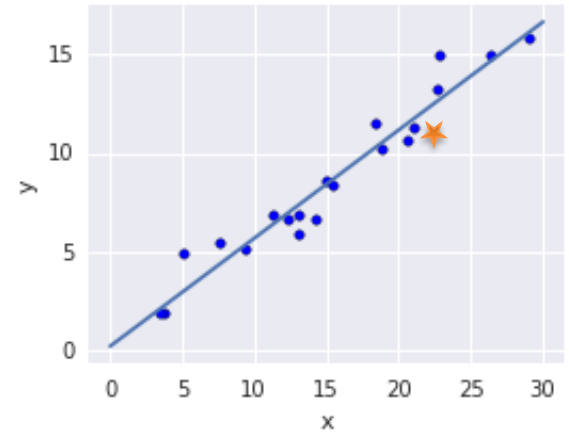
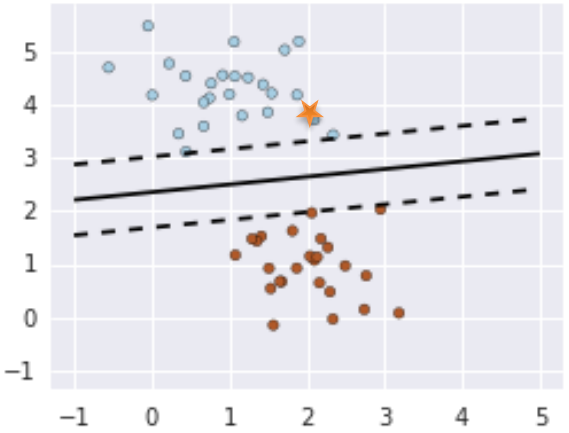
Machine learning

- Machine learning explores the study and construction of algorithms that can learn from and make predictions on data

- Unsupervised learning
  - clustering
  - dimensionality reduction

- Supervised learning
  - regression (label is continuous)
  - **classification** (label is discrete)

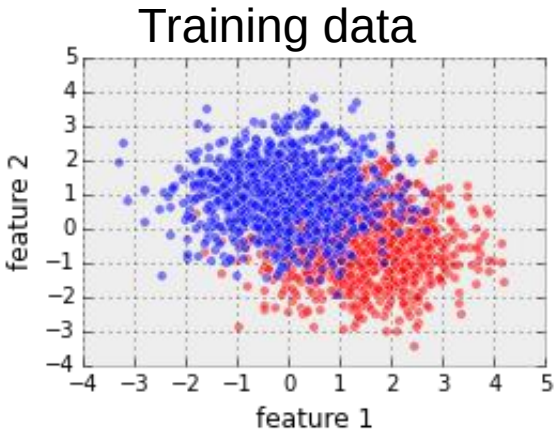




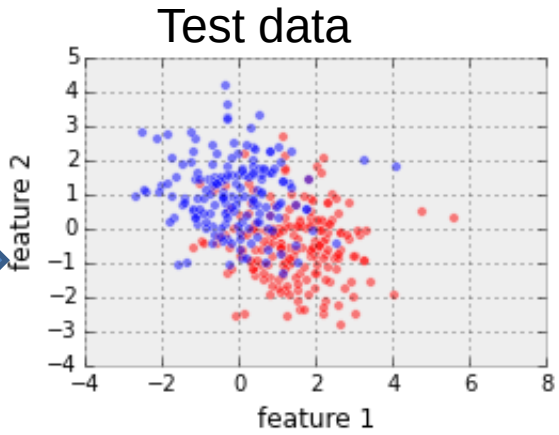
Machine learning – training and validating a binary classification model

Point:

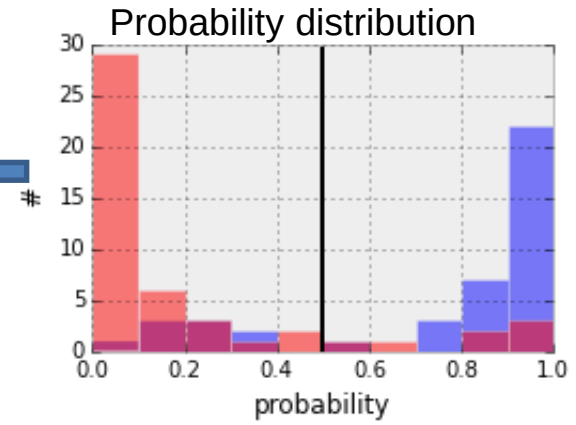
- Sample with features 1 and 2
- Color indicates class

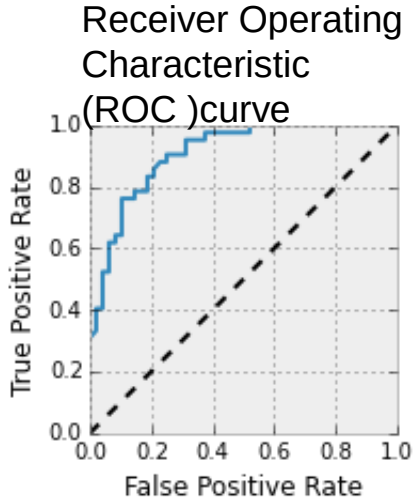


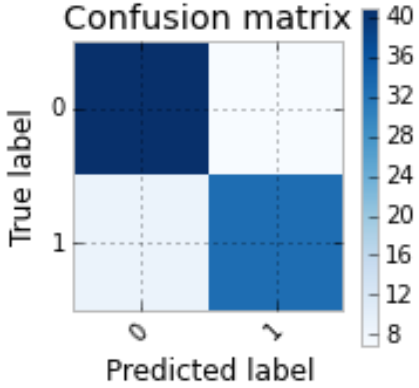
Train your model



Make predictions on test set



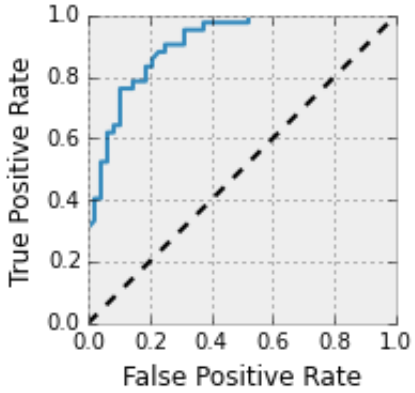




Machine learning – performance metrics

- Accuracy: fraction of correct predictions
- Area under the ROC curve (AUC)
  - 0.5 – performance of random selection
  - 1 – ideal classifier
  - > 0.9 diagnostic tests in medicine
- Confusion matrix
  - Sensitivity
    - Recall of positive, TP fraction
    - TP/ (TP + FN)
  - Specificity
    - Recall of negative, TN fraction
    - TN/ (TN + FP)

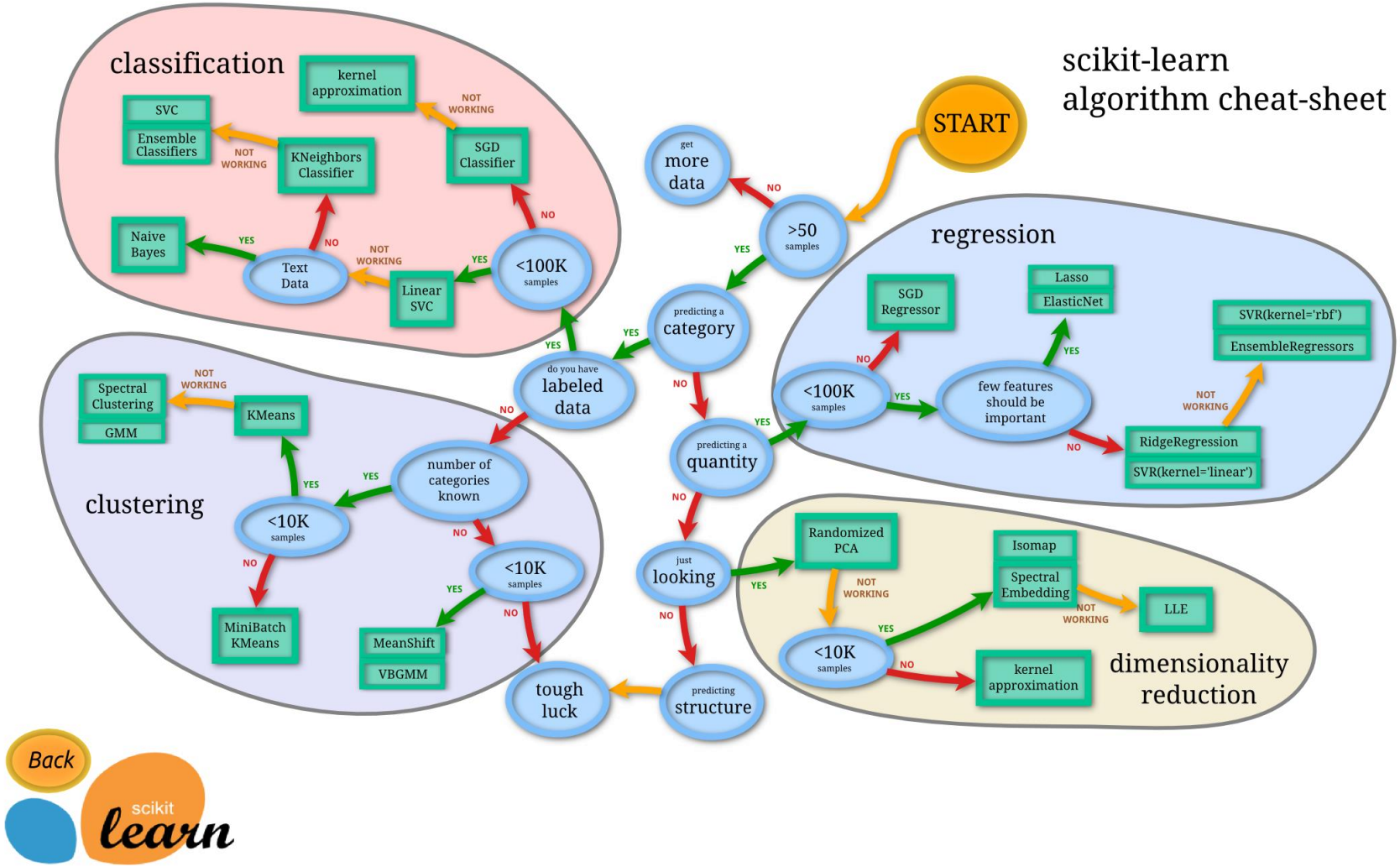
Receiver Operating Characteristic (ROC)curve



		Predicted	
		Positive	Negative
Actual	Positive	**TP**	**FN**
	Negative	**FP**	**TN**

TP – True positive, TN –True negative
FP – False positive, FN – False negative

Machine learning in Python
Scikit-learn



http://scikit-learn.org/stable/tutorial/machine_learning_map/