

# Support Vector Machines - Classification

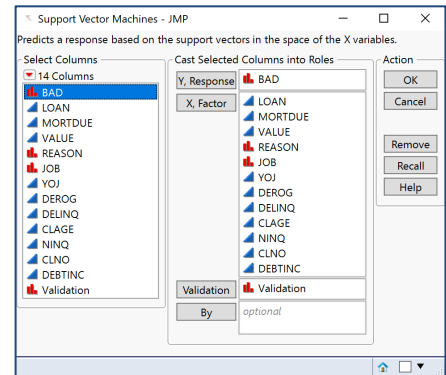


Build a boundary based statistical model to predict a categorical outcome (classify) as a function of multiple predictor variables. SVM is able to create much more flexible boundary shapes than the Classification Tree (Partition) and Discriminant Analysis method.

## Support Vector Machines

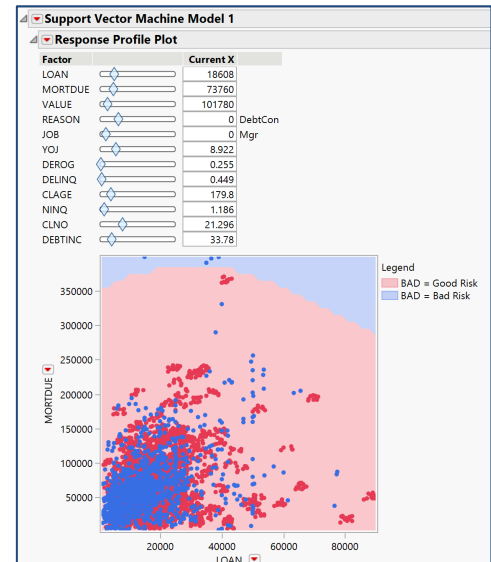
1. From an open JMP® table, select **Analyze > Predictive Modeling > Support Vector Machines**.
2. Add a nominal or ordinal response variable from **Select Columns** to the **Y, Response** role.
3. Add candidate predictor variables to the **X, Factor** role.
4. If desired, enter a validation column into the **Validation** role as shown in this example. Click **OK**.
5. The Model Launch control panel opens allowing a choice of a Kernel Function and associated options. Default settings were used for this example. Click **Go**.

Equity.jmp (Help &gt; Sample Data Library)



JMP displays:

- **Response Profile Plot** displaying the classification regions and the data values for two of the predictor variables. These can be changed to other variables by selecting the red triangle next to the variable name on the axes. Levels for all the remaining predictors can be changed with the sliders above the plot.
- **Model Summary** (not shown) and a **Confusion Matrix** detailing the classification performance.



Intepretation:

- There are 649 observations in the Vaidation Data. Of these, 45 (6.9%) where misclassified.  $45/(45+18) = 71\%$  of the Bad Risk customers were misclassified as Good Risk.  $0/(0+586) = 0\%$  of the Good Risk customers were misclassified as Bad Risk.

Note: The default rule is to classify an observation in the class with the highest estimated probability of being in that class (i.e., Prob > 0.50). It is advantageous to evaluate different cutoff values in order to minimize a specific type of misclassification rate over another. In this example it would be much better to choose a cutoff level to create a lower misclassification rate for Bad Risk Customers misclassified as Good Risk while accepting a higher misclassification rate for the other. This analysis can be performed by saving the predicted probabilities to the data table and using the calculator tool to create a conditional argument to conclude an outcome based on those predicted probability values.

Training				Validation			
Actual	Predicted Rate		Misclassification Rate	Actual	Predicted Rate		Misclassification Rate
BAD	Good Risk	Bad Risk		BAD	Good Risk	Bad Risk	
Good Risk	1.000	0.000	0.0550	Good Risk	1.000	0.000	0.0693
Bad Risk	0.621	0.379		Bad Risk	0.714	0.286	
Actual	Predicted Count			Actual	Predicted Count		
BAD	Good Risk	Bad Risk		BAD	Good Risk	Bad Risk	
Good Risk	1871	0		Good Risk	586	0	
Bad Risk	113	69		Bad Risk	45	18	

Additional options, such as **ROC and Lift Curves, Profilers, Save Predicteds, Save Prediction Formula, Save Probabilities**, as well as **Publish Probability Formulas** are accessible from the **red triangle**.

Visit **Predictive and Specialized Models > Support Vector Machines** in **JMP Help** to learn more.