# CSC 482A: Problem set 2: Due by 7:00pm Friday, October 25

Swapnil Daxini (V00861672)

October 26, 2019

1. The proof of this is quite trivial. Given that $\mathcal{F}$ contains a set of all convex polygons, if we want to shatter n points, we can create a convex polygon with its vertices as the n points. This can be done with any number of points (see **Fig. 1**). Thus, given that our class contains all convex polygons, we have that $VCdim(\mathcal{F}) = \infty$.

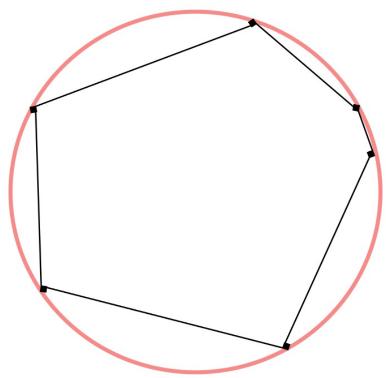

Figure 1: Creating a convex polygon that shatters 6 points. Using this method, it is possible to create a convex polygon that can shatter any number of points.

2. a) For this question, we first realize that our hyperplane has (d+1) degrees of freedom. For our classification, we have that:

$$f(x) = \mathbb{1}[\langle w, x\rangle + b > 0] = sign(\boldsymbol{w}\boldsymbol{x}^T + b)$$

This is because of the d dimensions and the parameter b. In order to show that the $VCdim \geq d+1$, we consider $X \in \mathbb{R}^{d+1 \times d+1}$ given by:

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_{d+1}^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 0 & \cdots & 1 \end{bmatrix}$$

where the offset $b$ is provided by the first term of each $x_i$. Given that $det(X) \neq 0$, we know that $X$ is invertible. If we set that $w = X^{-1}y$ where $y = [y_1, \ldots, y_{d+1}]$, we have that $Xw = y$, thus all the points will be correctly classified. So, we have that $VCdim \geq d+1$.

b) By Radon's theorem, we have that for any set of d + 2 points, we will have two sets A and B such that their convex hull intersect. In this case, if we were to label set A as positive and set B as negative, we could not have a linear separator that can classify this case as we an intersection of the convex hull (see **Fig. 2**). Thus we have that $VCdim < d+2$. Together with part a), this implies that $VCdim = d+1$.

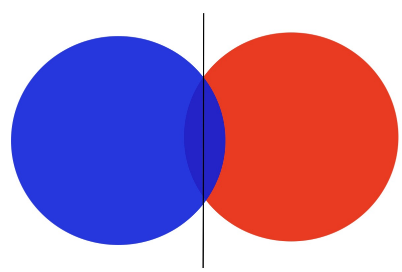

Figure 2: Two sets with overlapping convex hulls can not be partioned by a linear separator

c) We can use the hint provided in the question. We start by proving that for d + 2 points, we have that there exists:

$$\sum_{j=1}^{d+2} = \lambda_j x_j = 0 \tag{1}$$

and

$$\sum_{j=1}^{d+2} = \lambda_j = 0 \tag{2}$$

Consider d + 1 points, such that $x_2 - x_1, x_3 - x_1, \ldots, x_{d+2} - x_1$. Given that these are linearly dependent, we know that:

$$x_2 - x_1 = \sum_{j=3}^{d+2} \alpha_j (x_j - x_1) \tag{3}$$

Simplifying that we have that:

$$\sum_{j=1}^{d+2} = \lambda_j x_j \tag{4}$$

where $\lambda_1 = \alpha_3 + \ldots + \alpha_{d+2} - 1$, $\lambda_2 = 1$ and $\lambda_{3,\cdots,d+2} = -\alpha_{3,\cdots,d+2}$

Thus we now have a form that agrees with our initial statement.

To prove Radon's theorem, we can consider two sets A and B, such that $A = j|\lambda_j > 0$ and $B = j|\lambda_j < 0$

If we define $C = \sum_{j\in A} \lambda_j$, we know that $C = -\sum_{j\in A} \lambda_j$ because of Eq. 2. For a point x in the convex hull of A, given by $\sum_{j\in A} \lambda_j \frac{x_i}{C}$, we know that it will also be in the convex hull of B as we have $\sum_{j\in B} -\lambda_j \frac{x_i}{C}$. This is because $\sum_{j\in A} = \lambda_j x_j + \sum_{j\in B} = \lambda_j x_j = 0$ by Eq. 1. Thus we know that x is in the convex hull of A and B.

3. For this question we need to find $Pr(||X||_0 > \hat{s})$ given X is s-sparse.

   I am unsure about how to do this.

4. For this question, my intuition is to use the bounded differences inequality. In particular, we can use Theorem 1 on Lecture notes 10. I am not entirely sure about this question, but I will make an effort to solve using that.

   The bounded differences inequality states that for 2 points on the Euclidean ball, we have that:

$$\sup |g(x_1, \ldots, x_d) - g(x_1, \ldots, x_i', \ldots, x_d)| \leq c_i \tag{5}$$

   where g is given by:

$$g = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} ||X_i - X_j||_2 \tag{6}$$

   Given that we are considering the distance between points on the unit Euclidean ball, we have that g is equal to the $l_2$ norm. Since we are considering an unit Ball, we have $c_i = \frac{1}{n}$. Thus by theorem 1, we have that:

$$Pr(|g(X_1^n) - E[g(X_1^n)]| \geq t) \leq 2exp(-2nt^2) \tag{7}$$

   from which we get that:

$$|g(X_1^n) - E[g(X_1^n)]| \leq 2\sqrt{\frac{log(1/\delta)}{n}} \tag{8}$$

   I am not sure where to go next. Sorry, this is all I can do. Rough week!