CS 419M Introduction to Machine Learning	**Spring 2021-22**

Lecture 7: Binary Classification

Lecturer: Abir De *Scribe: Alakh Agrawal, Devanshu Sarraf, Aditya Kudre*

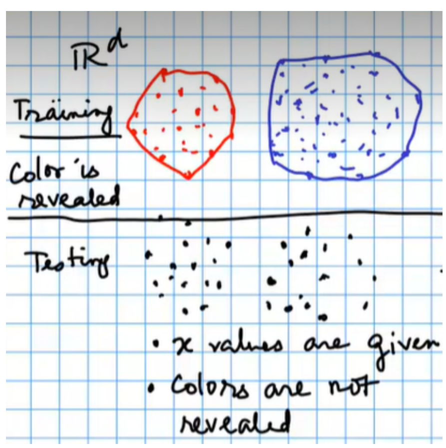
7.1 Classification

We will study the following under classification:

1. What is a classification task?
2. What are the classification models?
3. Depth of support vector machines

7.1.1 Binary Classification Problem

Suppose we are given a set of points and we need to classify them as **Red** or **Blue** based on their x coordinate.



Input: D $= \{(x_i, y_i) \mid y_i \in \{-1, +1\}\}$. **Find:** $m(x) \to y$

Let $P_m(y|x) = \frac{1}{1+e^{-w^\top xy}} \Rightarrow \max_w \prod P_m(y_i|x_i) \to \max_w \sum log(P_m(y_i|x_i)) \to \min_w \sum log(1 + e^{-w^\top xy})$

7.1.2 Trying out a Linear Model

Simplifying this problem as:
if, $w^\top x + b > 0 \Rightarrow y = +1$
else, $w^\top x + b < 0 \Rightarrow y = -1$

This can cause a problem if two points are too close, hence let's take a small Δ instead of 0
if, $w^\top x + b > \Delta \Rightarrow y = +1$
else, $w^\top x + b < -\Delta \Rightarrow y = -1$

This will lead to a problem while classifying the points around the margin and cannot be resolved for however small Δ.

The problem is that when we have points close to the margin, it may be possible that these points are not classified at all as we have a finite non zero Δ.

We can ignore the overlapping points,
Let $I^+ = \{i \mid y_i = 1\}$, $I^- = \{i \mid y_i = -1\}$, $S^+ \in I^+$,$S^- \in I^-$ such that $|S^+ \cup S^-| = n$,

$$\min_{\zeta_i} f(w) - \left(\sum_{i \in S^+} \mathbb{I}\left(w^T x_i + b \geq \Delta\right) + \sum_{i \in S^-} \mathbb{I}\left(w^T x_i + b \leq \Delta\right) \right)$$

7.1.3 A possible solution

We can have an additional loss function along with a usual convex loss function like mean squared error loss that would solve the issue of the above rare case.

This loss function will ensure that the number of points that are not classified or mis-classified is kept to a minimum number.

Thus, we have the following optimization problem:

$$min_{w,b,S,S^+ \bigcup S^- = S}[f(w,b) - \Sigma_{i \in S^+} I(w^T x_i + b > \Delta, y_i = 1) - \Sigma_{i \in S^-} I(w^T x_i + b < -\Delta, y_i = -1)]$$

Here, $f(w,b)$ is our usual loss function while $I(condition)$ is the indicator function and $I(condition) = 1$ if the *condition* is true else $I(condition) = 0$

Alternatively, we can also have the below optimization problem which involves a slightly different version of our additional loss function:

$$min_{w,b,S,S^+ \bigcup S^- = S}[f(w,b) + \Sigma_{i \in S^+} I(w^T x_i + b > \Delta, y_i = -1) + \Sigma_{i \in S^-} I(w^T x_i + b < -\Delta, y_i = 1)]$$

7.1.4 Another possible solution

We could instead modify the model as follows:

$\forall y_i = 1, w^T x_i + b > \Delta - \xi_i$
$\forall y_i = -1, w^T x_i + b < -\Delta + \xi_i$

Here, ξ_i is also a parameter that is to be trained and $\xi_i > 0 \ \forall \ i \in D$. Then, we have the following optimization problem:

$min_{w,b,\xi_i}[f(w,b) + \Sigma_{i \in D} \xi_i]$ such that $y_i(w^T x + b) > \Delta - \xi_i \ \forall \ i \in D$

7.2 Group Details and Individual Contribution

Alakh Agrawal (200040018): Section 7.1, 7.1.1, 7.1.2 (Binary Classification Problem, Trying out a Linear Model)
Devanshu Saraf (200040042): Section 7.1.2 (Binary Classification Problem, Trying out a Linear Model)
Aditya Kudre (200070039): Section 7.1.3, 7.1.4 (A possible solution, Another possible solution)