Assignment 4

Darwin Ding

September 26, 2016

Exercise 2.4

(a) Back in chapter 1, we defined $h(x) = sign(w^T x)$. We dealt with 2d space in chapter 1, but the math still applies to d+1 dimensional space. Our perceptron is a vector of size d+1, which represents a hyperplane in d+1 dimensional space.

The question can be reworded to, given any vector y of d+1 points:

$$y = [\pm 1, \pm 1, \pm 1, ...]$$

... and a d+1 by d+1 matrix X that represents d+1 points with d+1 dimensions each, can we generate a perceptron w (vector of size d+1) such that:

$$y = Xw$$

And in fact, this is quite straightforward to do. It is quite easy to create a d+1 by d+1 non-singular matrix, even if we require all 0th dimensions of our points to be 1 (see Chapter 1).

Simply see the following matrix, where all the lines are linearly independent:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 1 & 1 & 1 & 1 & \dots & 1 \end{bmatrix}$$

Due to its non-singularity, X^{-1} exists and we can simply set $w = X^{-1}y$. Because we can do this with at least one example using a non-singular matrix, we have shattered d+1 for perceptrons. $\boldsymbol{d_{VC} \geq d+1}$

(b) However, things change when you add the next point. When you have more points than dimensions, linear independence is impossible. Mathematically speaking, this means:

$$\begin{aligned} x_{d+2} &= \sum_{i=1}^{d+1} a_i x_i \\ \implies y_{d+2} &= sign(w^T x_{d+2}) \\ &= sign(\sum_{i=1}^{d+1} a_i w^T x_i) \end{aligned}$$

However the fact that we can now mathematically derive the sign of the d+2nd point based off of the other d+1 points shows a glaring hole in the dichotomies that we can implement.

Namely, for each point i from 1 to d+1, let's assign it +1 if $a_i > 0$, and -1 if $a_i < 0$. It doesn't matter so much if $a_i = 0$, but not all a_i can be 0 due to the linear dependence of point d+2.

Simplifying the above, we have $y_i = w^T x_i$, and $a_i w^T x_i = +a_i$ when $a_i > 0$, and $a_i w^T x_i = -a_i$ when $a_i < 0$.

As a result, $\sum_{i=1}^{d+1} a_i w^T x_i$ will always be positive! Therefore, if the d+2nd point is assigned the value -1, the dichotomy cannot be implemented. Therefore, d+2 cannot be shattered, because this process can be done with any set of d+2 points.