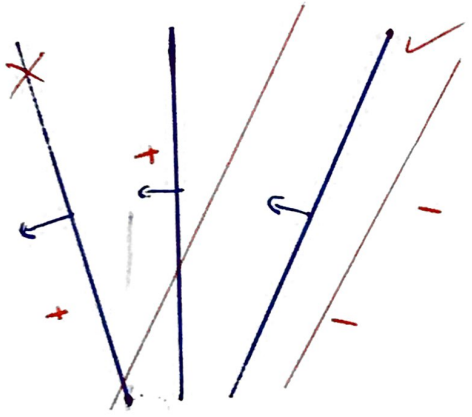
Unit 1 - lecture 3: Hinge loss, Margin boundaries and Regularization

Objectives:

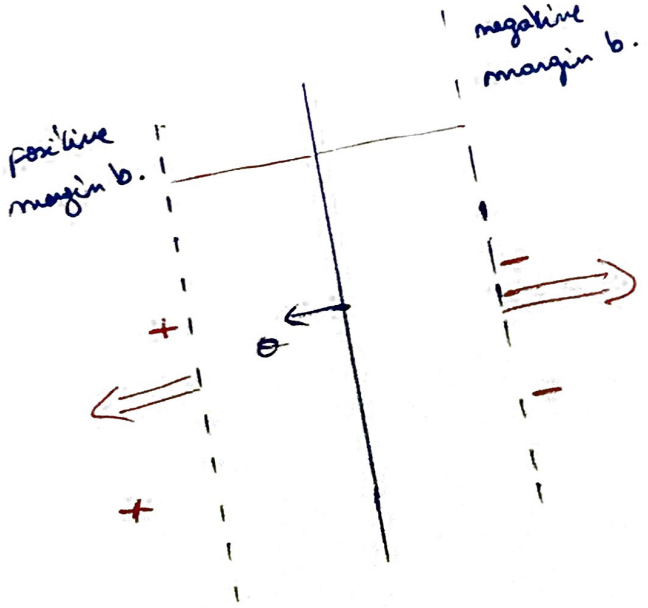
- Understand the need for maximizing the margin
- Pose linear classification as an optimization problem
- Understand Hinge loss, margin boundaries and regularization

Outline:

We are going to turn the problem of finding a linear classifier on the basis of the training set into an optimization problem that can be solved in many ways.



- leaving lots of space on both sides before hitting the training examples is a large margin classifier
- large margin classifier is more robust against noise in the examples.

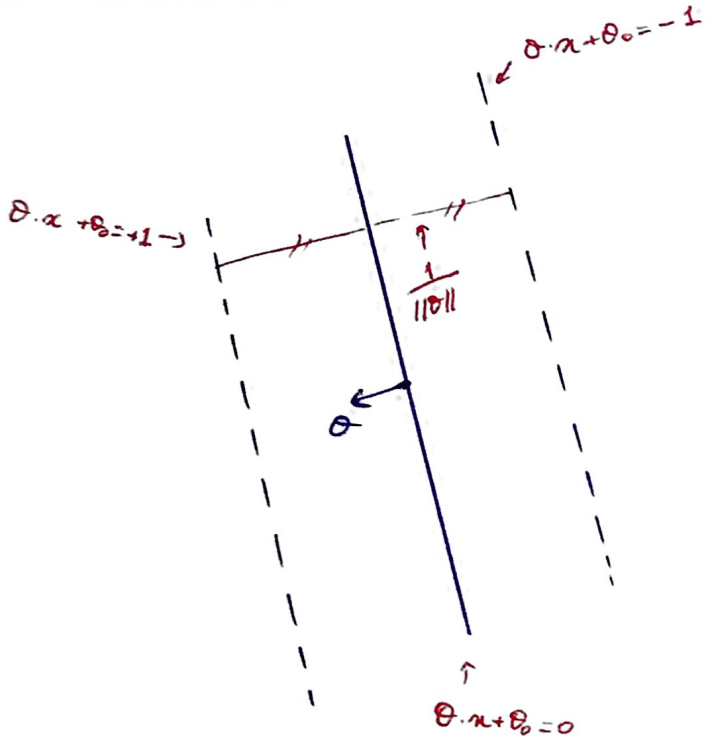


We will push these margin boundaries apart.

There are two aspects to this problem:

1. Favoring margin boundaries that are far apart from their decision boundaries.
   Regularization term.
2. As we are pushing these margin boundaries apart, trying to still fit the decision boundary between the set of examples, we might start violating the preference that all the training examples are outside this fat boundary.
   loss function

Margin boundary:



- We can use the norme of θ ($\|\theta\|$) to push the margin boundaries apart.
- if We increase $\|\theta\|$ ↗, The margin boundaries move closer to the decision boundary
- Our regularization goal is to maximize the distance that the margin boundaries are from the decision boundaries.

$$\max \frac{1}{\|\theta\|}$$

Objective function: To find large decision margin boundaries.

$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{loss}_h(y^i(\theta \cdot x^i + \theta_0)) + \underbrace{\frac{\lambda}{2}\|\theta\|^2}_{\text{Regularization}}$$

What is the loss function? or Hinge loss

$$\text{loss}_1\left(\underbrace{y^i(\theta \cdot x^i + \theta_0)}_{z \text{ agreement}}\right) = \begin{cases} 0 & \text{if } z \geq 1 \\ 1 - z & \text{if } z < 1 \end{cases}$$

measure how much that example violates their margin b. defined

What is the Regularization?

$$\max \frac{1}{\|\theta\|} \iff \min \|\theta\|^2$$

push the margin b. apart