PSS718 - Data Mining

Lecture 12 - Performance Evaluation

Asst.Prof.Dr. Burkay Genç

Hacettepe University

December 18, 2016

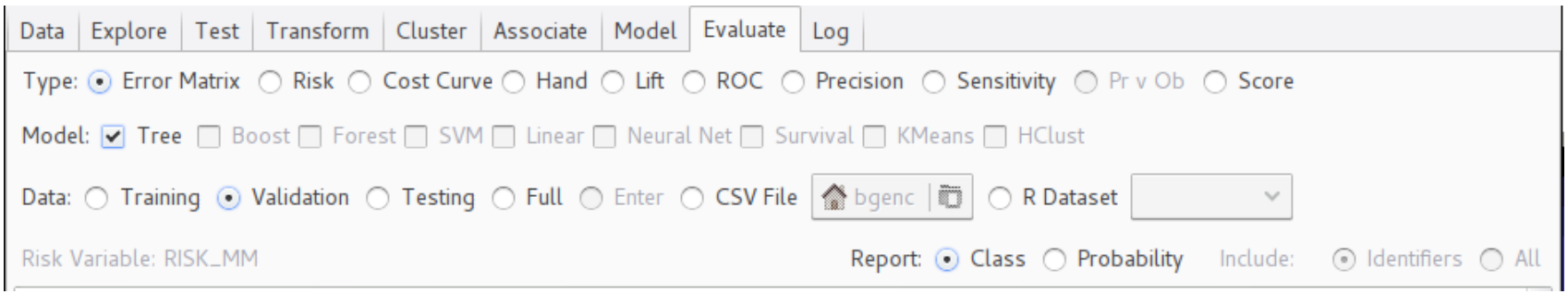
So far...

- We have described descriptive and predictive models
- To choose the best model, we must evaluate them
- This also helps identifying obvious variable decision errors
  - Trying to output RainTomorrow
  - Using RainAmountTomorrow as input

What to look at

- Confusion matrix
- Risk chart
- ROC curves
- Scoring datasets

User interface



Types of evaluations

- Error Matrix (aka. Confusion Matrix)
- Risk
- Cost Curve
- Hand
- Lift
- ROC
- Precision
- Sensitivity
- Pr v. Ob
- Score

Models to evaluate

- Only models that have been built will be available
- Last built model is automatically selected
- ...but any previous model can also be evaluated

Dataset used for evaluation

- Based on partitioning of the original data
  - Training : Generally not a good idea (biased)
  - Validation : Use for fine tuning
  - Testing : Real (final) evaluation data
  - Full : As a curiosity
- Enter : To enter new test data (only available with Score)
- CSV File
- R Dataset

Risk variable

- Chosen in the Data tab
- Used as a measure of how significant each observation is with respect to the target variable
- The risk chart makes use of this variable

Scoring

- Final row lets us choose between class output or probability output
- Include allows us to report all variables or just Ident variables

Cross-validation

- Some algorithms already apply cross validation
  - Decision tree (rpart)
  - Create 10 subsets
  - Choose 1 for validation, rest for training
  - Do for each subset, take average
- Then there is no need to further partition for validation
- Also, some algorithms use OOB
  - Random forest is an example
  - Again, no need for further validation partition
- Note that final testing is still important

Error rate

- Simplest measure
- Sum up the misclassifications
- Divide by total number of observations

True and False Positives and Negatives

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

- Often it is useful to differentiate error types
- Different error types may have different importance
- We control this with the loss matrix, such as in decision trees

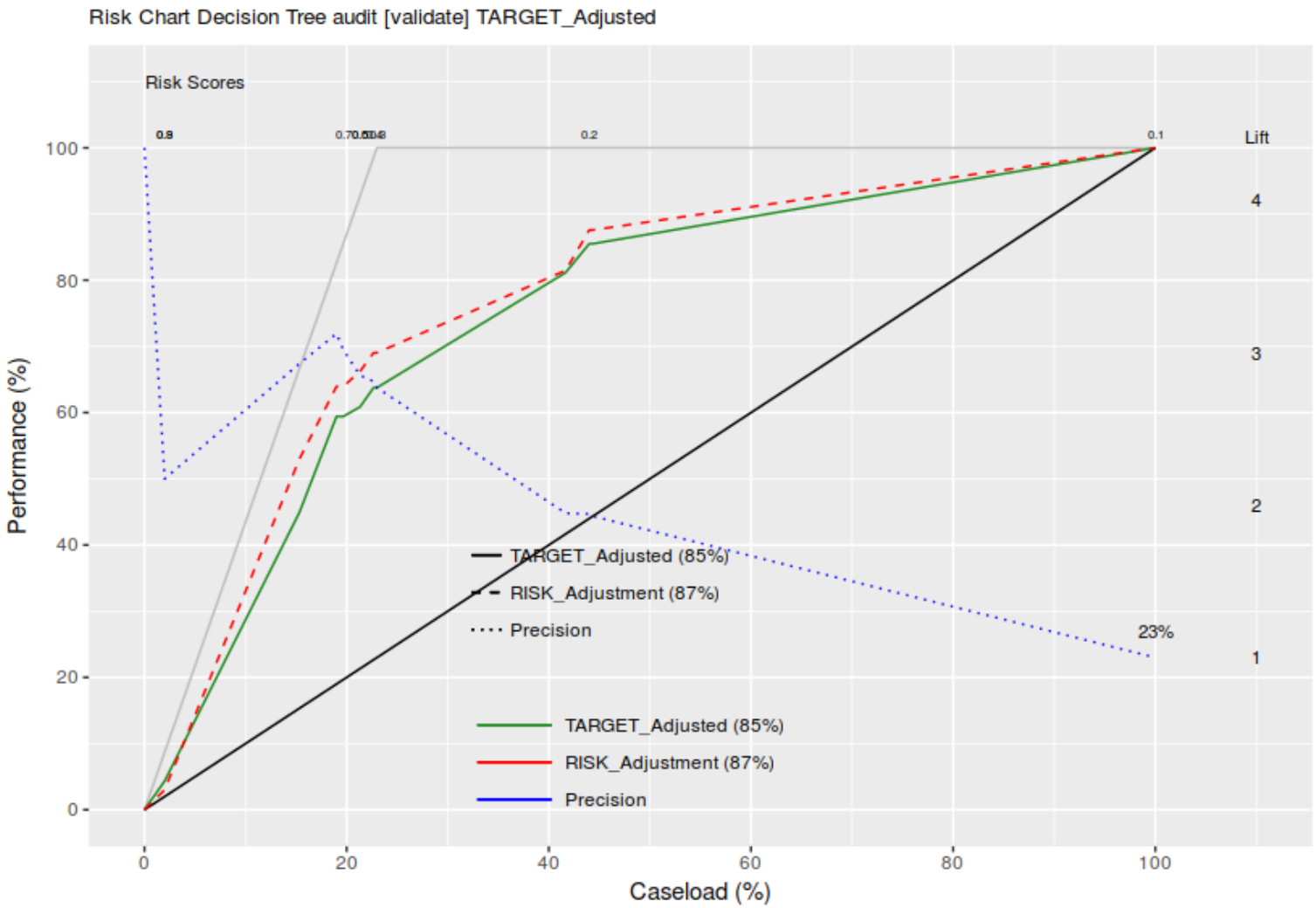
Precision, recall, sensitivity, specificity

- The **precision** of a model is the ratio of the number of true positives to the total number of predicted positives (the sum of the true positives and the false positives)
  - precision = TP / (TP + FP)
- **Sensitivity** (aka. recall) is the ratio of correctly classified positives to all actual positives
  - sensitivity = recall = TP / (TP + FN)
- **Specificity** is the ratio of correctly classified negatives to all actual negatives
  - specificity = TN / (TN + FP)

Risk Chart

- We will explain using the audit dataset
- In the audit dataset, target variable is TARGET_Adjusted
  - Yes means the taxpayers return was adjusted due to errors
  - No means the return requires no adjustments
- RISK_Adjustment is the amount of adjustment and will be the risk variable

Risk Chart


Risk Chart Decision Tree audit [validate] TARGET_Adjusted
Risk Scores
0.9
0.7
0.2
0.1
Lift
4
3
2
1
Performance (%)
100
80
60
40
20
0
TARGET_Adjusted (85%)
RISK_Adjustment (87%)
Precision
23%
TARGET_Adjusted (85%)
RISK_Adjustment (87%)
Precision
0
20
40
60
80
100
Caseload (%)

How does it work?

- Assume 100,000 tax payers
- Assume 24,000 requires adjustment
- We have money for 50,000 inspections
- On a random basis, we can catch 12,000 adjustments
- However, assume that you have a risk score attached to each taxpayer
- Now, you can prioritize based on this score to catch more adjustments

Reading the plot

- The diagonal shows the performance for random sampling
- The blue dashed line shows the lift: how much better the model is behaving compared to a random sampling process
- The green line tracks the target variable
- The dashed red line tracks the risk variable
- If the risk curve is well above the green curve, that is a positive thing: we catch high risk cases early

What risk analysis is not

- Risk is not used in modeling
- It is a way to measure performance
- You tune your model to increase risk performance
- The better model has a larger area under the curve

Comparing models under risk

- You can create the risk plots for each model and compare the charts
- Each plot includes the "area under the curve" values for both the target and the risk
- Higher the area, better the model

ROC Chart

- Plots True Positives vs False Positives
- Higher AUC is better