Data Mining II Homework III

Allen Rahrooh

05 February 2020

Exercise 4.6

Suppose we have N points in x_i in $\mathbb{R}^p$ in general position, with class labels $y_i \in \{-1, 1\}$. Prove that the perceptron learning algorithm converges to a separating hyperplane in a finite number of steps:

(a) Denote a hyperplane by $f(x) = \beta_1^T x + \beta_0 = 0$, or in more compact notation $\beta^T x^* = 0$, where $x^* = (x, 1)$ and $\beta = (\beta_1, \beta_0)$. Let $z_i = \frac{x_i^*}{||x_i^*||}$. Show that separability implies the existence of a β_{sep} such that $y_i = \beta_{sep}^T z_i \geq 1 \forall i$

By the definition, if points are separable there exists a β vector given that

$$\beta^T x_i^* > 0 \text{ when } y_i = +1 \text{ and}$$
$$\beta^T x_i^* < 0 \text{ when } y_i = -1$$

for $i = 1, 2, \cdots, N$. Also this is equivalent to $y_i \beta^T x_i^* > 0$.

Next we can divide the expression by $||x_i^*||$ which give us:

$$y_i \beta^T z_i > 0$$

We then let $m > 0$ to be the minimum value of the product for the whole training set. This would give us m as:

$$y_i \beta^T z_i \geq m$$

Then dividing the equation by m gives us:

$$y_i (\frac{1}{m} \beta)^T z_i \geq 1$$

Defining $\beta_{sep} = \frac{1}{m}\beta$ this proves that $y_i \beta_{sep}^T z_i \geq 1$ for all i.

(b) Given a current β_{old}, the perceptron algorithm identifies a point z_i that is misclassified, and produces the update $\beta_{new} \leftarrow \beta_{old} + y_i z_i$. Show that $||\beta_{new} - \beta_{old}||^2 \leq ||\beta_{old} - \beta_{sep}||^2 - 1$, and hence that the algorithm converges to a separating hyperplane in no more than $||\beta_{start} - \beta_{sep}||^2$ steps.

We start out using the following equation $\beta_{new} = \beta_{old} + y_i z_i$ which gives:

$$\beta_{new} - \beta_{sep} = \beta_{old} - \beta_{sep} + y_i zi$$

squaring the equation gives:

$$||\beta_{new} - \beta_{sep}||^2 = ||\beta_{old} - \beta_{sep}||^2 + y_i^2 ||z_i||^2 + 2y_i(\beta_{old} - \beta sep)^2 z_i$$

With $y_i = \underline{+}1$ and $||z_i||^2 = 1 \rightarrow y_i^2||z_i||^2 = 1$ this is for the second term on the right hand side of the equation. The third term is given by

$$2(y_i\beta_{old}^T z_i - y_i\beta_{sep}^T z_i)$$

With $y_i z_i$ we get a misclassified vector $\beta_{old} \rightarrow y_i\beta_{old}^T z_i < 0$. The β_{sep} is a vector that can classify all points $\rightarrow y_i\beta_{sep}^T z_i \geq 1$. With that we can write the equation as:

$$2(y_i\beta_{old}^T z_i - y_i\beta_{sep}^T z_i) \leq 2(0-1) = -2$$

This shows that:

$$||\beta_{new} - \beta_{sep}||^2 \leq ||\beta_{old} - \beta_{sep}||^2 + 1 - 2 = ||\beta_{old} - \beta_{sep}||^2 - 1$$

Exercise 4.7

Consider the criterion

$$D^*(\beta, \beta_0) = -\sum_{i=1}^{N} y_i(x_i^T\beta + \beta_0)$$

a generalization of (4.41) where we sum over all the observations. Consider minimizing D^* subject to $||\beta|| = 1$. Describe the criterion in words. Does it solve the optimal separating hyperplane problem?

With the constraint $||\beta|| = 1$ the equation $x_i^T\beta + \beta_0$ is equal to the signed distance from the point x_i from the given hyperplane ($x^T\beta + \beta_0 = 0$) With the minimization of $D(\beta, \beta_0)$ the signed distance of x_i will take a high positive value when $y_i = +1$ and a very negative value when $y_i = -1$. With these conditions it does create a hyperplane but does not solve the optimal seperating hyperplane problem because the optimal seperating hyperplane has a pointwise constraint with the following equation:

$$y_i(x_i^T\beta + \beta_0) \geq M, i = 1, \cdots, N$$