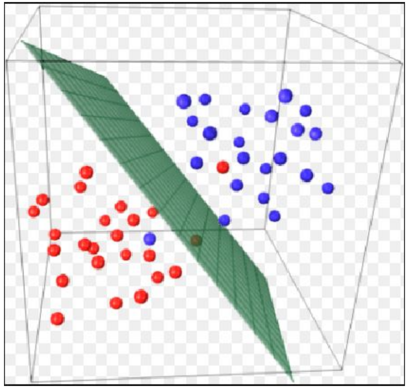
Support Vector Machines (SVMs)

Agenda

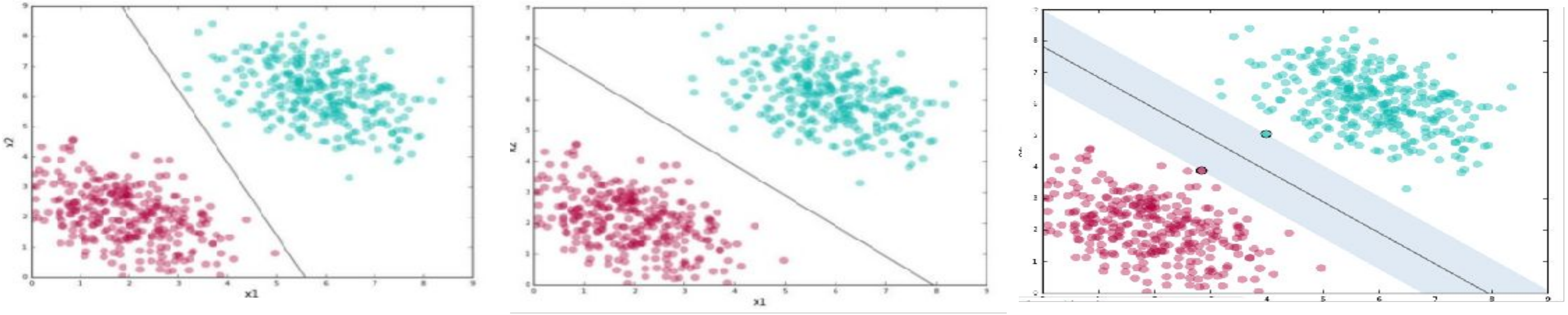
1. Introduction to SVM
2. How SVM looks in 2D space
3. Kernel SVM
4. SVM Parameters
5. Industry Applications of SVM

Introduction to SVM

Known as maximum-margin hyperplane, finds the linear model with max margin. Unlike the linear classifiers, objective is not minimizing sum of squared errors but finding a line/plane that separates two or more groups with maximum margins.



Introduction to SVM



- First line does separate the two sets but id too close to both red & green data points.
- Chances are that when this model is put in production, variance in both cluster data may force some data points on wrong side

Introduction to SVM

- The second line doesn't look so vulnerable to the variance. The two points nearest from different clusters define the margin around the line and are support vectors.
- SVMs try to find the second kind of line where the line is at max distance from both the clusters simultaneously.

The Perceptron

- The perceptron is an algorithm used to produce a binary classifier. That is, the algorithm takes binary classified input data, along with their class membership, and outputs a line that attempts to separate data of one class from data of the other: data points on one side of the line are of one class and data points on the other side are of the other.

How SVM looks in 2D space

- In a 2d space, the separating plane is a line.
- As described, the Perceptron tries to do the same
- The fig1 at the right shows a possible line.
- This line when put to classification for unseen data, is more prone to errors because of the variance in data.
- The SVM uses a line that looks more like fig2
- The hyperplane is a plane that acts as the decision boundary.

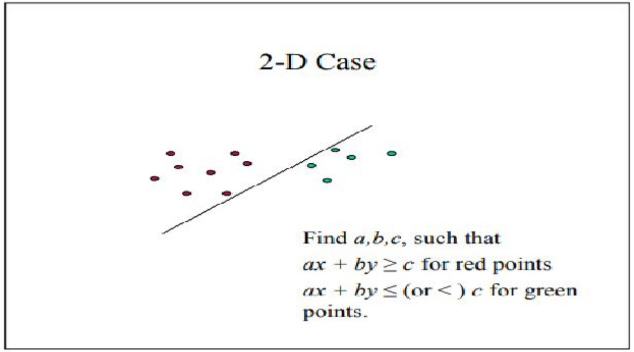


Fig1 - A separating line

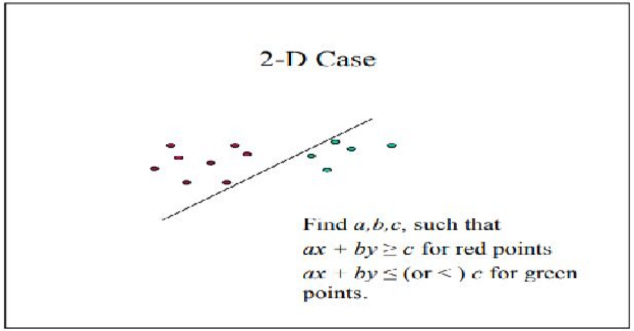


Fig2 - An optimal separating line (SVM)

How SVM looks in 2D space

Which Hyperplane to pick?

- Lots of possible solutions for a,b,c.
- Some methods find a separating hyperplane, but not the optimal one (e.g., neural net).
- But: Which points should influence optimality?
    - All points?
        - - Linear regression
        - - Neural nets
    - Or only “difficult points” close to decision boundary
        - - Support vector machines

Kernel SVM

- We know that when we transform the mathematical space from 2 dimensions to higher dimension, the probability of linearly separating the data points increases.

- Now given that we have only 2 dimensions x_1 and x_2, how to create more dimensions out of it? We transform higher mathematical space into dimensions which are polynomials.

- x_1, x_2, x_{12}, x_{22},...each one of them is 1 dimension in the mathematical space. Kernel SVM (KSVM) takes data point to higher mathematical space where they become linearly separable and then draw the plane through the data points.

SVM Parameters

- Kernel
  - Used to specify the type of kernel we choose to describe the data points. We need kernels as data when scaled on higher dimensions has a higher probability of being linearly separable. In sklearn, we can use many kernels such as – rbf, poly, sigmoid, linear, precomputed etc.
- C
  - Defines the misclassification error of the model
  - If set to high, will have a very high penalty for the misclassified points and vice versa.
- Gamma
  - Defines the radius of influence of data points in classification.
  - By Increasing gamma, we have a tight radius of influence of the data points in the classification.

These parameters need to be tried for different values to come to the optimum value/ highest accuracy model.

Industry Applications of SVM

1. Text (and hypertext) categorization
2. Image classification
3. Bioinformatics (Protein classification, Cancer classification)
4. Handwritten character recognition

greatlearning

Power Ahead

Happy Learning !