MACHINE LEARNING

Shoaib Farooq

Department of Computer Science

Disclaimer: This presentation includes contents available online including images copied from Google search and contents of presentations of other professors. I don't claim any image or text to be my own. All the credit goes to the original authors.

Instructor Information

Instructor	Shoaib Farooq	E-mail	m.shoaib1050@gmail.com
 GitHub	https://github.com/SHOAIB1050		
 YouTube	https://www.youtube.com/c/pakithub		
 LinkedIn	https://www.linkedin.com/in/shoaib-farooq-b5b190105/		

Let's Start .....

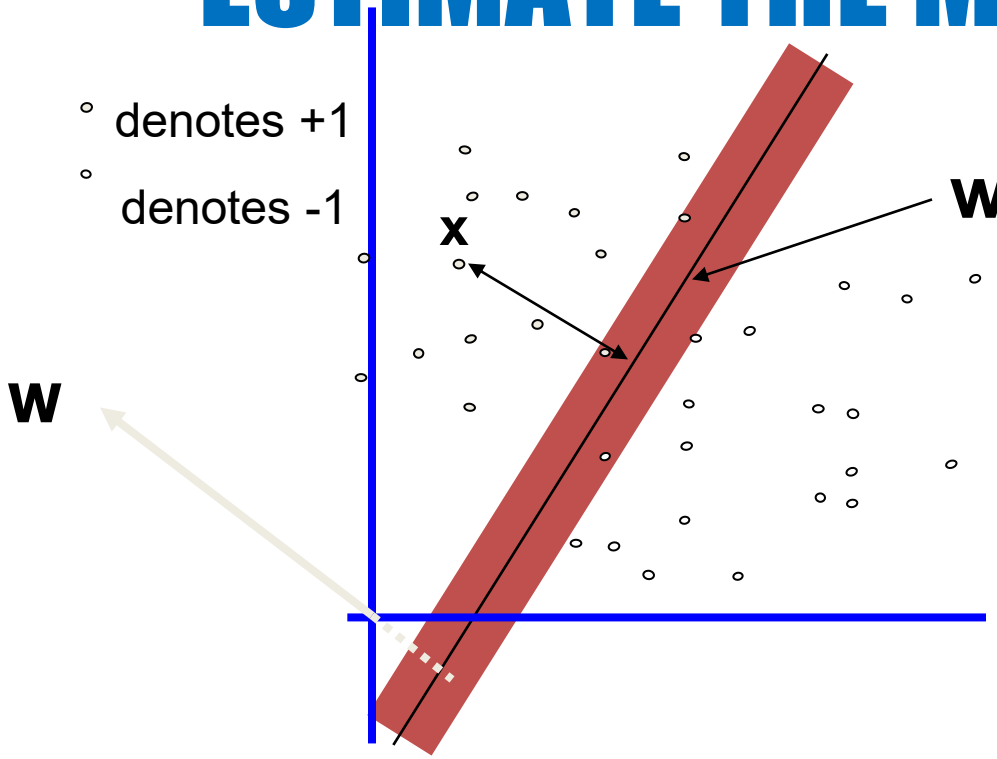
Lecture #10

GOALS

This Lecture Will Cover:

- **Evaluation of Support Vector Machine**

ESTIMATE THE MARGIN



x – Vector

w – Normal Vector

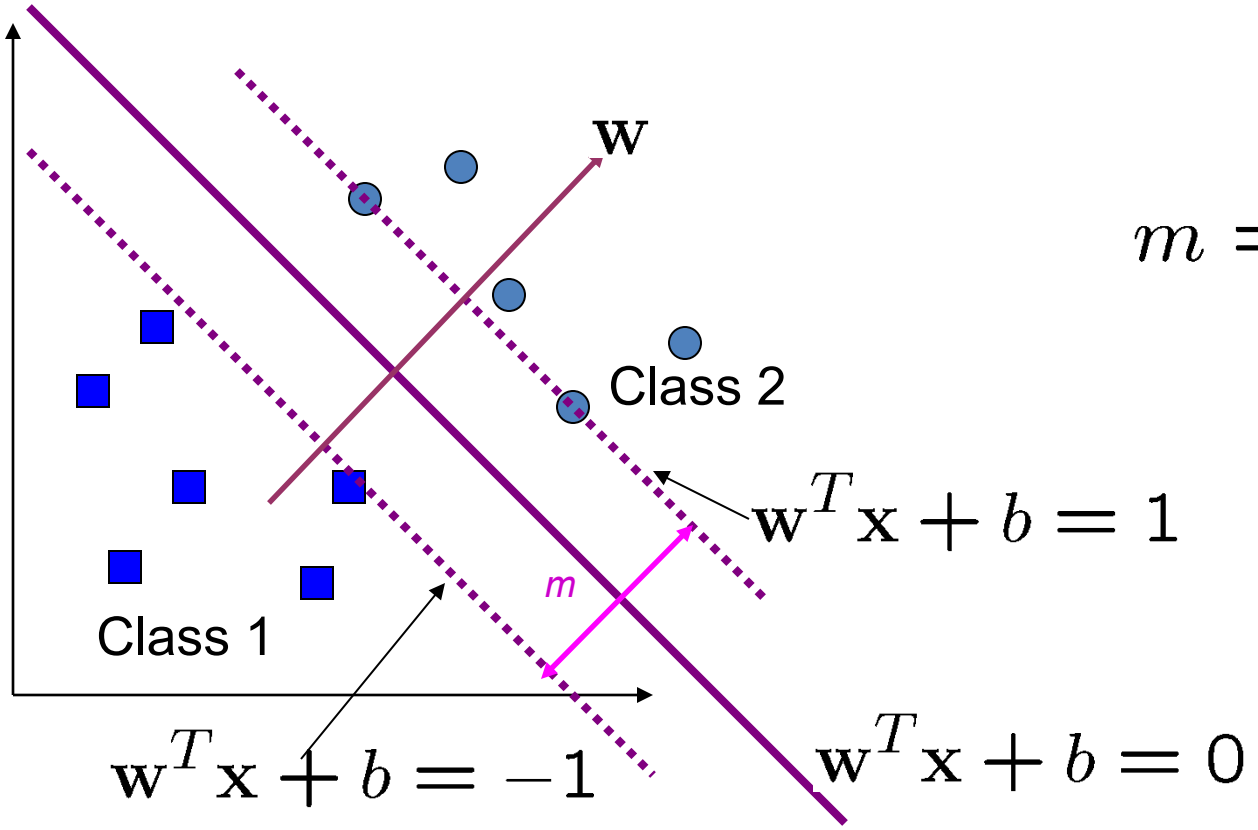
b – Scale Value

- What is the distance expression for a point **x** to a line **wx**+b= 0?

$$d(\mathbf{x}) = \frac{|\mathbf{x}\cdot\mathbf{w}+b|}{\sqrt{\|\mathbf{w}\|_2^2}} = \frac{|\mathbf{x}\cdot\mathbf{w}+b|}{\sqrt{\sum_{i=1}^{d} w_i^2}}$$

LARGE-MARGIN DECISION BOUNDARY

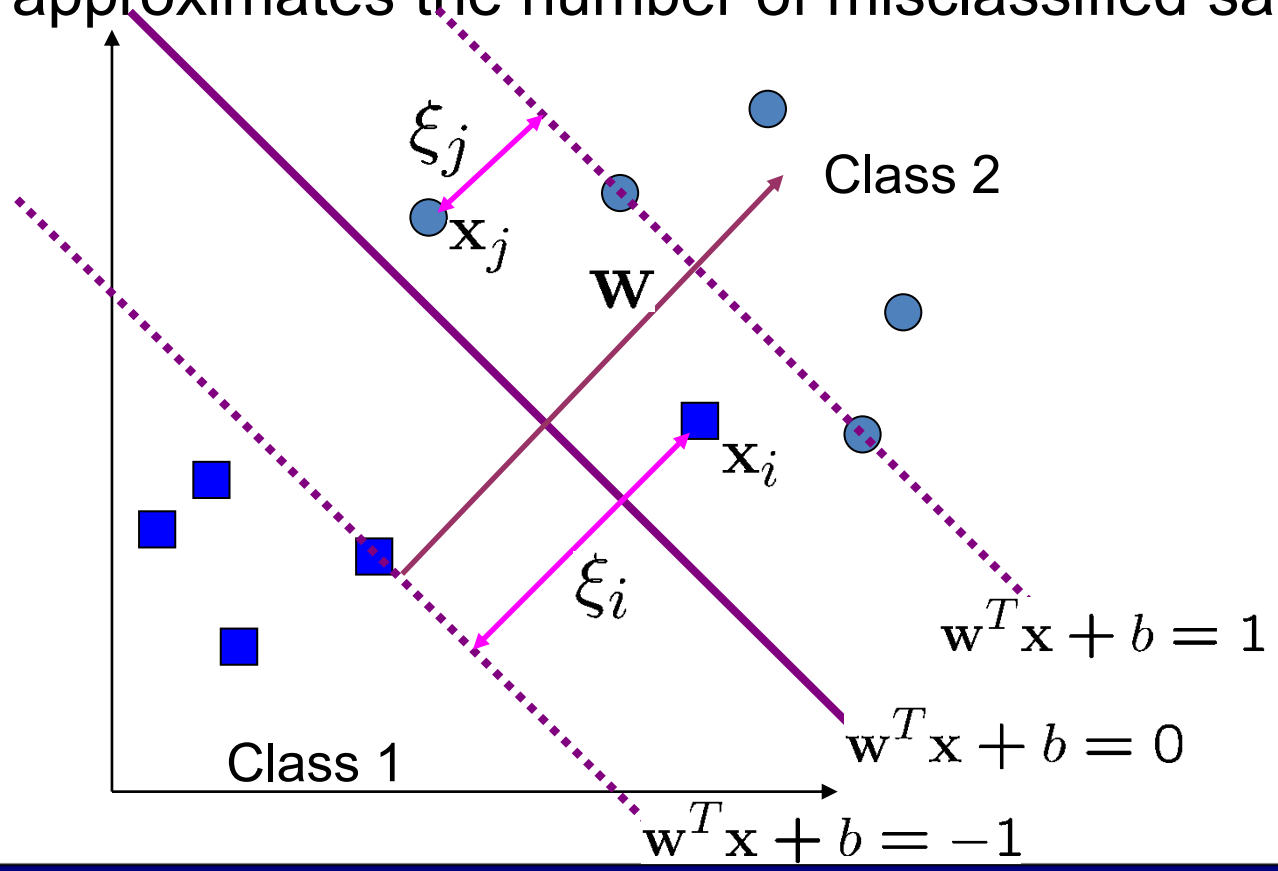
- The decision boundary should be as far away from the data of both classes as possible
  - We should maximize the margin, *m*



$$m = \frac{2}{||\mathbf{w}||}$$

ALLOWING ERRORS IN OUR SOLUTIONS

- We allow "error" ξ_i in classification; it is based on the output of the discriminant function $\mathbf{w}^T\mathbf{x}+b$
- ξ_i approximates the number of misclassified samples



SOFT MARGIN HYPERPLANE

In Support Vector Machines (SVM), a **fixed soft margin** allows for some degree of misclassification by introducing a *penalty* for misclassified points. This technique helps to make the SVM model more robust when dealing with non-linearly separable data or data that may contain noise.

SOFT MARGIN HYPERPLANE

Soft Margin Concept

In a perfect, linearly separable dataset, a hard margin SVM will maximize the margin between two classes without allowing any misclassification. However, in real-world data, perfect separation isn’t always possible, so a soft margin SVM introduces slack variables ξ to permit some misclassifications. These slack variables represent how far each misclassified point is from the margin.

SOFT MARGIN HYPERPLANE

- If we minimize $\sum_i \xi_i$, ξ_i can be computed by

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + b \geq 1 - \xi_i & y_i = 1 \\ \mathbf{w}^T \mathbf{x}_i + b \leq -1 + \xi_i & y_i = -1 \\ \xi_i \geq 0 & \forall i \end{cases}$$

  - ξ_i are "slack variables" in optimization
  - Note that ξ_i=0 if there is no error for $\mathbf{x}_i$
  - ξ_i is an upper bound of the number of errors
- We want to minimize $\frac{1}{2}||\mathbf{w}||_2 + C\sum_{i=1}^{n} \xi_i$
  - *C* : tradeoff parameter between error and margin

SOFT MARGIN HYPERPLANE

- If $0 < \xi_i \leq 1$ then data point falls inside the boundary but on the correct side of the hyperplane.
- If $\xi_i > 1$ then data point falls on the wrong side of the hyperplane.
- If $\xi_i < -1$ then data point falls on the wrong side of the hyperplane.

SOFT MARGIN HYPERPLANE

Regularization Parameter C:
The degree of tolerance for misclassified points is controlled by a hyperparameter C, which is known as the **penalty parameter** or **regularization parameter**. This parameter controls the trade-off between maximizing the margin and minimizing classification errors. A **fixed** value of C is used throughout the training to penalize any misclassified point.

- When C is **high**, the model tries to fit the data with minimal errors, resulting in a smaller margin (more like a hard margin).
- When C is **low**, the model allows more points to fall within or beyond the margin, resulting in a wider margin but potentially more misclassifications.

SOFT MARGIN HYPERPLANE

Objective Function for Fixed Soft Margin:

The goal is to find a hyperplane that maximizes the margin while minimizing misclassification. The objective function with the soft margin is:

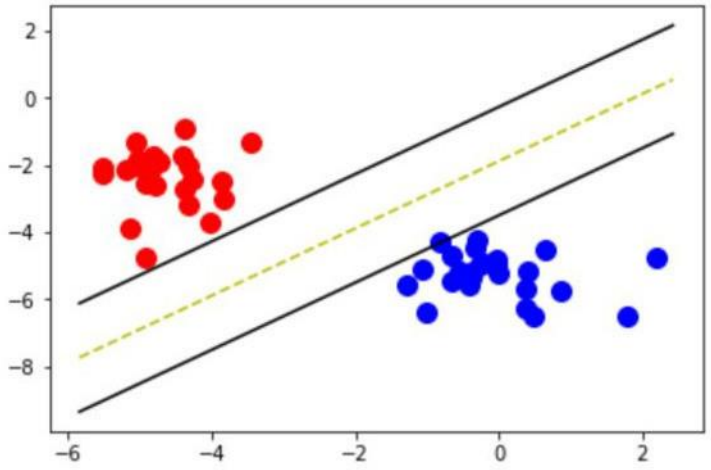
$$\text{Minimize } \frac{1}{2}||\mathbf{w}||_2 + C\sum_{i=1}^{n}\xi_i$$

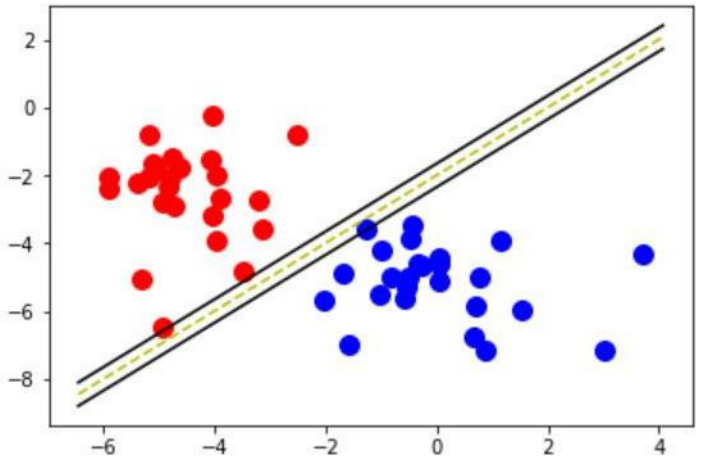
$$\text{subject to } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

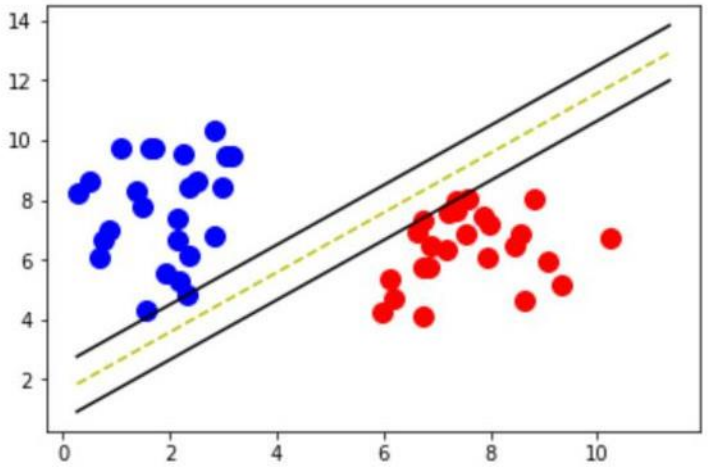
Choosing a Fixed Margin:

For a fixed soft margin, you choose a single value for C that will apply to all points in the dataset. The best choice of C is typically determined through cross-validation.

SOFT MARGIN HYPERPLANE







MULTICLASS CLASSIFICATION

- SVM can solve binary classification problems
- Multiclass classification problems can be broken down into multiple binary classification problems
  - *one-vs-one*
  - *one-vs-rest*
- The number of classifiers *needed to train* *for n number of classes is:*

$$\frac{n * (n - 1)}{2}$$

ADVANTAGES OF SVM

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Can deal with imbalanced data
- It is memory efficient as it uses a subset of training points in the decision function (called support vectors).
- Different Kernel functions can be used input transformation

Thank You ☺