DSS
Decision Sciences & Systems

Technische Universität München

Tutorial Business Analytics

Tutorial 8: Evaluation

Decision Sciences & Systems (DSS)
Department of Informatics
TU München

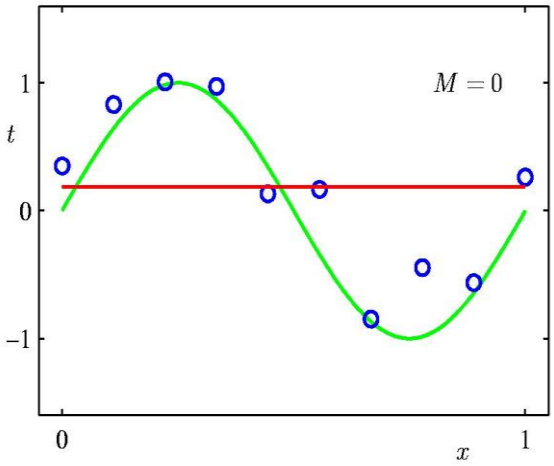
Tutorial Business Analytics

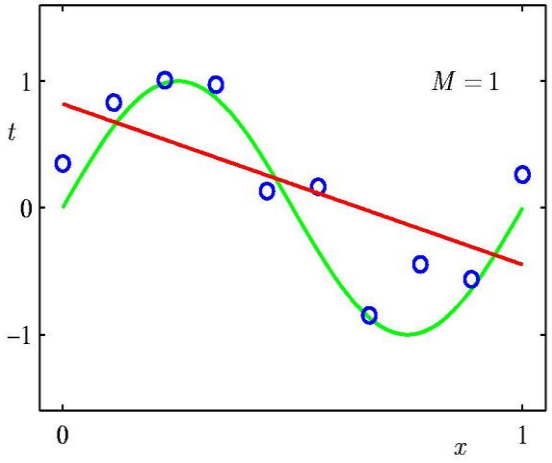
Agenda

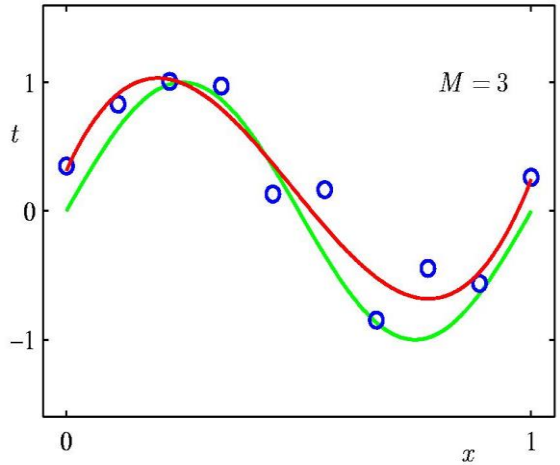
- Bias-Variance Tradeoff
- Evaluation of classifiers
  - Evaluation on training set
  - Holdout
  - Cross-Validation
- Metrics for classifiers
  - Confusion matrix
  - Gain Curves
  - Lift Curves
  - ROC Curves

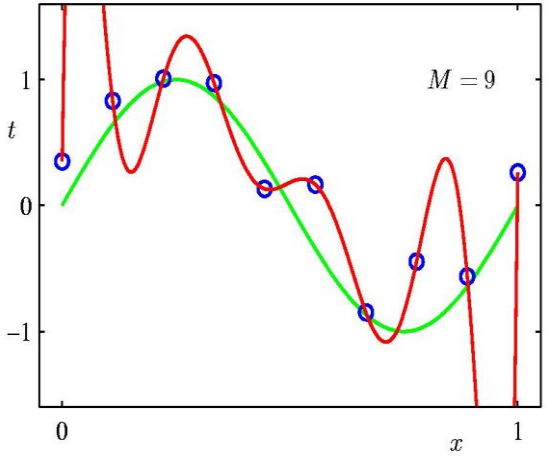
Tutorial Business Analytics

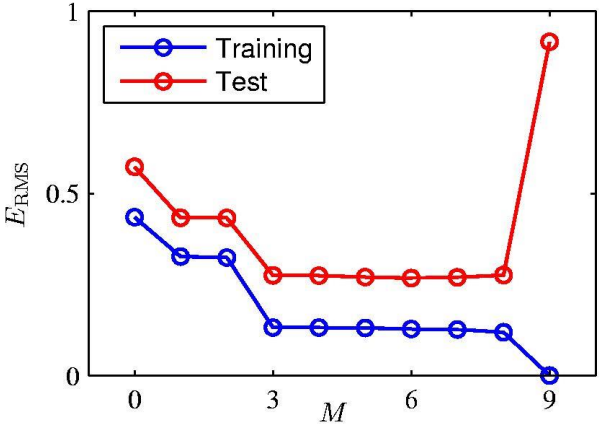
Bias-Variance Tradeoff











Root mean squared error

Tutorial Business Analytics

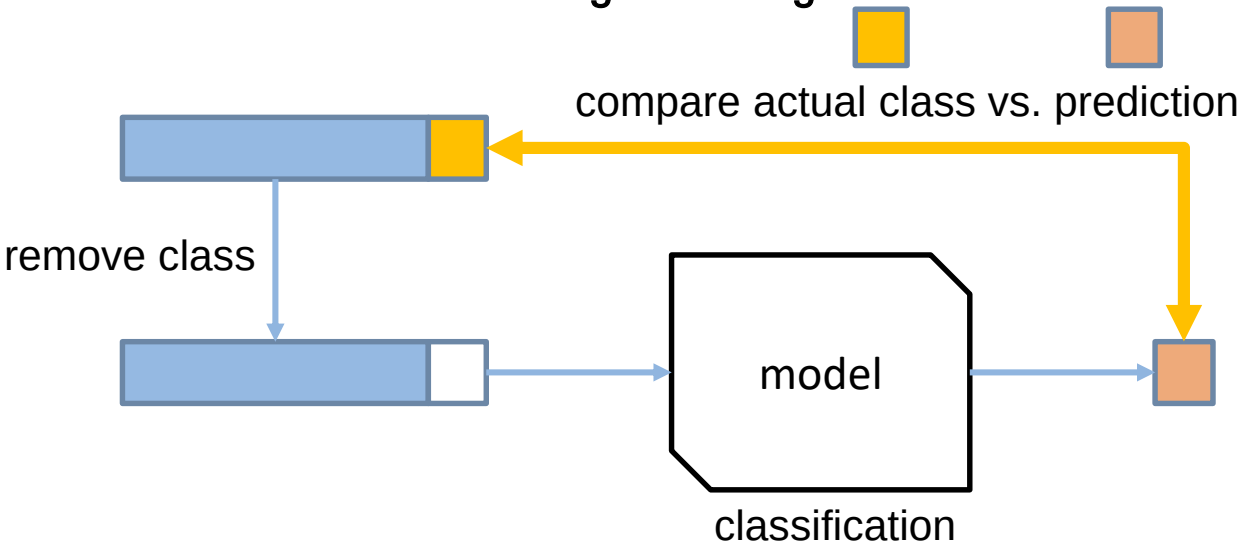
Evaluation – "How good is your model on new data?"

The model classifies an instance, of which the class is *already known*.
Then, *compare* model's prediction *to actual class*.

Which Instances to use for testing?

Instances with known class can be used for training and testing

Competition for classified instances: **training vs. testing**

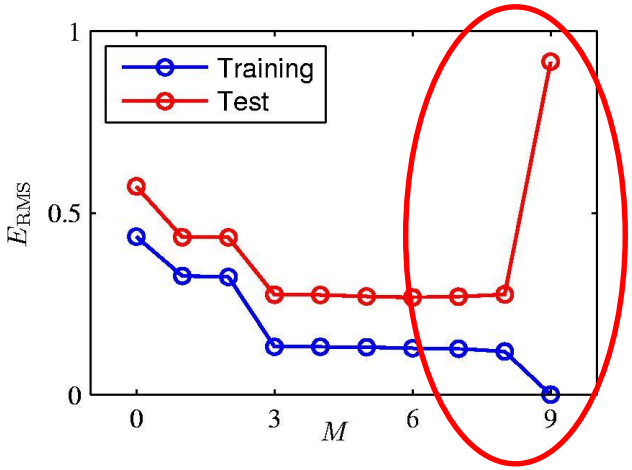


Tutorial Business Analytics

Evaluation using training sets

Using training data also for evaluation:

- Overfitting
- Evaluation is too optimistic, the actual error rate is way higher



Tutorial Business Analytics

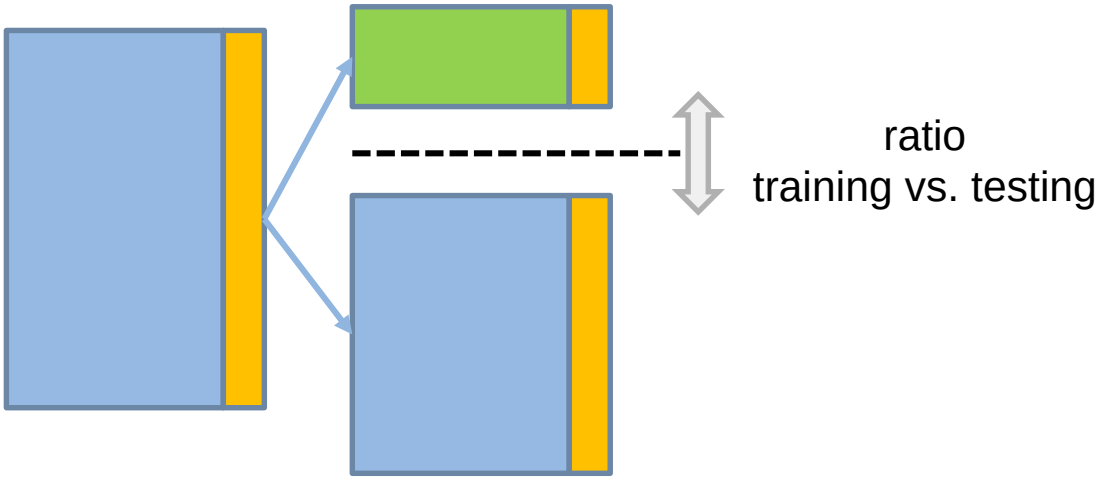
Holdout

Holding out certain instances for evaluation, not for training:

- Reduces the number of training instances
- Composition (e.g. distribution of the classes) influences results

Solution: Stratified holdout

- More reliable results using stratification
- Some instances may never be used for training/testing



Tutorial Business Analytics

k-fold Cross-Validation

- Partition the data set into k complementary subsets
- Train on $k-1$ subsets, test on 1 subset
- Every subset is used $k-1$ times for training and 1 time for testing
- Stratification of subsets is possible

Example: k = 5

Average error,

$$\hat{e} = \frac{1}{k}\sum_{i=1}^{k} e_i$$

Tutorial Business Analytics

Leave One Out Validation

- k-fold cross validation with k = N, where N is the number of training instances
- Deterministic, a lot of time required for computing
- Extreme class distribution of test data
- Example: N = 5000
  each of which is used 4999 times for training and 1 time for testing
  5000 models need to be trained

Tutorial Business Analytics

Statistical tests – t-test

Results of a **validation** may be considered as random chance

- Only *significant* difference between classifiers is interesting

A paired t-test checks whether mean accuracy of two classifiers are significantly **different**

- H_0: $d = 0$
- H_1 : $d \neq 0$ (depend on question)

$$\bar{d} = \frac{1}{k}\sum_{i=1}^{k} d_i \qquad s_d = \sqrt{\frac{1}{k-1}\sum_i \left(d_i - \bar{d}\right)^2} \qquad t = \frac{\bar{d}}{s_d / \sqrt{k}} \sim t_{k-1}$$

Tutorial Business Analytics

Confusion Matrix

Result of an evaluation containing

- the true class
- the class predicted by the classifier

True Class	Predicted Class	
1	1	True Positive
1	0	False Negative
0	1	False Positive
0	0	True Negative

Confusion Matrix

		Predicted Class	
		Positive	**Negative**
True Class	**Positive**	True Positive	False Negative
	Negative	False Positive	True Negative

Tutorial Business Analytics

Predicting Probabilities

Classifiers can predict probabilities instead of classes

- A probability is assigned to every class
- Classify using a cutoff
- Helpful in case errors have different impacts (costs)
  - Minimize cost by selecting a suitable cutoff-value

Example:

- The cost of FN is 5 times the cost of FP
- Adjust the classifier to predict positive rather than negative
- Decrease the cutoff-value

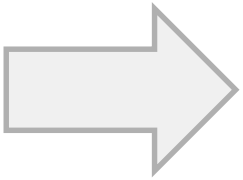
		Predicted Class	
		Positive	**Negative**
True Class	**Positive**	0	5
	Negative	1	0

Tutorial Business Analytics

Predicting Probabilities

The probability of a positive class is displayed, instead of the predicted class (cutoff-value=0.5)

True Class	Predicted Class
0	0
0	1
1	1
1	0
0	0
1	0
0	0
1	1
0	1
1	0



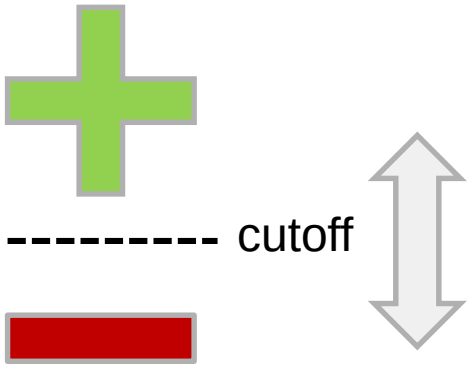
True Class	Predicted Class	Probability
0	0	0.34
0	1	0.56
1	1	0.63
1	0	0.45
0	0	0.47
1	0	0.17
0	0	0.26
1	1	0.89
0	1	0.52
1	0	0.43

Tutorial Business Analytics

Predicting Probabilities

- Sort by probability:

True Class	Probability
+	0.89
+	0.63
-	0.56
-	0.52
-	0.47
+	0.45
+	0.43
-	0.34
-	0.26
+	0.17

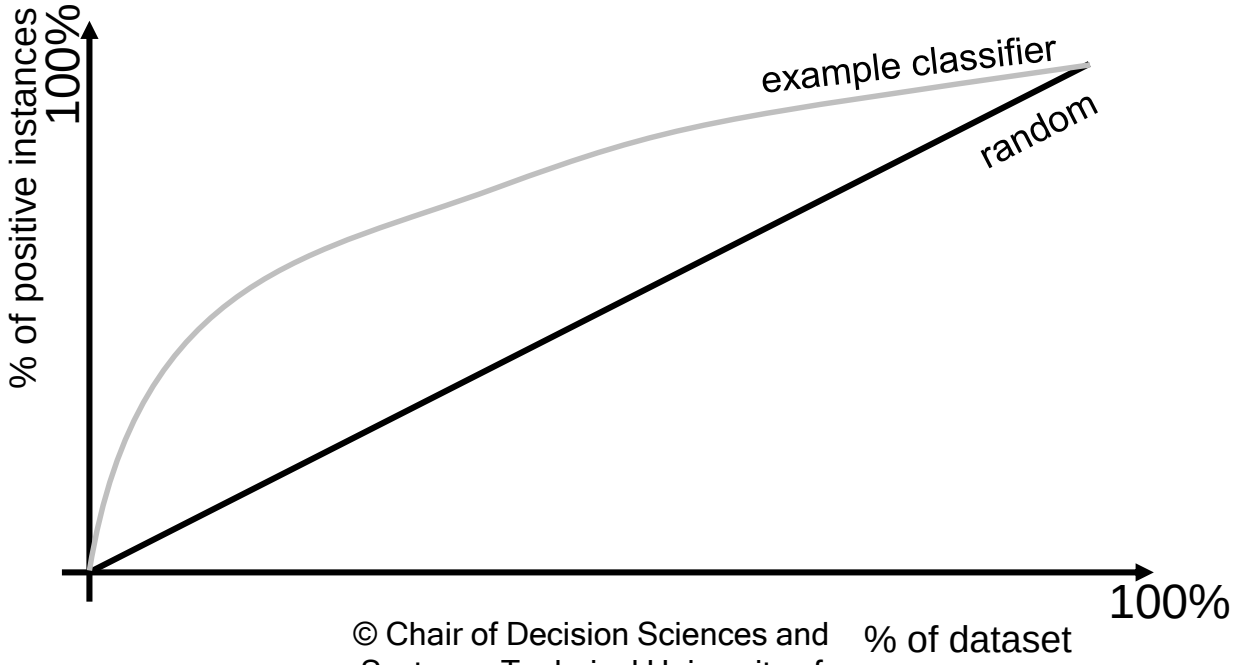


Tutorial Business Analytics

Gain Curve

To visualize results of different cutoffs

- Instances are sorted by probability (descending)
- x-axis: percentage of the data set (or number of instances)
- y-axis: percentage (or number) of positive instances in the fraction (subsample)

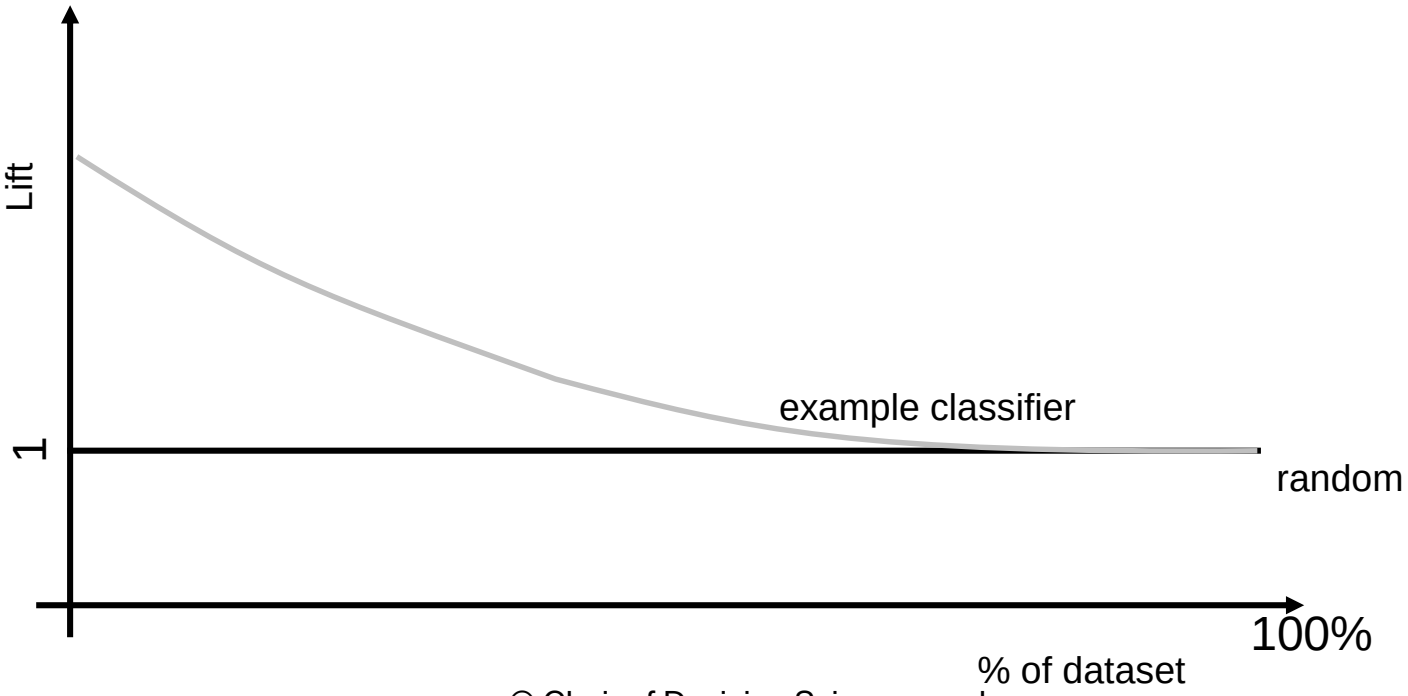


Tutorial Business Analytics

Lift Curve

Displays the factor between the classifier and random value for every part of the gain curve

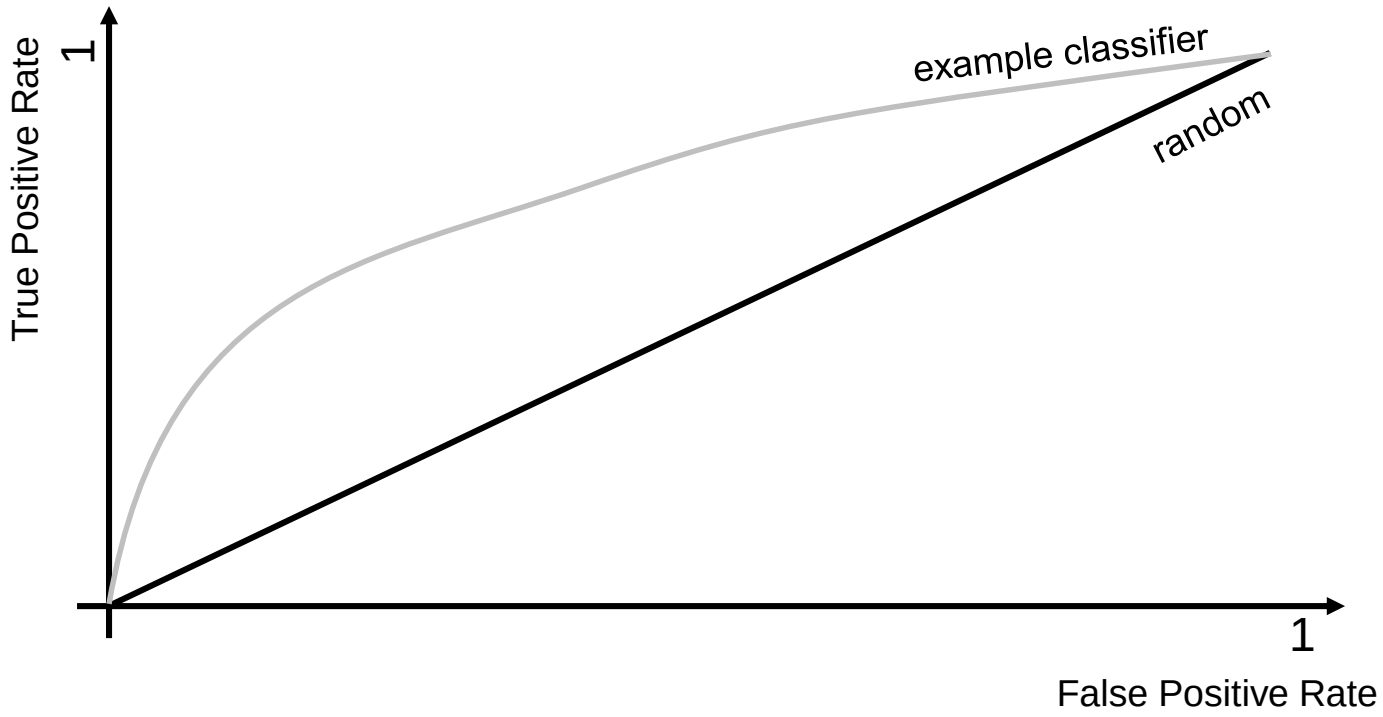
- x-axis: percentage of the data set (or number of instances)
- y-axis: $\frac{Gain\ at\ x}{x}$



Tutorial Business Analytics

ROC Curve

Displays the ratio of False Positive Rate and True Positive Rate



Tutorial Business Analytics

Agenda

- Bias-Variance Tradeoff
- Evaluation of classifiers
  - Evaluation on training set
  - Holdout
  - Cross-Validation
- Metrics for classifiers
  - Confusion matrix
  - Gain Curves
  - Lift Curves
  - ROC Curves