Table of Contents:

Problem Introduction

Consider the problem of predicting how a movie viewer would rate the various movies out there `The Netflix Prize` (https://www.thrillist.com/entertainment/nation/the-netflix-prize).

Given we know what movies a viewer like (comedy, action, romance,..) & what does the movie features (same factors), we can estimate How much did the viewer rate the movie?

Formalization

The learning problem can be formalized as:

Given:

- Input \(x\) (viewer likings)
- Output \(y\) (rating)
- Data: \((x_1, y_1), (x_2, y_2), (x_3, y_3), ...\)

We want to find the

- Hypothesis: \(g: X \mapsto Y\) where \(g \in H \)

Assuming

- Target function \(f: X \mapsto Y\)

The hypothesis set and learning algorithm are informally referred to as the learning model.

So, how do we find the hypothesis ? One way is, Say data is linearly separable.

Perceptron Learning Algorithm

Say we have two classes **Class A** & **Class B**. Then we say

$$x_i \in A \text{ if } \sum_{i=1}^{d} w_i x_i > threshold$$

$$x_i \in B \text{ if } \sum_{i=1}^{d} w_i x_i < threshold$$

Therefore the linear formular \(g \in H\) can be written as,

$$h(x) = sign\left(\left(\sum_{i=1}^{d} w_i x_i\right) - threshold\right)$$

Thus, if \(h(x_i) > 0 \) then \(x_i \in A \) else \(x_i \in B \).

So we have \(x_i\), the training data, we need to find optimal \(w_i\) & \(threshold\) such that we can classify the data as best as we can.

Making some notation changes,

$$h(x) = sign\left(\left(\sum_{i=1}^{d} w_i x_i\right) + x_0 w_0\right)$$

In vector form, the perceptron implements,

$$h(x) = sign(w^T x)$$

So, how do we find w ?

References: