

Classification Threshold Explorer

Exploring Model Thresholds

The Classification Threshold Explorer Add-In is a tool for exploring and selecting thresholds for a binary classification model.

Figure 1.1



Contents

About the Classification Threshold Explorer	3
Example of the Classification Threshold Explorer.....	3
Launch the Classification Threshold Explorer	4
Data Format	5
The Classification Threshold Explorer Report	5
ROC Curve.....	5
State vs. Score Plot.....	6
Performance Measures	7
Confusion Matrix	7
Error Tradeoff Curve	8
Overall Lift Curve.....	9
Overall Cum Gains Curve	10
Classification Threshold Explorer Options.....	10
Example of a Cost Based Threshold.....	11
References.....	12

About the Classification Threshold Explorer

This add-in was envisioned by Mia Stephens of JMP and Karen Copeland of Boulder Statistics. The idea was to develop a tool for exploring cut-offs in classification models to be used in the teaching of classification models as well as in the practice of using such models. Our goal was to tie together visual information and performance measures to make the process of threshold selection meaningful to the user and consumer of classification models.

This add-in was scripted in the summer of 2018 by JMP Intern Tarek Zikry.

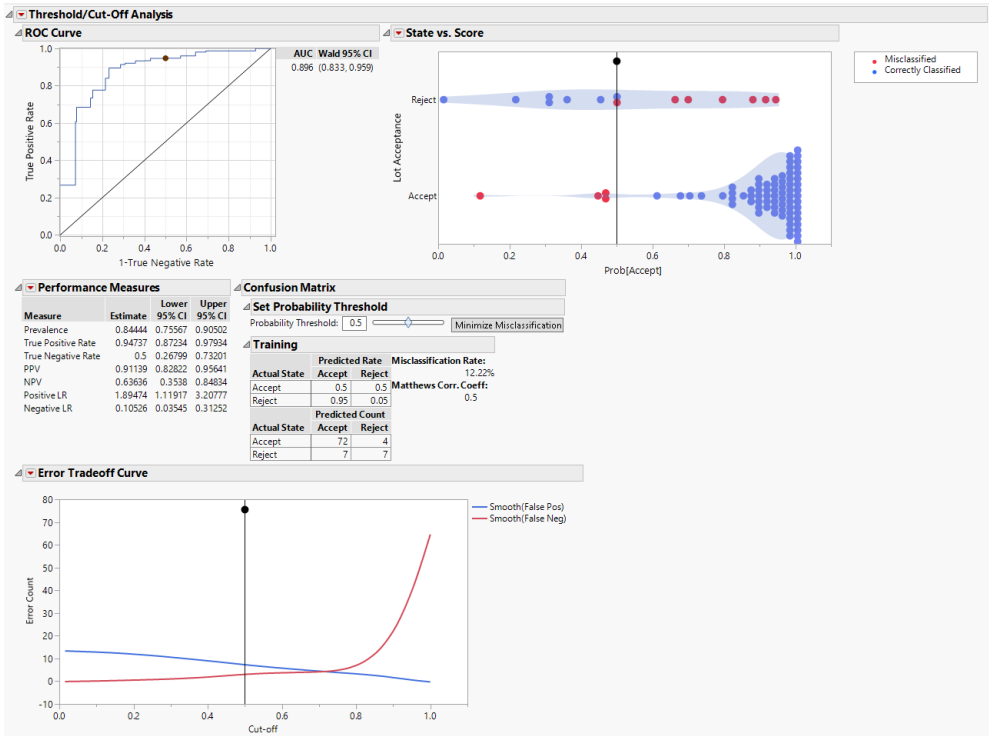
Example of the Classification Threshold Explorer

This example uses a data set from the production of tablets.

1. Select **Help > Sample Data Library** and open **Tablet Production.jmp**
2. Select **Rows > Clear Row States**.
3. Select **Analyze > Fit Model**.
4. Select **Lot Acceptance** and click **Y**.
5. Select **Inlet Temp, Exhaust Temp, and Spray Rate** and click **Add**.
6. Select **Personality > Nominal Logistic**.
7. Set **Target Level** to **Accept**, and click **Run**.
8. Click the **Nominal Logistic Fit for Lot Acceptance** red triangle menu and select **Save Probability Formula**.
9. Select **Add-Ins > Classification Threshold Explorer**.
10. Select **Lot Acceptance** and click **Y, Response**.
11. Select **Prob[Accept]** and click **X, Predicted Value**.
12. Set **Target Level** to **Accept**, and click **OK**.

Classification Threshold Explorer
Launch the Classification Threshold Explorer

Figure 1.2 Initial Classification Threshold Explorer Report



Launch the Classification Threshold Explorer

Launch the Classification Threshold Explorer by selecting Add-Ins > Classification Threshold Explorer.

Figure 1.3 Classification Threshold Explorer Launch Window

The launch window for the Classification Threshold Explorer includes the following sections:

- Description:** A text box containing "Add-In to explore model classification cut-offs".
- Choose Naming Set:** A dropdown menu set to "Performance Terminology: General".
- Select Columns:** A list of available columns including API Lot No, API Particle Size, Mill Time, Screen Size, Mag. Stearate Supplier, Lactose Supplier, Sugar Supplier, Talc Supplier, Blend Time, and Blend Speed.
- Cast Selected Columns into Roles:** Fields for Y, Response (required), X, Predicted Value (required), and Validation (optional). It also includes Set alpha level (0.05) and Target Level (Accept).
- Action:** Buttons for OK, Cancel, Remove, Recall, and Help.
- Visual Accessibility:** A checkbox at the bottom.

Performance Terminology Select a terminology set for performance measures.

Y, Response The column to be analyzed. This column must have two categories and have a nominal type.

X, Predicted Value The column of your model score. This column must be continuous.

Validation A column to define validation groups for comparisons. The column should be of nominal type and must have at least 2 and no more than 3 categories. The categories are labeled as Training, Validation, and Test in the report.

Tip: Use the Validation variable for any grouping variable of interest.

Set alpha level Enter the alpha level for confidence intervals.

Target Level Select the target level of your response.

Visual Accessibility Enables the use markers and dashed lines in reports for visual accessibility.

Data Format

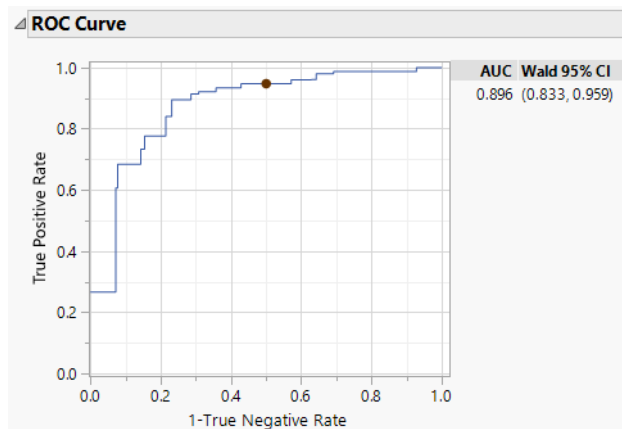
Two columns of data are required for the Classification Threshold Explorer Add-In; a column of nominal type for the responses or state and a continuous column of model scores.

The Classification Threshold Explorer Report

The Classification Threshold Explorer Report is an interactive report enabling you to evaluate different thresholds on the performance metrics of a classification model.

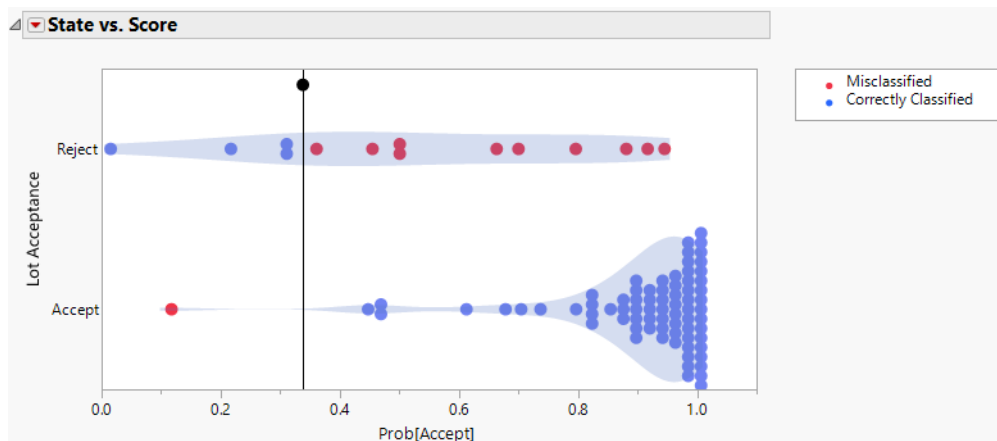
ROC Curve

The ROC Curve is a receiver operating characteristic curve or a plot of sensitivity versus 1 - specificity. The ROC curve describes the accuracy of a classification model across all possible thresholds. The dot on the plot shows the current threshold. Wald confidence intervals are included in the Area Under the Curve (AUC) report. The labels on the ROC Curve plot depend on the selected performance terminology. If a validation column is used, the ROC plot has a curve for each level of the validation column.

Figure 1.4 ROC Curve

State vs. Score Plot

The State vs. Score plot is used to visualize the predicted scores within each category. Move the slider to adjust the threshold. The data points are colored by their classification state. If a validation column is used, the State vs. Score Plot will have plots data with the state nested within each of the validation column.

Figure 1.5 State vs. Score Plot

Performance Measures

There are three sets of terminology for the performance measures. Performance measures summarize the ability of a classification model to categorize the outcomes. The performance measures are based on counts in the confusion matrix. The report includes Score confidence intervals for the first 5 measures. The confidence intervals for the likelihood ratios are based on the method of Simel, et. al. (1991).

Table 1.1 Confusion Matrix

	Predicted Positive	Predicted Negative
True Condition Positive	True Positive (TP)	False Negative (FN)
True Condition Negative	False Positive (FP)	True Negative (TN)

Table 1.2 Performance Metrics

General	Marketing/Business	Medical/Diagnostics	Calculation
Prevalence	Prevalence	Prevalence	$(TP+FN) / N$
True Positive Rate	Recall	Sensitivity	$TP / (TP+FN)$
True Negative Rate	Specificity	Specificity	$TN / (FP+TN)$
PPV	Precision	PPV	$TP / (TP+FP)$
NPV	NPV	NPV	$TN / (FN+TN)$
Positive LR	Positive LR	Positive LR	$Sensitivity / FPR$
Negative LR	Negative LR	Negative LR	$FNR / Specificity$

Note: When a validation column is used, multiple performance measure tables are provided. Each table has a red triangle menu to open individual lift or cumulative gain plots.

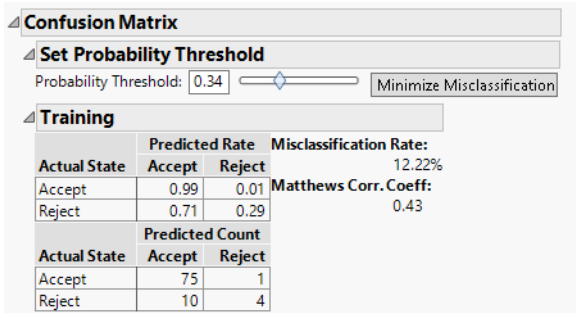
Confusion Matrix

The confusion matrix shows the 2x2 table of counts based on the current threshold. Use the Probability Threshold slider or text box to adjust the Threshold. The Misclassification Rate is the percent of misclassified rows.

Matthew’s Correlation Coefficient (MCC) is a performance metric to describe the results of the confusion matrix. The measure takes into account both correct and incorrectly classified observations and can be useful when the categories are of different sizes. The MCC falls between -1 and +1 with -1 indicating no agreement between the classification and truth and +1 perfect classification.

$$MCC = \frac{TP \times TN - PF \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Figure 1.6 Confusion Matrix



Note: When a validation column is used, multiple confusion matrices are provided.

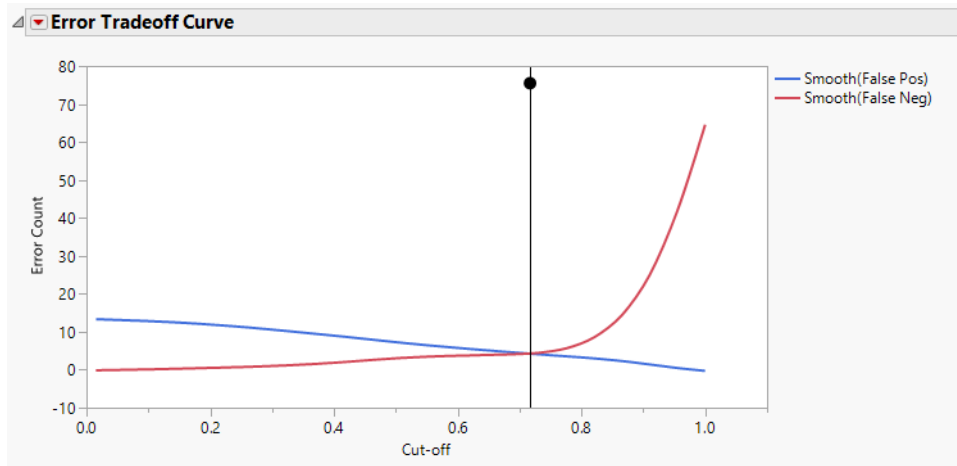
Error Tradeoff Curve

The tradeoff curve is a plot of the false positive counts and false negative counts across all thresholds. The curves cross at the threshold that minimizes overall misclassification.

Error Tradeoff Curve Options

Set Y-axis Enables changing the Y axis on from counts to rates.

Figure 1.7 Error Tradeoff Curve

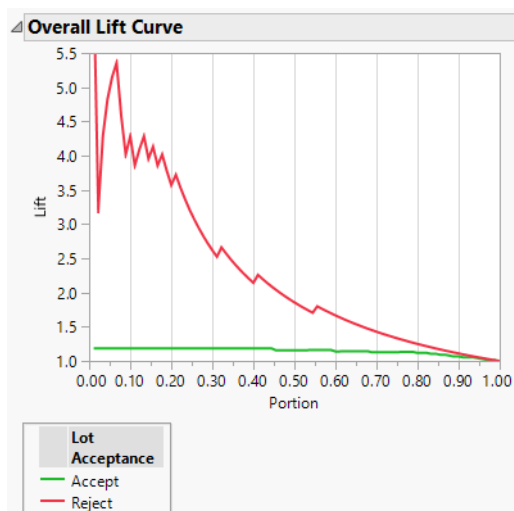


Note: When a validation column is used, each level is a stacked row in the Error Tradeoff Curve.

Overall Lift Curve

The lift curve illustrates the effectiveness of a predictive model. The higher the lift curve at a given portion, the better the model is at correctly classifying outcomes within this portion.

Figure 1.8 Lift Curve



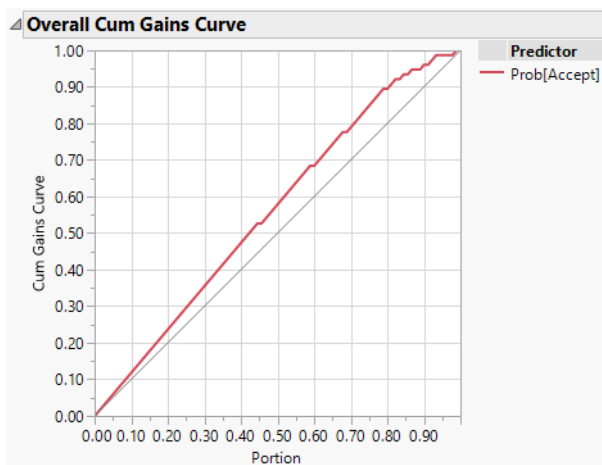
Note: When a validation column is used, access individual lift curves from the Performance Measures red triangle menus.

Overall Cum Gains Curve

Available only in JMP Pro.

The cumulative gains curve plots the proportion of the target level that is identified against the proportion of all responses. The cumulative gain curve for a perfect model would reach 1.0 at the prevalence of the target in the sample.

Figure 1.9 Cumulative Gains Curve



Note: When a validation column is used access cumulative gains curves from the Performance Measures red triangle menus.

Classification Threshold Explorer Options

Lift Curve Shows or hides a lift curve.

Cum Gains Curve Shows or hides a cumulative gain curve for your selected target.

Cost Based Threshold Launches a dialog for calculating an optimal threshold based on costs. Finds the threshold that maximizes (sensitivity - $m(1-\text{specificity})$) where

$$m = \left(\frac{FP_c - TN_c}{FN_c - TP_c} \right) \left(\frac{1-p}{p} \right).$$

For more information about this cost-based threshold method, see Zweig and Campbell (1993).

Save Classification Column Saves a column to your data table with the classification for the selected threshold. The column name includes the threshold enabling you to save additional columns for different thresholds.

Note: The red triangle menus for the State vs. Score Plot is the standard Graph Builder menu.

Performance Measures Enables the selection of terminology for the performance measures

Example of a Cost Based Threshold

When costs can be assigned to classification outcomes, an optimal threshold to minimize costs can be calculated. Consider again the Tablet Production sample data table. Suppose that accepting a bad lot is twice as costly to your company then rejecting a good lot. That is, you would rather scrap good material than to ship bad material.

Continuing from the [“About the Classification Threshold Explorer”](#) on page 3:

1. Select **Cost Based Threshold** from the Lot Acceptance Threshold/Cut-Off Analysis red triangle menu.
2. Enter 0.85 for the Prevalence. This is our expected rate of good lots.
3. Enter 2 for False Positive Cost and 1 for the False Negative Cost. That is, a false positive (accepting and then shipping a bad lot) costs 2 times as much as a false negative (scrapping a good lot).
4. Enter 0 for the True Positive and True Negative costs. We do not care to use these costs in our threshold determination.
5. Click **Calculate**.

Figure 1.10 Cost Based Threshold Calculation

Prevalence:

Enter Costs on a Common Scale,
Where: $FP_c \neq TN_c$, $FN_c \neq TP_c$
If $TN_c > FP_c$ then $TP_c > FN_c$
If $TN_c < FP_c$ then $TP_c < FN_c$
A cost can be zero

True Positive Cost: True Negative Cost:

False Positive Cost: False Negative Cost:

Optimal Threshold: 0.7

The optimal threshold for this cost configuration is at 0.7.

References

- Simel, D. L., Samsa, G. P. and Matchar, D. B. (1991). "Likelihood ratios with confidence: sample size estimation for diagnostic test studies." *Journal of Clinical Epidemiology* 44:763-770.
- Zweig, M. H., and Campbell, G. (1993). "Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine." *Clinical Chemistry* 39:561-577.
<https://www.ncbi.nlm.nih.gov/pubmed/8472349>. Retrieved July 29, 2018