

# CS419M (Spring 2018): Assignment 2

## Theory

Dimple Kochar, 16D070010

February 11, 2018

### 1 Problem 1: What's my VC dimension?

$\mathbb{R}^3$  -  $H$  is the set of axis-parallel rectangles. - VC Dimension is 6  
Consider the 6 points  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ ,  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ . If we draw a bounding box for these points, then by excluding/including each point by moving a face of the box, we can get any labeling for the points. So, the VC dimension is at least 6. For 7 points, consider the bounding box. If the bounding rectangle has at least one point in its interior, then we cannot accomplish the labeling where the interior point is labeled - and the rest are labeled +. If none of the points are in the interior, then at least two must be on the same face of the rectangle by the pigeonhole principle and then cannot have opposite labels.

### 2 Problem 2: Perceptron with margin

#### 2.1 A.

The weight update step in the perceptron algorithm, given by  $w^{t+1} \leftarrow w^t + y_i x_i$  where  $x_i$  is a training example in  $\mathbb{R}^d$  with label  $y_i \in \{-1, 1\}$ , is performed when  $y_i(w^t \cdot x_i) \leq 0$ . The "perceptron with margin" algorithm is similar to the perceptron algorithm: The main difference is that a weight update is performed on  $(x_i, y_i)$  when  $y_i(w^t \cdot x_i) < \eta$  where  $\eta > 0$  is a predefined margin parameter. Therefore, initializing  $w_1 = y_1 x$ , where  $x$  is the first example seen and initialize  $t$  to 1, we predict positive if  $\frac{w_t x}{\|w_t\|} \geq \frac{\eta}{2}$  and predict negative

if  $\frac{w_t x}{\|w_t\|} \leq -\frac{\gamma}{2}$  (since  $\eta = \frac{\gamma}{2}$ ) and an example to be a margin mistake when  $\frac{w_t x}{\|w_t\|} \in (-\frac{\gamma}{2}, \frac{\gamma}{2})$ . On a mistake (incorrect prediction or margin mistake), we update as in the standard  $w^{t+1} \leftarrow w^t + y_i x_i; t \leftarrow t + 1$

So, each update increases  $w_t \cdot w$  by at least  $\gamma$ . The increase in  $\|w_t\|$ , since we are updating on some examples where the angle is more than  $90^\circ$ ; For the original algorithm, we had:  $\|w_{t+1}\|^2 \leq \|w_t\|^2 + 1$ . Since  $\sqrt{1+x} \leq 1 + \frac{x}{2}$ , we get, that  $\|w_{t+1}\| \leq \|w_t\| + \frac{1}{2\|w_t\|}$ . For the new algorithm, we instead get  $\|w_{t+1}\| \leq \|w_t\| + \frac{1}{2\|w_t\|} + \frac{\gamma}{2}$  which we can see by breaking each  $x$  into its orthogonal part (for which the original statement holds) and its parallel part (which adds at most  $\gamma/2$  to the length of  $w_t$ ). We can now solve this directly, but just to get a simple upper bound, just notice that if  $\|w_t\| \geq \frac{\gamma}{2}$  then  $\|w_{t+1}\| \leq \|w_t\| + \frac{3\gamma}{4}$ . So, after  $M$  updates we have:  $\|w_{M+1}\| \leq \frac{\gamma}{2} + \frac{3M\gamma}{4}$ . Solving  $M\gamma \leq \frac{\gamma}{2} + \frac{3M\gamma}{4}$  we get  $M \leq \frac{2}{\gamma} + \frac{4}{\gamma}$ .

## 2.2 B.

The upper bound of the perceptron with margin is more than the perceptron algorithm which means that it will take larger number of steps (at max) for it to converge than the standard perceptron. Suppose we are handed a set of examples  $S$  and we want to actually find a large-margin separator for them. However, if we only need to approximately maximize the margin, then another approach is to use Perceptron. In particular, suppose we cycle through the data using the Perceptron algorithm, updating not only on mistakes, but also on examples  $x$  that our current hypothesis gets correct by margin less than  $\gamma/2$ .

## 2.3 C.

$K(x, x')$  such that for some  $\phi : X \rightarrow \mathbb{R}^m$  ( $m$  could even be infinite) we have  $K(x, x') = (1 + x^T \cdot x')^k$  for integer  $k \geq 1$  where  $K(x, x') = \phi(x)^T \phi(x')$