

Machine learning

1. Supervised learning
 - ▶ Text classification
 - ▶ Sequence labeling
 - ▶ Structure learning
 - ▶ Sequence to sequence
 - ▶ Learning to rank
2. Unsupervised learning
 - ▶ Cluster analysis
 - ▶ Embedding models
 - ▶ Topic modelling
3. Reinforcement learning
4. Other learning paradigms:
 - ▶ Semi-supervised learning
 - ▶ Active learning

Today: Text classification

Examples:

- ▶ Spam filtering
- ▶ Language identification
- ▶ Topic / genre / sentiment / intent analysis
- ▶ Author profiling or identification

Methods:

- ▶ Rules
- ▶ `from sklearn import *`
- ▶ Neural networks

Tasks:

- ▶ Binary classification
- ▶ Multiclass classification
- ▶ Multilabel classification (categorization)

Text classification

Metrics:

- ▶ Accuracy
- ▶ Precision, recall, f-measure
- ▶ ROC-AUC

Datasets:

- ▶ Reuters-21578 Text Categorization Collection
- ▶ IMDb dataset
- ▶ newsgroup20
- ▶ ...