ブースティングモデルでは、次の表示オプションを指定できます。

全般**表示されたビジュアル**

表示されるオブジェクトの要素を選択します。

グラフのレイアウト

オブジェクト内のサブプロットをキャンバスに表示する方法を指定します。**自動調整**は、キャンバス上のすべてのサブプロットを自動的に揃えます。**積み上げ**は、一度に1つのプロットのみが表示されるスライドデッキのように、サブプロットを表示します。**積み上げ**が指定されている場合、サブプロット間の移動にスクロールバーではなくコントロールバーが用意されています。

表示する統計量

モデルに表示する評価統計量を指定します。パーティション変数を使用している場合、パーティションごとに次の当てはめの統計量(SSE を除く)を使用できます。オブジェクトツールバー

には、使用可能な各パーティションタイプ(学習、検証、およびテスト)のサブメニューが含まれています。

- ASE
- 観測平均
- SSE
- C 統計量
- 累積 % 捕捉
- 累積 % イベント
- 累積リフト
- F1 スコア
- FDR
- FPR
- ゲイン
- ガンマ分布
- ジニ
- KS (Youden)
- リフト
- 誤分類率
- 誤分類率 (イベント)
- Tau

使用可能な当てはめの統計量の詳細情報については[当てはめの統計量](#)を参照してください。

注: ASE、観測平均、および SSE は、メジャー応答でのみ使用できます。残りの当てはめ統計量は、カテゴリー応答でのみ使用できます。

数値を省略表記

大きなメジャー値が省略表記の数値を使用することを指定します。たとえば、**1,100,000,000** は **1.1B** のように表示されます。**スケール**と**精度の桁**を指定できます。

ヘッダー

表示するフィールド

表示されるヘッダーの要素を指定します。

オブザベーションの形式

モデルで使用されたオブザベーションに関する情報の表示方法を指定します。2 つの異なる値が指定された精度の桁で表示される場合、明確にするために追加の桁が追加されます。

反復プロット / 部分従属プロット

表示するプロット

誤差プロットまたは部分従属プロットを表示するかどうかを指定します。

X 軸

部分従属プロットに表示される変数を指定します。

凡例の表示

反復プロットに凡例を表示するかどうかを指定します。

評価プロット**テストパーティションの表示**

テストパーティションの評価プロットを表示するかどうかを指定します。このオプションは、テストパーティションでパーティション ID を指定する場合にのみ使用できます。

表示するプロット

どの評価プロットを表示するかを指定します。**混同行列**、**リフト**、**ROC**、**カットオフプロット**、または**誤分類**を選択します。

カットオフプロット統計量

表示される評価統計量を指定します。**カットオフ**オプションの値は、常に垂直線として表示されます。次の評価統計量が使用可能です。

- 精度
- F0.5 スコア
- F1 スコア
- 偽発見率
- 偽陰性率
- 偽陽性率
- Kolmogorov-Smirnov (KS)
- 感度
- 特異度

Y 軸

標準リフトプロットまたは累積リフトプロットを表示するかどうかを指定します。

凡例の表示

評価プロットに凡例を表示するかどうかを指定します。

勾配ブースティング 結果

変数の重要度

変数の重要度プロットには、誤差平方値の残差合計の変化に寄与することで測定される各変数の相対的重要度が表示されます。

反復プロット

反復プロットは、モデル全体のツリー数にわたって、カテゴリ応答の誤分類率またはメジャー応答の平均平方誤差のいずれかを表示します。

部分従属プロット

部分従属(PD)プロットは、モデルの予測が特定の予測変数の値にどのように依存するかを理解するのに役立ちます。このプロットは、特定の予測変数の値が変更したときに応答の値がどのように変更するかを判断するのに役立ちます。Y 軸では、カテゴリ応答モデルはイベントの予測確率を表示し、メジャー応答モデルは予測応答値を表示します。どちらの場合も、値はそのビンのオブザベーションの平均です。多項モデルの場合、イベントレベルを変更して、他の関係を調べることができます。X 軸には、指定された予測変数の値が表示されます。デフォルトで表示される予測変数は、重要度が最大の予測変数です。プロットには、点推定と 95%信頼区間を示す線が表示されます。プロットの下部にあるヒートマップは、サンプリングされたオブザベーションの密度を示しています。一部のビンのオブザベーションがゼロになる場合があります。部分従属プロットは、パーティション変数を使用している場合でも、すべてのオブザベーションのサンプルを使用して生成されます。

評価プロット

混同行列には、カテゴリの応答モデルの分類結果が表示されます。モデルの作成後、各オブザベーションには観測値と予測値があります。各観測/予測ペアの総数が計算されます。混同行列は、各ペアに含まれるオブザベーションの数を表示します。完全なモデルは常に観測値を予測し、すべての値は行列の対角線上にあります。非対角値は誤分類を表します。セルは、各レベルの観測値の数に対する各セルの値の割合に基づいて陰影が付けられます。暗い影付きのセルは、観測水準の予測の集中度を示しています。バイナリ応答の場合、混同行列の情報は誤分類プロットと同じです。

リフトは、各百分位のビン内の捕捉応答のパーセントとモデルの応答の平均パーセントの比率です。同様に、累積リフトは、現在の百分位のビンまでのすべてのデータを使用して計算されます。

受信者操作特性(ROC)チャートには、偽陽性分類および偽陰性分類を回避するモデルの能力が表示されます。偽陽性分類とは、オブザベーションが実際には非イベントである場合(タイプ I エラーとも呼ばれる)、イベントとして特定されたことを意味します。偽陰性分類とは、オブザベーションが実際にイベントである場合(タイプ II エラーとも呼ばれる)、非イベントとして特定されたことを意味します。

モデルの**特異度**は真陰性率です。偽陽性率を派生させるには、特異度を 1 から引きます。**1 - 特異度**とラベル付けされた偽陽性率は、ROC チャートの X 軸です。モデルの**感度**は真陽性率です。これは ROC チャートの Y 軸です。したがって、ROC チャートは、偽陽性率が変化すると真陽性率がどのように変化するかをプロットしています。

優れた ROC チャートには、非常に急峻な初期スロープがあり、すぐに横ばいになります。つまり、オブザベーションの誤分類ごとに、著しく多くのオブザベーションが正しく分類されています。偽陽性も偽陰性もない完璧なモデルでは、ROC チャートは(0,0)から始まり、垂直に(0,1)に上がり、水平に(1,1)になります。この例では、誤分類が 1 回発生する前に、モデルはあらゆるオブザベーションを正しく分類します。

ROC チャートには、その解釈に役立つ 2 つの線があります。最初の線は、傾きが 1 のベースラインモデルです。この線は、誤って分類するのと同率でオブザベーションを正しく分類するモデルを模倣しています。理想的な ROC チャートは、ベースラインモデルと ROC チャート間の距離が最大になっています。オブザベーションを正しくではなく誤って分類するモデルは、ベースラインモデルよりも低くなります。2 番目の線は、ROC チャートとベースラインモデルの Kolmogorov-Smirnov 値の差が最大になっている偽陽性率の垂直線です。

カットオフプロットでは、カットオフ値がモデルのパフォーマンスにどのように影響するかを視覚化できます。カットオフ値は**カットオフ**オブションで指定され、プロットでは垂直線で表されます。カットオフ値は、オブザベーションをイベントとしてモデル化するか、非イベントとしてモデル化する

かを決定します。垂直線をドラッグしてカットオフ値を調整することで、モデルの再評価が行われます。

誤分類プロットには、応答変数の各値に対していくつのオブザベーションが正しくまた誤って分類されたかが表示されます。応答変数がバイナリでない場合、勾配ブースティングモデルは、イベントではないすべてのレベルを等しいと見なします。著しい数の誤分類は、モデルがデータに適合していないことを示している可能性があります。

メジャー応答の場合、評価プロットは、ビニングされたデータに対する平均予測応答値および平均観測応答値をプロットします。このプロットを使用して、モデルに強いバイアスがあることが明らかになります。平均予測値および平均観測値の大きな差はバイアスを示す可能性があります。

詳細テーブル

オブジェクトツールバーのをクリックすると、詳細テーブルがキャンバスの下部に表示されます。詳細テーブルには、次の情報が含まれています。

変数の重要度

モデルの各変数の重要度情報および標準偏差を表示します。

反復履歴

誤差関数の反復結果を表示します。このタブには、各反復時の誤差関数の値が表示されます。

混同行列

混同行列の生成に使用されるモデルの正しい分類と誤った分類の要約を提供します。

リフト

リフトプロットの生成に使用されるビニングされた評価結果を一覧表示します。

ROC

ROC プロットの生成に使用される結果を一覧表示します。

カットオフ統計

カットオフ値の変更に伴って変化するいくつかの統計の値を一覧表示します。

誤分類

モデルの正しい分類と誤った分類の要約を表示します。

評価

評価プロットの生成に使用されるビニングされた評価結果を表示します。

評価統計量

モデルに対して計算された評価統計量の値を表示します。

評価要約

モデルに対して計算された評価統計量の説明を表示します。応答の種類とパーティション変数の割り当ての有無によって、表示される出力が決まります。

注: 評価詳細テーブルは、メジャー応答でのみ使用できます。リフト、ROC、および誤分類の詳細テーブルは、カテゴリ応答でのみ使用できます。

ニューラルネットワークについて

ニューラルネットワークの概要

ニューラルネットワークは、人間の脳の生物学的構造を模倣するように設計された統計的モデルです。ニューラルネットワークは、予測子(入力変数)、隠れ層、出力層、それらの間の接続で構成されています。予測子は出力層に直接接続できます。これらの接続の作成は、デフォルトの活性化関数によって決まります。隠れ層ごとに活性化関数を指定できます。

ニューラルネットワークの作成

ニューラルネットワークを作成するには、次の手順を実行します。

- 1 左ペインで、をクリックしてオブジェクトを選択します。をキャンバスにドラッグしてニューラルネットワークを作成します。
- 2 右ペインのをクリックします。単一の変数を**応答変数**として指定します。
- 3 **予測子**にメジャー変数またはカテゴリ変数を 1 つ以上指定します。
- 4 (オプション)**パーティション ID** 変数または**重み**変数を指定します。

注: パーティション ID を設定するには、パーティション変数が必要です。新しいパーティション変数の作成については、“[Create a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。既存のデータアイテムにパーティション変数を割り当てる方法については、“[Assign a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。

ニューラルネットワークオプション

ニューラルネットワークでは、次のオプションを指定できます。

全般

イベントレベル

関心のあるイベントレベルを選択できます。カテゴリ応答変数に 2 つ以上のレベルが含まれている場合、SAS Viya 機械学習は評価統計量を計算する際に関心のあるレベルのすべてのオブザベーションをイベントとして、他のすべてのオブザベーションを非イベントとして扱います。

自動調整

自動調整アルゴリズムを制御する制約を指定できます。制約は、アルゴリズムの実行時間、アルゴリズムの実行回数、指定できるモデル評価の数を決定します。

自動調整アルゴリズムは、最適なモデルを生成する次のパラメーターの値を選択します。**学習率、アニーリング率、L1、L2、隠れ層の数、隠れ層ニューロン**詳細は、“[自動調整](#)” (6 ページ)を参照してください。

自動停止メソッド

モデル構築アルゴリズムの早期終了を制御するメソッドを指定します。**スタグネーション**の場合、指定された数の連続した検証エラー計算に改善がない場合、アルゴリズムは終了します。**ゴール**の場合、検証エラーが指定された**ゴール**よりも小さい場合、アルゴリズムは終了します。このオプションは、パーティション ID を指定する場合にのみ使用できます。

自動停止メソッドオプションを**なし**に設定して自動調整を開始した場合、オプション値は**スタグネーション**に変更され、**自動停止反復値**は**4**になります。そうでない場合、自動調整では値は変わりません。

自動停止反復

スタグネーション自動停止メソッドをトリガーするのに必要な連続した検証エラー計算回数を指定します。1 から 10,000 の範囲の整数値を指定します。このオプションは、パーティション ID を指定する場合にのみ使用できます。

ゴール値

ゴール自動停止メソッドをトリガーするのに必要な許容できる最大の検証エラーを指定します。検証エラーがこの値よりも小さい場合、モデル構築アルゴリズムは終了します。このオプションは、パーティション ID を指定する場合にのみ使用できます。

欠損値を含める

欠損値がモデルに含まれているかどうかを指定します。欠損しているカテゴリ値は、自身の測定レベルとして扱われます。欠損しているメジャー変数は、その変数の平均値に転換されます。

分布

応答変数をモデル化するために使用される分布を指定します。このオプションは、応答がメジャー変数の場合にのみ使用できます。可能な値は、**正規分布**、**ガンマ分布**、および **Poisson 分布**です。

出力活性化関数

メジャー応答を使用するときに出力層を作成するのに使用される活性化関数を指定します。指数分布活性化関数は、ガンマ分布または Poisson 分布とともに使用されます。このオプションは、メジャー応答と**標準分布**を指定した場合にのみ使用できます。可能な値は、**双曲線正接**、**恒等**、および**正弦**です。

標準化

メジャー予測子の標準化に使用される方法を指定します。可能な値は次のとおりです。

- **なし**標準化は適用されません。
- **標準偏差**: 平均値が 0 で標準偏差が 1 になるように効果変数を変換します。
- **中範囲**: 実効値を範囲[0、1]に線形に変換します。

最大反復回数

モデルの学習が停止する前に使用される最適化の最大反復回数を指定します。

最大時間(秒)

最大タイムアウト値を指定します。

最適化メソッド

ニューラルネットワークを学習するのに使用される最適化メソッドを指定します。**SGD** または **LBFGS** いずれかのメソッドを指定できます。

学習率

確率的勾配降下の最適化に使用される学習率を指定します。0 より大きく 100 以下の値を指定する必要があります。このオプションは、**最適化手法**が **SGD** の場合にのみ使用できます。

アニーリング率

確率的勾配降下最適化に使用されるアニーリング率パラメーターを指定します。0 から 100 の範囲の値を指定します。このオプションは、**最適化手法**が **SGD** の場合にのみ使用できます。

L1

L1 の正則化パラメーターを指定します。

L2

L2 の正則化パラメーターを指定します。

隠れ層

隠れ層の数

モデル内の隠れ層の数を指定します。**最適化メソッド**が LBFGS の場合、最大許容値は 2 です。**最適化メソッド**が SGD の場合、最大許容値は 5 です。

入力ニューロンとターゲットニューロンとの直接接続を可能にする

ネットワーク内で各入力ニューロンが各出力ニューロンに接続されることを指定します。

ニューロン

隠れ層のニューロンの数を指定します。

活性化関数

その入力の加重和に基づいて各ニューロンの出力に使用される活性化関数を指定します。これは隠れ層ごとに変更できます。活性化関数として、**Tanh**、**Identity**、**Sine**、**Exponential**、**Logistic**、**ReLU**、または **Softplus** のいずれかの関数を指定できます。

部分従属

オブザベーションの数

入力データからサンプリングするオブザベーションの最大数を指定します。これらのオブザベーションは、部分従属プロットに表示されるモデル予測を生成するために使用されます。

ビンの数

部分従属プロットでメジャー変数を分類するために使用するビンの数を指定します。

評価

ビンの数

評価で使用するビンの数を指定します。5 から 100 の範囲の整数値を指定します。

カットオフ

計算された確率がイベントと見なされる値を指定します。計算されたカットオフ値を使用するには、**最適カットオフ**を選択します。独自のカットオフ値を設定するには、**固定カットオフ**を選択します。

統計パーセント点

観測平均、リフト、累積リフト、累積 % 捕捉、累積 % イベント、およびゲインの計算に使用される百分位のビンの深さを指定します。

許容値

百分位を推定する反復アルゴリズムの収束を決定するのに使用される許容値を指定します。アルゴリズムの精度を高めるには、小さい値を指定します。

ニューラルネットワークモデルの表示プロパティ

ニューラルネットワークでは、次の表示プロパティを指定できます。

全般

表示されたビジュアル

表示されるオブジェクトの要素を選択します。

グラフのレイアウト

オブジェクト内のサブプロットをキャンバスに表示する方法を指定します。**自動調整**は、キャンバス上のすべてのサブプロットを自動的に揃えます。**積み上げ**は、一度に1つのプロットのみが表示されるスライドデッキのように、サブプロットを表示します。**積み上げ**が指定されている場合、サブプロット間の移動にスクロールバーではなくコントロールバーが用意されています。

表示する統計量

モデルに表示する評価統計量を指定します。パーティション変数を使用している場合、パーティションごとに次の当てはめの統計量(SSEを除く)を使用できます。オブジェクトツールバーには、使用可能な各パーティションタイプ(学習、検証、およびテスト)のサブメニューが含まれています。

- ASE
- 観測平均
- SSE
- C 統計量
- 累積 % 捕捉
- 累積 % イベント
- 累積リフト
- F1 スコア
- FDR
- FPR
- ゲイン
- ガンマ分布
- ジニ
- KS (Youden)
- リフト
- 誤分類率
- 誤分類率 (イベント)
- Tau

使用可能な当てはめの統計量の詳細情報については[当てはめの統計量](#)を参照してください。

注: ASE、観測平均、および SSE は、メジャー応答でのみ使用できます。残りの当てはめ統計量は、カテゴリ応答でのみ使用できます。

数値を省略表記

大きなメジャー値が省略表記の数値を使用することを指定します。たとえば、1,100,000,000 は 1.1B のように表示されます。スケールと精度の桁を指定できます。

ヘッダー

表示するフィールド

表示されるヘッダーの要素を指定します。

オブザベーションの形式

モデルで使用されたオブザベーションに関する情報の表示方法を指定します。2つの異なる値が指定された精度の桁で表示される場合、明確にするために追加の桁が追加されます。

ネットワークダイアグラム

ニューロンのラベル

ネットワークダイアグラムのニューロンにラベルを付けるかどうかを指定します。

ニューロンのレイアウト

ネットワークダイアグラムの層内で、どのようにニューロンをプロットするかを指定します。

水平間隔

水平レイアウトが固定に設定されている場合にノード間に存在する水平スペースの量を指定します。水平レイアウトが比例に設定されている場合、このオプションは水平間隔と垂直間隔の比率を指定します。

垂直間隔

ノード間に存在する垂直方向のスペースの量を指定します。

表示するリンクのパーセント

ネットワークダイアグラムに表示される接続リンクの数を指定します。

水平レイアウト

ニューロン層間の幅が固定されているか、ネットワークダイアグラムの高さに比例しているかを指定します。

表示するニューロンの数

ネットワークダイアグラムの各層内に表示されるニューロンの最大数を指定します。

凡例の表示

ネットワークダイアグラムに凡例を表示するかどうかを指定します。

反復 / 相対重要度 / 部分従属プロット

表示するプロット

反復プロット、相対重要度プロット、検証誤差プロット、または部分従属プロットのいずれを表示するかを指定します。

X 軸

部分従属プロットに表示される変数を指定します。

凡例の表示

凡例を反復プロット、相対重要度プロット、または検証誤差プロットのいずれに表示するかを指定します。

評価プロット

テストパーティションの表示

テストパーティションの評価プロットを表示するかどうかを指定します。このオプションは、テストパーティションでパーティション ID を指定する場合にのみ使用できます。

表示するプロット

どの評価プロットを表示するかを指定します。**混同行列**、**リフト**、**ROC**、**カットオフプロット**、または**誤分類**を選択します。

カットオフプロット統計量

表示される評価統計量を指定します。**カットオフオプション**の値は、常に垂直線として表示されます。次の評価統計量が使用可能です。

- 精度
- F0.5 スコア
- F1 スコア
- 偽発見率
- 偽陰性率
- 偽陽性率
- Kolmogorov-Smirnov (KS)
- 感度
- 特異度

Y 軸

標準リフトプロットまたは累積リフトプロットを表示するかどうかを指定します。

凡例の表示

評価プロットに凡例を表示するかどうかを指定します。

ニューラルネットワークの結果

ネットワークプロット

ネットワークプロットは、ニューラルネットワークの入力ノード、非表示ノード、接続、および出力ノードを表示します。ノードは円で表され、ノード間のリンクは2つの円を結ぶ線です。円のサイズは、モデルに対するノードの絶対値を表します。色は、その絶対値が正か負かを示します。同様に、2つのノード間の線のサイズはリンクの強さを示し、色はその値が正か負かを示します。

ニューラルネットワークを変更するには、ネットワークプロットを右クリックし、次のいずれかのオプションを選択します。

- **隠れ層の追加**により、新しい隠れ層がニューラルネットワークに挿入され、モデルが再構築されます。
- **隠れ層の編集**は、変更する隠れ層を指定します。**隠れ層の編集**ウィンドウでは、**ニューロンの数**と**活性化関数**を指定します。
- **隠れ層の削除**により、隠れ層がニューラルネットワークから削除され、モデルが再構築されます。

反復プロット

反復プロットは、モデル作成プロセスの各手順での目的関数値の変化を表示します。プロットの縦線は、各パフォーマンス反復の最初の内部反復を表します。目的関数値は各垂直線で増加する可能性があります、パフォーマンスの反復内では常に減少するはずですが。

相対重要度プロット

相対重要度プロットは、各入力変数の重要度値をプロットします。変数は、採点された学習データに適用されるとき、最初の分割対数値値を使用してランク付けされます。このプロットは、変数が重要でないと判断される場合は空の場合があります。

部分従属プロット

部分従属(PD)プロットは、モデルの予測が特定の予測変数の値にどのように依存するかを理解するのに役立ちます。このプロットは、特定の予測変数の値が変更したときに応答の値がどのように変更するかを判断するのに役立ちます。Y 軸では、カテゴリ応答モデルはイベントの予測確率を表示し、メジャー応答モデルは予測応答値を表示します。どちらの場合も、値はそのビンのオブザベーションの平均です。多項モデルの場合、イベントレベルを変更して、他の関係を調べることができます。X 軸には、指定された予測変数の値が表示されます。デフォルトで表示される予測変数は、重要度が最大の予測変数です。プロットには、点推定と 95%信頼区間を示す線が表示されます。プロットの下部にあるヒートマップは、サンプリングされたオブザベーションの密度を示しています。一部のビンのオブザベーションがゼロになる場合があります。部分従属プロットは、パーティション変数を使用している場合でも、すべてのオブザベーションのサンプルを使用して生成されます。

評価プロット

混同行列には、カテゴリの応答モデルの分類結果が表示されます。モデルの作成後、各オブザベーションには観測値と予測値があります。各観測/予測ペアの総数が計算されます。混同行列は、各ペアに含まれるオブザベーションの数を表示します。完全なモデルは常に観測値を予測し、すべての値は行列の対角線上にあります。非対角値は誤分類を表します。セルは、各レベルの観測値の数に対する各セルの値の割合に基づいて陰影が付けられます。暗い影付きのセルは、観測水準の予測の集中度を示しています。バイナリ応答の場合、混同行列の情報は誤分類プロットと同じです。

リフトは、各百分位のビン内の捕捉応答のパーセントとモデルの応答の平均パーセントの比率です。同様に、累積リフトは、現在の百分位のビンまでのすべてのデータを使用して計算されます。

受信者操作特性(ROC)チャートには、偽陽性分類および偽陰性分類を回避するモデルの能力が表示されます。偽陽性分類とは、オブザベーションが実際には非イベントである場合(タイプ I エラーとも呼ばれる)、イベントとして特定されたことを意味します。偽陰性分類とは、オブザベーションが実際にイベントである場合(タイプ II エラーとも呼ばれる)、非イベントとして特定されたことを意味します。

モデルの**特異度**は真陰性率です。偽陽性率を派生させるには、特異度を 1 から引きます。**1 - 特異度**とラベル付けされた偽陽性率は、ROC チャートの X 軸です。モデルの**感度**は真陽性率です。これは ROC チャートの Y 軸です。したがって、ROC チャートは、偽陽性率が変化すると真陽性率がどのように変化するかをプロットしています。

優れた ROC チャートには、非常に急峻な初期スロープがあり、すぐに横ばいになります。つまり、オブザベーションの誤分類ごとに、著しく多くのオブザベーションが正しく分類されています。偽陽性も偽陰性もない完璧なモデルでは、ROC チャートは(0,0)から始まり、垂直に(0,1)に上がり、水平に(1,1)になります。この例では、誤分類が 1 回発生する前に、モデルはあらゆるオブザベーションを正しく分類します。

ROC チャートには、その解釈に役立つ 2 つの線があります。最初の線は、傾きが 1 のベースラインモデルです。この線は、誤って分類するのと同率でオブザベーションを正しく分類するモデルを模倣しています。理想的な ROC チャートは、ベースラインモデルと ROC チャート間の距離が最大になっています。オブザベーションを正しくではなく誤って分類するモデルは、ベースラインモデルよりも低くなります。2 番目の線は、ROC チャートとベースラインモデルの Kolmogorov-Smirnov 値の差が最大になっている偽陽性率の垂直線です。

カットオフプロットでは、カットオフ値がモデルのパフォーマンスにどのように影響するかを視覚化できます。カットオフ値は**カットオフ**オプションで指定され、プロットでは垂直線で表されます。カットオフ値は、オブザベーションをイベントとしてモデル化するか、非イベントとしてモデル化するかを決定します。垂直線をドラッグしてカットオフ値を調整することで、モデルの再評価が行われます。

誤分類プロットには、応答変数の各値に対していくつのオブザベーションが正しくまた誤って分類されたかが表示されます。応答変数がバイナリでない場合、ニューラルネットワークモデルは、イベントではないすべてのレベルを等しいとみなします。著しい数の誤分類は、モデルがデータに適合していないことを示している可能性があります。

メジャー応答の場合、評価プロットは、ビニングされたデータに対する平均予測応答値および平均観測応答値をプロットします。このプロットを使用して、モデルに強いバイアスがあることが明らかになります。平均予測値および平均観測値の大きな差はバイアスを示す可能性があります。

検証エラー

ニューラルネットワークの場合、検証誤差プロットは学習プロセスの各反復での検証エラーの変化を表示します。

詳細テーブル

オブジェクトツールバーのをクリックすると、詳細テーブルがキャンバスの下部に表示されます。詳細テーブルには、次の情報が含まれています。

モデル情報

モデルの概要を示します。

反復履歴

目的関数および損失関数の反復結果を表示します。このタブには、各反復時の目的関数および損失関数の値が表示されます。

収束

収束の理由を示します。

混同行列

混同行列の生成に使用されるモデルの正しい分類と誤った分類の要約を提供します。

リフト

リフトプロットの生成に使用されるビニングされた評価結果を一覧表示します。

ROC

ROC プロットの生成に使用される結果を一覧表示します。

カットオフ統計

カットオフ値の変更に伴って変化するいくつかの統計の値を一覧表示します。

誤分類

モデルの正しい分類と誤った分類の要約を表示します。

評価

評価プロットの生成に使用されるビニングされた評価結果を表示します。

評価統計量

モデルに対して計算された評価統計量の値を表示します。

評価要約

モデルに対して計算された評価統計量の説明を表示します。応答の種類とパーティション変数の割り当ての有無によって、表示される出力が決まります。

注: 評価詳細テーブルは、メジャー応答でのみ使用できます。リフト、ROC、および誤分類の詳細テーブルは、カテゴリ応答でのみ使用できます。

サポートベクターマシンについて

サポートベクターマシンの概要

サポートベクターマシンは、2つのクラス間の余白を最大化する1組の超平面を構築することによって分類を実行するために使用される機械学習モデルです。SAS Viya 機械学習は、境界との距離および境界のどちらの側にあるかに基づいて、各オブザベーションに連続確率出力を割り当てます。

サポートベクターマシンの作成

サポートベクターマシンを作成するには、次の手順を実行します。

- 1 左ペインで、をクリックしてオブジェクトを選択します。をキャンバスにドラッグして、サポートベクターマシンを作成します。
- 2 右側ペインでをクリックします。1つのカテゴリ変数を**応答変数**として指定します。
- 3 **予測子**にメジャー変数またはカテゴリ変数を1つ以上指定します。
- 4 (オプション) **パーティション ID**を指定します。

注: パーティション ID を設定するには、パーティション変数が必要です。新しいパーティション変数の作成については、“[Create a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。既存のデータアイテムにパーティション変数を割り当てる方法については、“[Assign a Partition Variable](#)” (*SAS Visual Analytics: Working with Statistics Objects*)を参照してください。

サポートベクターマシンのオプション

サポートベクターマシンでは、次のオプションを指定できます。

全般

イベントレベル

関心のあるイベントレベルを選択できます。カテゴリ応答変数に 2 つ以上のレベルが含まれている場合、SAS Viya 機械学習は関心のあるレベルのすべてのオブザベーションをイベントとして、他のすべてのオブザベーションを非イベントとして扱います。

自動調整

自動調整アルゴリズムを制御する制約を指定できます。制約は、アルゴリズムの実行時間、アルゴリズムの実行回数、指定できるモデル評価の数を決定します。

自動調整アルゴリズムは、最適モデルを生成するカーネル関数およびペナルティ値の値を選択します。詳細は、“[自動調整](#)” (6 ページ)を参照してください。

注: 自動調整は、線形、3 次、および 2 次カーネルでのみ実行できます。学習アルゴリズムオプションは自動的に内点に設定されます。

欠損値を含める

欠損値がモデルに含まれているかどうかを指定します。欠損しているカテゴリ値は、自身の測定レベルとして扱われます。欠損しているメジャー変数は、その変数の平均値に転換されます。

メジャーの予測子を標準化

間隔変数を拡大するかどうか指定します。

カーネル関数

入力を拡大された特徴空間にマッピングするために使用されるカーネル関数を指定します。

- **線形** $K(u, v) = u^T v$.
- **2 次** $K(u, v) = (u^T v + 1)^2$ 。ヘッシアン行列でゼロ値のエントリを避けるために 1 が追加されています。
- **3 次** $K(u, v) = (u^T v + 1)^3$ 。ヘッシアン行列でゼロ値のエントリを避けるために 1 が追加されています。
- **シグモイド** $K(u, v) = \tanh(p_1 u^T v + p_2)$.
- **動径基底関数** $K(u, v) = \exp(-(|u - v|^2 \div 2\sigma^2))$.

シグモイドパラメーター 1

シグモイド関数のスケール(係数)パラメーターを指定します。

シグモイドパラメーター 2

シグモイド関数の位置(バイアス)パラメーターを指定します。

RBF パラメーター

ガウス動径基底関数(RBF)カーネルのスケールパラメーター sigma または**自動**を指定します。**自動**値は、メジャー予測子の数に各カテゴリ予測子のカテゴリ数を組み合わせた値の平方根です。

学習アルゴリズム

モデルの学習アルゴリズムの方法を指定します。シグモイドおよび動径基底関数カーネルは、**アクティブセット**に設定する必要があります。

- **アクティブセット**
- **座標降下**
- **内点**

注: 反復履歴は、アクティブセットまたは座標降下法では使用できません。

ペナルティの種類

座標降下法のペナルティの種類を指定します。

ペナルティ値

ペナルティ値を指定します。ペナルティ値は、モデルの複雑さと学習のエラーのバランスを保ちます。ペナルティ値が大きいほど、学習データのオーバーフィットを犠牲にしてより堅牢なモデルが作成されます。

許容値

モデル学習のカスタム許容値を指定します。許容値は、サポートベクター数とモデル精度のバランスを保ちます。許容値が大きすぎるとサポートベクターの数が少なくなりすぎ、値が小さすぎると学習データのオーバーフィットが発生します。

サポートベクター最大数

アクティブセット法の最大サポートベクター数を指定します。アクティブセット手法の性質上、サポートベクターの実際数は、指定した数よりもわずかに多くなる場合があります。

最大反復回数

モデルの学習が停止する前に使用される最適化の最大反復回数を指定します。

部分従属

オブザベーションの数

入力データからサンプリングするオブザベーションの最大数を指定します。これらのオブザベーションは、部分従属プロットに表示されるモデル予測を生成するために使用されます。

ビンの数

部分従属プロットでメジャー変数を分類するために使用するビンの数を指定します。

評価

ビンの数

評価で使用するビンの数を指定します。5 から 100 の範囲の整数値を指定します。

カットオフ

計算された確率がイベントと見なされる値を指定します。計算されたカットオフ値を使用するには、**最適カットオフ**を選択します。独自のカットオフ値を設定するには、**固定カットオフ**を選択します。

統計パーセント点

観測平均、リフト、累積リフト、累積 % 捕捉、累積 % イベント、およびゲインの計算に使用される百分位のビンの深さを指定します。

許容値

百分位を推定する反復アルゴリズムの収束を決定するのに使用される許容値を指定します。アルゴリズムの精度を高めるには、小さい値を指定します。

サポートベクターマシンモデル表示オプション

サポートベクターマシンでは、次の表示オプションを指定できます。

全般

表示されたビジュアル

表示されるオブジェクトの要素を選択します。

グラフのレイアウト

オブジェクト内のサブプロットをキャンバスに表示する方法を指定します。**自動調整**は、キャンバス上のすべてのサブプロットを自動的に揃えます。**積み上げ**は、一度に1つのプロットのみが表示されるスライドデッキのように、サブプロットを表示します。**積み上げ**が指定されている場合、サブプロット間の移動にスクロールバーではなくコントロールバーが用意されています。

表示する統計量

モデルに表示する評価統計量を指定します。パーティション変数を使用している場合、パーティションごとに次の当てはめの統計量を使用できます。オブジェクトツールバーには、使用可能な各パーティションタイプ(学習、検証、およびテスト)のサブメニューが含まれています。

- C 統計量
- 累積 % 捕捉
- 累積 % イベント
- 累積リフト
- F1 スコア
- FDR
- FPR
- ゲイン
- ガンマ分布
- ジニ
- KS (Youden)
- リフト
- 誤分類率 (イベント)
- Tau

使用可能な当てはめの統計量の詳細情報については[当てはめの統計量](#)を参照してください。

数値を省略表記

大きなメジャー値が省略表記の数値を使用することを指定します。たとえば、**1,100,000,000** は **1.1B** のように表示されます。**スケール**と**精度の桁**を指定できます。

ヘッダー

表示するフィールド

表示されるヘッダーの要素を指定します。

オブザベーションの形式

モデルで使用されたオブザベーションに関する情報の表示方法を指定します。2つの異なる値が指定された精度の桁で表示される場合、明確にするために追加の桁が追加されます。

部分従属プロット

部分従属を表示

部分従属プロットを表示するかどうかを指定します。

X 軸

部分従属プロットに表示される変数を指定します。

評価プロット

テストパーティションの表示

テストパーティションの評価プロットを表示するかどうかを指定します。このオプションは、テストパーティションでパーティション ID を指定する場合にのみ使用できます。

表示するプロット

どの評価プロットを表示するかを指定します。**混同行列**、**リフト**、**ROC**、**カットオフプロット**、または**誤分類**を選択します。

カットオフプロット統計量

表示される評価統計量を指定します。**カットオフ**オプションの値は、常に垂直線として表示されます。次の評価統計量が使用可能です。

- 精度
- F0.5 スコア
- F1 スコア
- 偽発見率
- 偽陰性率
- 偽陽性率
- Kolmogorov-Smirnov (KS)
- 感度
- 特異度

Y 軸

標準リフトプロットまたは累積リフトプロットを表示するかどうかを指定します。

凡例の表示

評価プロットに凡例を表示するかどうかを指定します。

サポートベクターマシンの結果

相対重要度プロット

相対重要度プロットは、各入力変数の重要度値をプロットします。変数は、採点された学習データに適用されるとき、最初の分割対数値を使用してランク付けされます。このプロットは、変数が重要でないと判断される場合は空の場合があります。

部分従属プロット

部分従属(PD)プロットは、モデルの予測が特定の予測変数の値にどのように依存するかを理解するのに役立ちます。このプロットは、特定の予測変数の値が変更したときに応答の値がどのように変更するかを判断するのに役立ちます。Y 軸では、カテゴリ応答モデルはイベントの予測確率を表示し、メジャー応答モデルは予測応答値を表示します。どちらの場合も、値はそのビンのオブザベーションの平均です。多項モデルの場合、イベントレベルを変更して、他の関係を調べることができます。X 軸には、指定された予測変数の値が表示されます。デフォルトで表示される予測変数は、重要度が最大の予測変数です。プロットには、点推定と 95%信頼区間を示す線が表示されます。プロットの下部にあるヒートマップは、サンプリングされたオブザベーションの密度を示しています。一部のビンのオブザベーションがゼロになる場合があります。部分従属プロットは、パーティション変数を使用している場合でも、すべてのオブザベーションのサンプルを使用して生成されます。

評価プロット

混同行列には、カテゴリの応答モデルの分類結果が表示されます。モデルの作成後、各オブザベーションには観測値と予測値があります。各観測/予測ペアの総数が計算されます。混同行列は、各ペアに含まれるオブザベーションの数を表示します。完全なモデルは常に観測値を予測し、すべての値は行列の対角線上にあります。非対角値は誤分類を表します。セルは、各レベルの観測値の数に対する各セルの値の割合に基づいて陰影が付けられます。暗い影付きのセルは、観測水準の予測の集中度を示しています。バイナリ応答の場合、混同行列の情報は誤分類プロットと同じです。

リフトは、各百分位のビン内の捕捉応答のパーセントとモデルの応答の平均パーセントの比率です。同様に、累積リフトは、現在の百分位のビンまでのすべてのデータを使用して計算されます。

受信者操作特性(ROC)チャートには、偽陽性分類および偽陰性分類を回避するモデルの能力が表示されます。偽陽性分類とは、オブザベーションが実際には非イベントである場合(タイプ I エラーとも呼ばれる)、イベントとして特定されたことを意味します。偽陰性分類とは、オブザベーションが実際にイベントである場合(タイプ II エラーとも呼ばれる)、非イベントとして特定されたことを意味します。

モデルの**特異度**は真陰性率です。偽陽性率を派生させるには、特異度を 1 から引きます。**1 - 特異度**とラベル付けされた偽陽性率は、ROC チャートの X 軸です。モデルの**感度**は真陽性率です。これは ROC チャートの Y 軸です。したがって、ROC チャートは、偽陽性率が変化すると真陽性率がどのように変化するかをプロットしています。

優れた ROC チャートには、非常に急峻な初期スロープがあり、すぐに横ばいになります。つまり、オブザベーションの誤分類ごとに、著しく多くのオブザベーションが正しく分類されています。偽陽性も偽陰性もない完璧なモデルでは、ROC チャートは(0,0)から始まり、垂直に(0,1)に上がり、水平に

(1,1)になります。この例では、誤分類が 1 回発生する前に、モデルはあらゆるオブザベーションを正しく分類します。

ROC チャートには、その解釈に役立つ 2 つの線があります。最初の線は、傾きが 1 のベースラインモデルです。この線は、誤って分類するのと同率でオブザベーションを正しく分類するモデルを模倣しています。理想的な ROC チャートは、ベースラインモデルと ROC チャート間の距離が最大になっています。オブザベーションを正しくではなく誤って分類するモデルは、ベースラインモデルよりも低くなります。2 番目の線は、ROC チャートとベースラインモデルの Kolmogorov-Smirnov 値の差が最大になっている偽陽性率の垂直線です。

カットオフプロットでは、カットオフ値がモデルのパフォーマンスにどのように影響するかを視覚化できます。カットオフ値は**カットオフ**オプションで指定され、プロットでは垂直線で表されます。カットオフ値は、オブザベーションをイベントとしてモデル化するか、非イベントとしてモデル化するかを決定します。垂直線をドラッグしてカットオフ値を調整することで、モデルの再評価が行われます。

誤分類プロットには、応答変数の各値に対していくつのオブザベーションが正しくまた誤って分類されたかが表示されます。応答変数がバイナリでない場合、サポートベクターマシンは、イベントではないすべてのレベルを等しいと見なします。著しい数の誤分類は、モデルがデータに適合していないことを示している可能性があります。

詳細テーブル

オブジェクトツールバーのをクリックすると、詳細テーブルがキャンバスの下部に表示されます。詳細テーブルには、次の情報が含まれています。

モデル情報

モデルの概要を示します。

反復履歴

各反復の相補および実現可能の統計量を表示します。

学習

モデルの成分の簡単な説明を表示します。

当てはめ統計量

当てはめの統計量すべてを一覧表示します。

混同行列

混同行列の生成に使用されるモデルの正しい分類と誤った分類の要約を提供します。

リフト

リフトプロットの生成に使用されるビンニングされた評価結果を一覧表示します。

ROC

ROC プロットの生成に使用される結果を一覧表示します。

カットオフ統計

カットオフ値の変更に伴って変化するいくつかの統計の値を一覧表示します。

誤分類

モデルの正しい分類と誤った分類の要約を表示します。

評価統計量

モデルに対して計算された評価統計量の値を表示します。

評価要約

モデルに対して計算された評価統計量の説明を表示します。応答の種類とパーティション変数の割り当ての有無によって、表示される出力が決まります。