K-Means

Linear regression

Logistic regression

Mean Shift

Nearest neighbors
KNeighborsClassifier

Support Vector Classifier
SVM

Cross Validation
KFold

Principal component analysis
Randomized PCA

Naive Bayes
Multinomial NB

Support Vector Regression
SVR

Stochastic Gradient Descent
SGD Regressor

Neural networks
MLP Classifier

- Classification
- Linear: predicts dichotomous values.

“Is this Iris setosa or Iris versicolor?”

- Regression
- Predicting linear data.
- Challenge: underfitting?
- Weight of each feature is same.

“Based on how tall someone is, what could their shoe size be?”

- Clustering
- Finding similarities in n-dimensional data.
- Uses euclidean distance: data needs to be integers
- We know how many types we want

“What are the *n* types of employees are there in a company?”

- Classification
- Effective in high-dimensional spaces (using different kernels).

“Is this setosa, versicolor or virginica?”

- Classification
- Easy interpretation, low calculation time.
- Predictive power a bit lower than Random Forest.

“Is this setosa, versicolor or virginica?”

- Clustering
- Discover blobs
- We don’t know how many types we want

“What types of criminals can we identify?”

- Classification
- Text classification (Natural Language Processing).

“Is this sentence related to sports?”

- Dimension reduction

“I want to describe this glass of wine, but it has so many different characteristics!”

- Model validation

“Is linear regression or machine learning better for my problem?”

- Classification
- Weakness: prone to overfitting.

“Is this a cat or not?”

- Regression
- Escape shallow local minima more easily than “regular” gradient descent

“What is the price of a house of this size?”

- Regression
- Weight of each feature reflects features correlation

“Based on it’s physical measures, how old is this abalone?”