Support-vector Machines

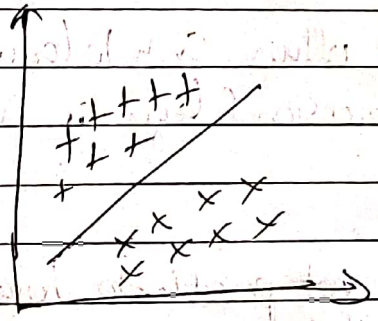
→ SVM was considered the cool-kid before Deep learning entered the game

→ Let's consider our old example classifying cars as cheap & expensive.

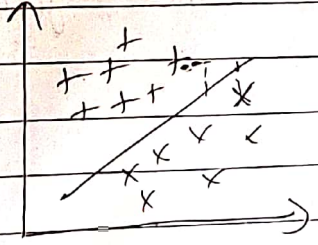
→ Let's plot our data on 2-D graph. which means we are choosing only 2 features. Cost is cheap.

→ Let's consider (+) as Expensive (X) as cheap and the equation would be $w_1x_1 + w_2x_2$.
Here w_1 & w_2 are the weights.

→ If the value of above linear equation is < than 0 then it is cheap and greater than 0 then it is Expensive. So the linear graph would be



→ Confidence of each label can be represented as perpendicular distance between the separation line



→ This minimum value is called Margin, which the smallest distance of a given point to the Seperator line

→ This minimum value is optimal picked by choosing the Seperator that maximizes the margin

→ Given a optimal margin lined seperator, then the points with smallest margin are called Support vectors

→ SVM Cannot only learn from Lined seperable line but also from non-linear seperator.

→ which means more than 2 features can be picked and be represented in a 3-D plane

→ There is one advantage. Data which is not represented properly in 2-D plane. Can be seperated properly in 3D plane

→ A Kernel is which allow SVM to learn transformed features in higher dimensional layer in a much more Computational way.

→ In Some Cases it is not possible to have a Compulsory margin of 1. So in that Cases we allow few points whose margin is less than 1 to be allow and assign a hyperparameter to that part which determines how much our model should focus on minimizing this new term.