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

Lecture 7: Binary Classification

Lecturer: Abir De *Scribe: Group 1 and 2*

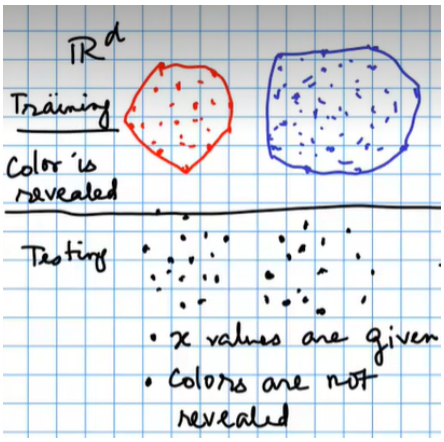
7.1 Classification

We will study the following under classification:

1. What is a classification task?
2. What are the classification models?

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: $\mathrm{D} = \{(x_i, y_i) \mid y_i \in \{-1, +1\}\}$. **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 Visualizing Non-Overlapping and Overlapping Case

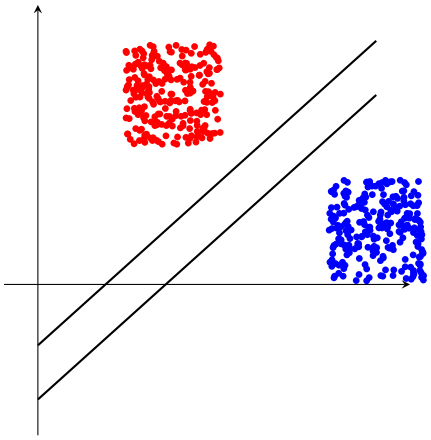


Figure 7.1: Non-Overlapping

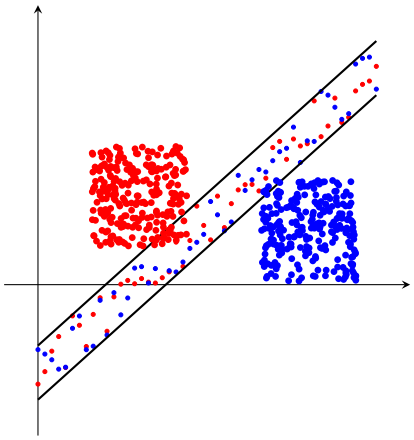


Figure 7.2: Overlapping

7.1.3 A simple Linear Model(Ignoring Overlapping Points)

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 $|S^+ \cup S^-| = n$,

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

7.1.4 Including additional loss function to solve overlapping case.

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:

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

7.1.5 Adding Slack variable to solve the overlapping case

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 Summary

1. We have defined the classification task in section 7.1.1 where the training set has (x_i, y_i) points, and the task of classification model is to predict the label of the test instance (x_j).
2. Section 7.1.2 showed the visual difference between the Non-Overlapping and Overlapping case of classification task.
3. Here in section 7.1.3 we represent a simpler model where we just ignored the overlapping of points.
4. Going further in section 7.1.3 and 7.1.4 we formulate the models considering the non overlapping case.

7.3 Group Details

Group	Name	Roll number
Group 1	Alakh Agrawal	200040018
Group 1	Devanshu Saraf	200040042
Group 1	Aditya Kudre	200070039
Group 2	Dadhichi Telwadkar	20D070083
Group 2	Sameer Parwani	20D180029
Group 2	Pratham Kingar	200040075
Group 2	Samyak Ajmera	19B080023
Group 2	Aakash Jha	200040002