



Artificial Intelligence Club: Classification Methods

Week 9



Created by Allgirlithm Outreach: <https://www.allgirlithm.org/>

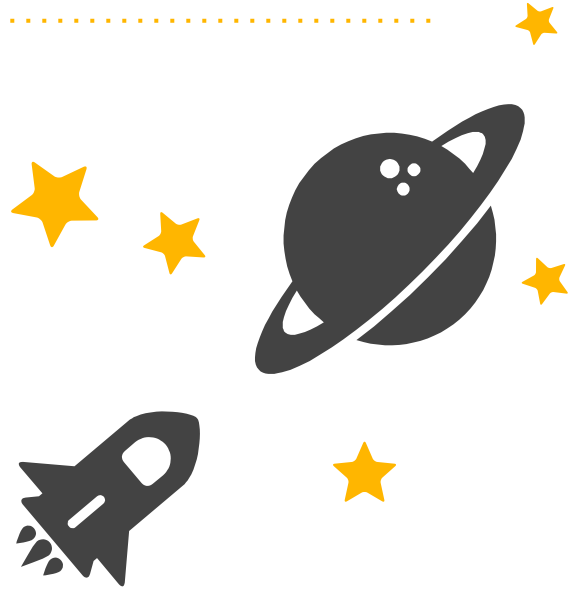
allgirlithm



Today's Agenda

- Supervised learning revisited
- Three methods
 - K-nearest neighbors
 - Support vector machines
 - Random forests and decision trees

Supervised Learning



*Machine is trained using labelled data
Correct/incorrect "answer" is known*

1

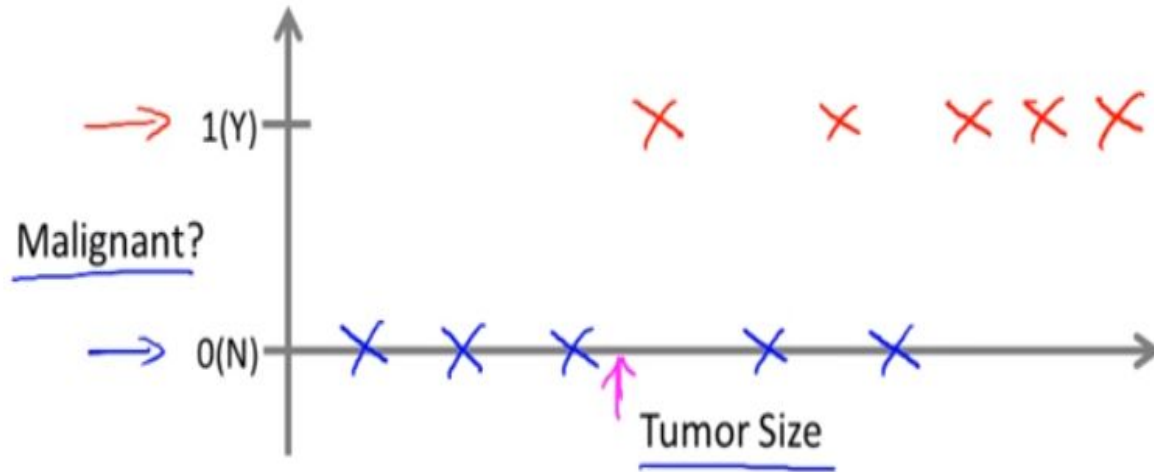
K-Nearest Neighbors

Birds of a feather flock together

Supervised Learning Example



Breast cancer (malignant, benign)

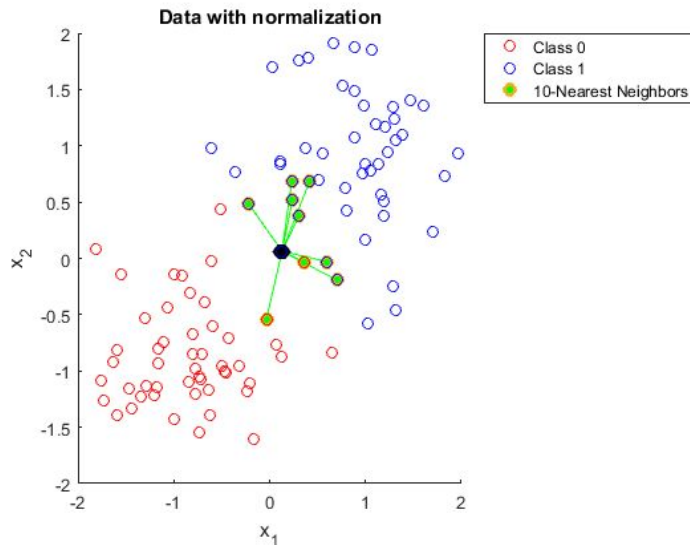


Classification
Discrete valued
output (0 or 1)

K-Nearest Neighbors



- Similar things are grouped close to each other
- Finds distance between test data and labelled data
- Selects number of examples (K) closest to the test data
 - Most frequent label: classification
 - Average label: regression



2

Support Vector Machines

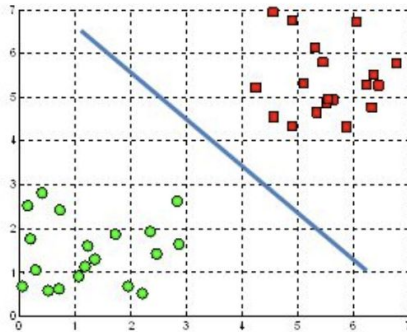
Finds a hyperplane that splits sections of data

Support Vector Machines

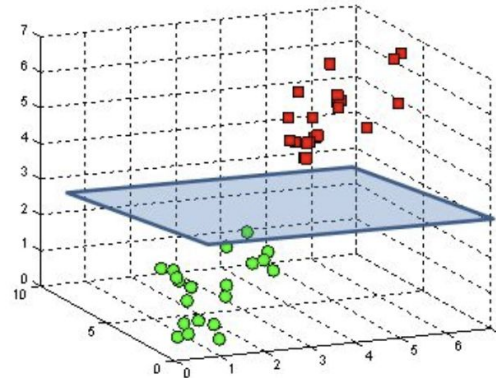


- For classification purposes
- Tuning parameter: kernel, regularization, gamma, margin

A hyperplane in \mathbb{R}^2 is a line



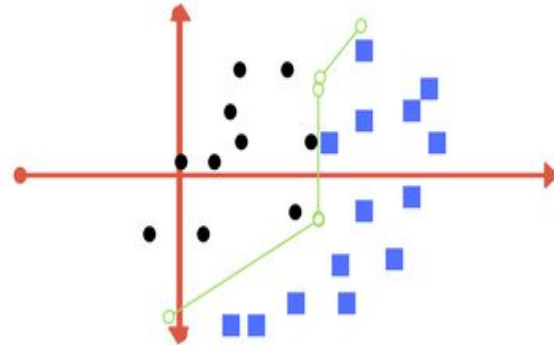
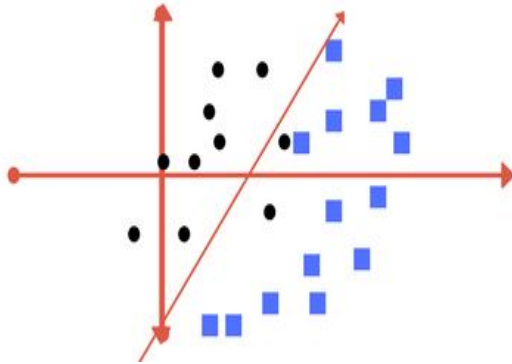
A hyperplane in \mathbb{R}^3 is a plane





SVM Parameters

- Kernel - can change line of separation (linear, polynomial, exponential)
- Regularization - how much you want to avoid misclassifying



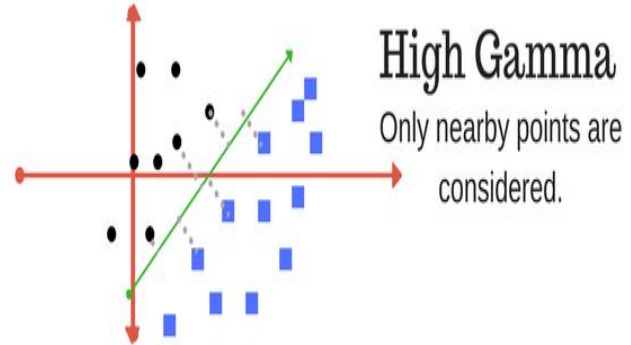
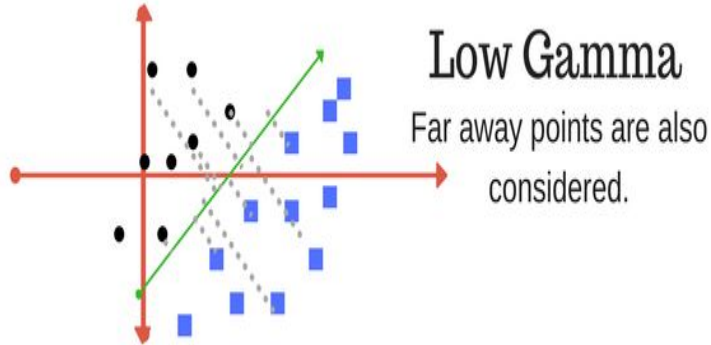
Left: low regularization

Right: high regularization



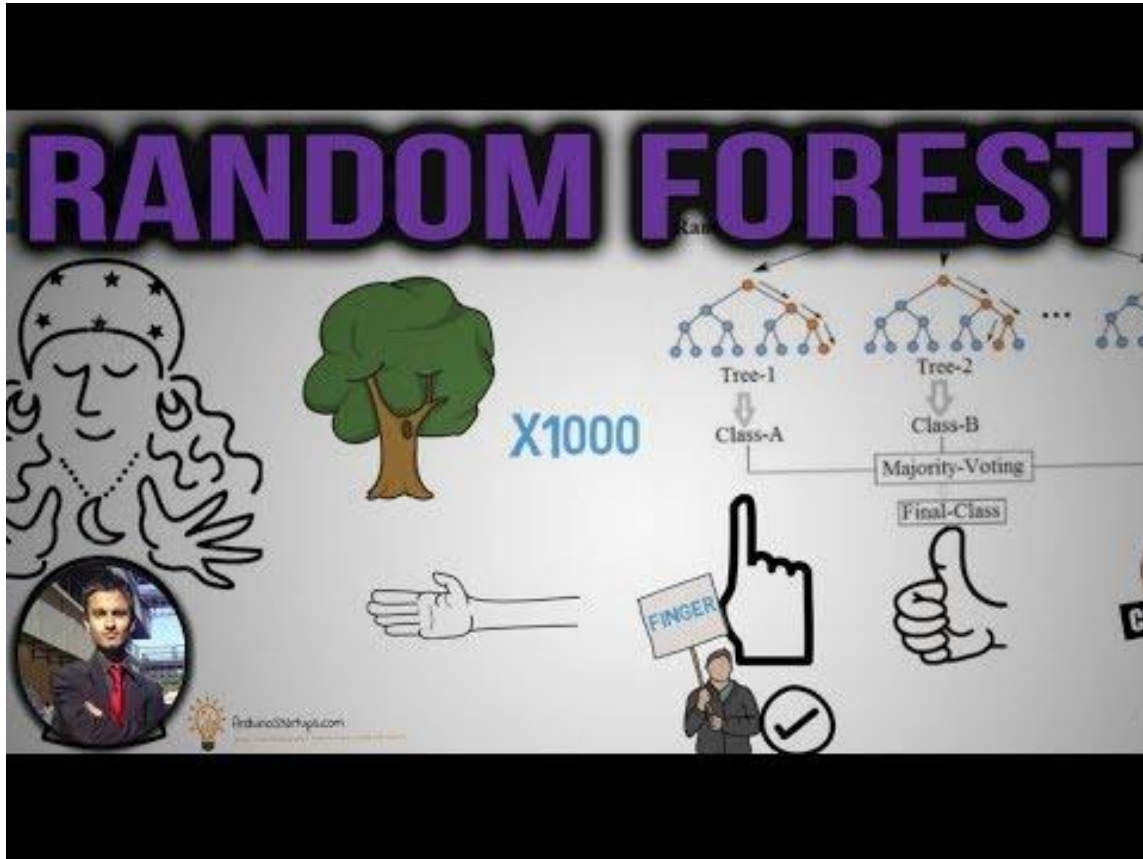
SVM Parameters

- Gamma - what types of data points are considered based on distance
- Margin - separation of the line to the closest class points



Random Forests and Decision Trees

Classifies data based on multiple features





Sources

- <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>
- <https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>
- https://www.youtube.com/watch?v=D_2LkhMJcfY