

classification
supervised machine learning

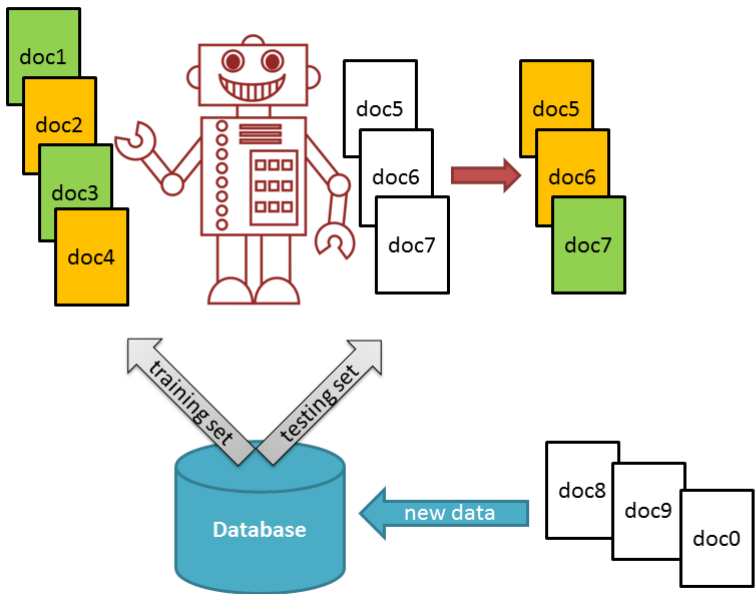
DTL SU @ AU

kristoffer | nielbo
kln@cas.au.dk
github.com/kln-courses/tmgu17
tmgu17.slack.com

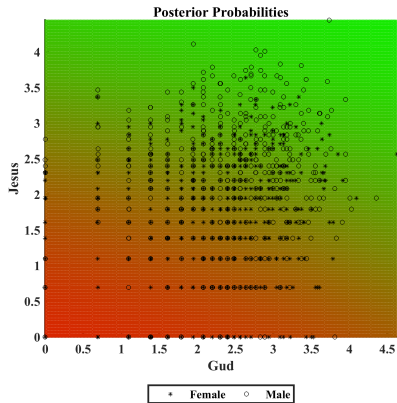
DAI|IMC|AARHUS UNIVERSITY

classification

- given labeled data (supervised learning), a classification algorithm will output a solution that categorizes new examples → associate labels with subsets of the data
- while clustering (unsupervised learning) searches for groups within the corpus, classification learns to map a collection of documents onto a categorical class values or labels → find mapping function
- data (features) with class values (\sim labeled data), excellent opportunity to make use of metadata
- vast majority of models are **black box models**
- workflow: separate data set in training and test subsets (training, test, and validation) → train model → test model → apply model to new data



classification in the humanities



labeled data the correct class information is available

- metadata is readily available, e.g. author, genre, year of publication
- labels from an external source/databases, e.g. reviews, ratings, reads
- annotate data (expert or raters)

evaluation

Estimate performance (error rate) of a classifier (the lower the error the better). Often several classifiers are compared

most metrics are developed for binary classification problems

confusion matrix: table for describing performance of classifier on training and/or testing data

		True	
		positive	negative
Predicted	positive	TP	FP
	negative	FN	TN

True Positive: Correctly assigns positive class membership

True Negative: Correctly rejects class membership

False Positive: Fail to rejects class membership

False Negative: Reject class membership incorrectly

we train a Naive Bayes classifier on 1500 verses of the KJV Bible labeled with collection data (NT: New Testament OT: Old Testament)

confusion matrix for binary classification problem:

	NT	OT
NT	644	89
OT	106	661

, verses: $644 + 106 + 89 + 661 = 1500$

accuracy

measures in how many cases the predicted class conformed with the correct class:

$$\frac{TP + NP}{TP + TN + FP + FN}$$

	NT	OT
NT	644	89
OT	106	661

, accuracy: $\frac{(644 + 661)}{1500} = 0.87$ (87%)

precision

measures the number of selected verses that are relevant, i.e., how certain are we that a classified verse is correctly classified (\sim how many time did the model positively predict a class): $\frac{TP}{TP + FP}$

	NT	OT
NT	644	89
OT	106	661

, $precision_{NT} = \frac{(644)}{644 + 89} = 0.88$

for each class label: How many of the items that got the label should have gotten it? How many should have gotten other labels?

recall

recall measures the number of relevant verses that are selected, i.e., how good is the classifier at detecting verses within a given class: $\frac{TP}{TP + FN}$

	NT	OT
NT	644	89
OT	106	661

$$, \text{recall}_{NT}: \frac{(644)}{644 + 106} = 0.86$$

for each class label: How many items that should have gotten the label did get it? How many were missed?

composite (general) measure of a classifier's accuracy

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$F_1 : 2 \times \frac{.88 \times .86}{.88 + .86} = 0.87$$

F is the harmonic mean of precision and recall.

