

K 近傍法

JMP PRO この近接性に基づいたアルゴリズムを使い、類似のオブザベーションの応答(近傍点)に基づいて、新しいオブザベーションの質的変数の応答を予測(分類)したり、連続変数の応答を予測したりすることができます。

K 近傍法

- データテーブルから、分析 > 予測モデル > K 近傍法を選択します。
- 列の選択から質的変数または連続変数の応答を選択し、Y, 目的変数を選択します。この例では質的変数を用います。
- 候補予測変数を選択し、X, 説明変数をクリックします。
- 必要な場合、検証データの割合を入力するか、右図のように検証列を選択して検証をクリックします。
- OK をクリックすると、JMP は以下の内容を表示します：
 - K の値の範囲内で誤分類されたオブザベーションの度数と割合を示したグラフと表。
 - 誤分類率が最小になっている K について、分類性能の詳細を示した混同行列。
 - 混同行列の値を図示したマザイク図(未掲載)。

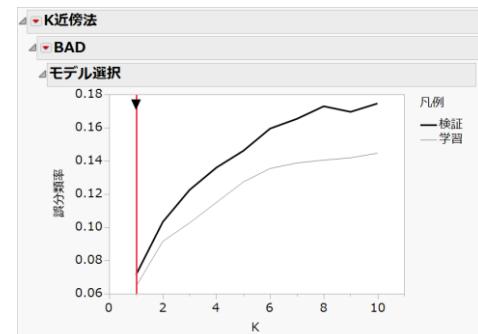
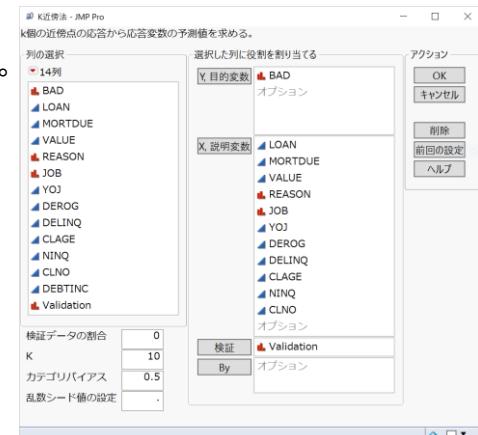
リスク水準(Bad/Good)を予測した K 近傍法の結果:

- 検証データは 1,192 例。誤分類率が最小となるのは 1 つの近傍点のみを予測に用いた場合で、 $85/1192 = 7.1\%$ が誤分類された。全体の誤分類のうち、 $0/(0+917) = 0\%$ の Good Risk のオブザベーションが Bad Risk と誤分類された。また $85/(85+190) = 31\%$ の Bad Risk のオブザベーションが Good Risk と誤分類された。
- テストデータ(未掲載)は 1,192 例。誤分類率が最小となるのは 1 つの近傍点のみを予測に用いた場合で、 $69/1192 = 5.7\%$ が誤分類された。全体の誤分類のうち、 $0/(0+949) = 0\%$ の Good Risk のオブザベーションが Bad Risk と誤分類された。また $69/(69+174) = 28\%$ の Bad Risk のオブザベーションが Good Risk と誤分類された。テストデータのオブザベーションはモデルの学習にも選択にも用いたれなかったため、この結果が将来のデータの誤分類率の最も正確な推定であるとしばしば考えられる。

注意:

近傍行の保存、近傍距離の保存等の追加のオプションを使うには一番上の赤い三角ボタンをクリックします。予測値の保存、予測式の保存、予測式を発行等の追加のオプションを使うには最上部近くの応答変数名の横の赤い三角ボタンをクリックします。K 近傍法の追加情報に関しては、JMP のヘルプか予測モデルおよび発展的なモデル(ヘルプ > JMP ドキュメンテーションライブラリ以下)で「K 近傍法」と検索してご確認ください。

例: Equity.jmp (ヘルプ > サンプルデータライブラリ)



学習				検証				
K	度数	R2乗	誤分類率	K	度数	R2乗	誤分類率	
1	3576	0.4079	0.06432	230 *	1	1192	0.44996	0.07131
2	3576	0.45873	0.09144	327	2	1192	0.47812	0.10319
3	3576	0.44273	0.10263	367	3	1192	0.45851	0.12249
4	3576	0.42115	0.11493	411	4	1192	0.42015	0.13591
5	3576	0.38727	0.12724	455	5	1192	0.40146	0.14597
6	3576	0.35814	0.13535	494	6	1192	0.37543	0.15940
7	3576	0.33587	0.13870	496	7	1192	0.35076	0.16527
8	3576	0.31932	0.14038	502	8	1192	0.32274	0.17282
9	3576	0.29532	0.14178	507	9	1192	0.31301	0.16946
10	3576	0.27695	0.14457	517	10	1192	0.29547	0.17450

最良の混同行列 K=1					
学習		検証			
実測値	予測値	度数	実測値	予測値	度数
BAD	Good Risk	Bad Risk	BAD	Good Risk	Bad Risk
Good Risk	2894	11	Good Risk	917	0
Bad Risk	219	452	Bad Risk	85	190
実測値	予測値	割合	実測値	予測値	割合
BAD	Good Risk	Bad Risk	BAD	Good Risk	Bad Risk
Good Risk	0.996	0.004	Good Risk	1.000	0.000
Bad Risk	0.326	0.674	Bad Risk	0.309	0.691