SVM Problem Set

Names:

1. Prove that the *geometric margin* (physical distance between the point and the hyperplane) for a given example $(\vec{x}_i, y_i)$ and a given hyperplane $(\vec{w} \cdot \vec{x} + b = 0)$ is

$$\gamma_i = y_i \left(\frac{\vec{w}}{\|\vec{w}\|} \cdot \vec{x}_i + \frac{b}{\|\vec{w}\|} \right)$$

We did a few of the steps in class – your task here is to explain those steps in your own words (a picture would be very helpful) and fill in the missing steps.

2. The Lagrangian for our SVM problem is

$$\mathcal{L}(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{n} \alpha_i \big[y_i(\vec{w} \cdot \vec{x}_i + b) - 1 \big]$$

After we take the gradient with respect $\vec{w}$ and the derivative with respect to b, we end up with these two equations

$$\vec{w} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i \quad , \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

Use these equations to demonstrate that

$$\mathcal{L}(\vec{w}, b, \vec{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j \vec{x}_i \cdot \vec{x}_j$$

(The RHS is what we redefine as $W(\vec{\alpha})$, since it no longer depends on $\vec{w}$ and b.)

3. Why is the "max of the mins" always less than or equal to the "min of the maxs"? In other words, let $f(x, y)$ be a function of two variables – why is

$$\max_x \ \min_y f(x, y) \leq \min_y \ \max_x f(x, y) \ ?$$

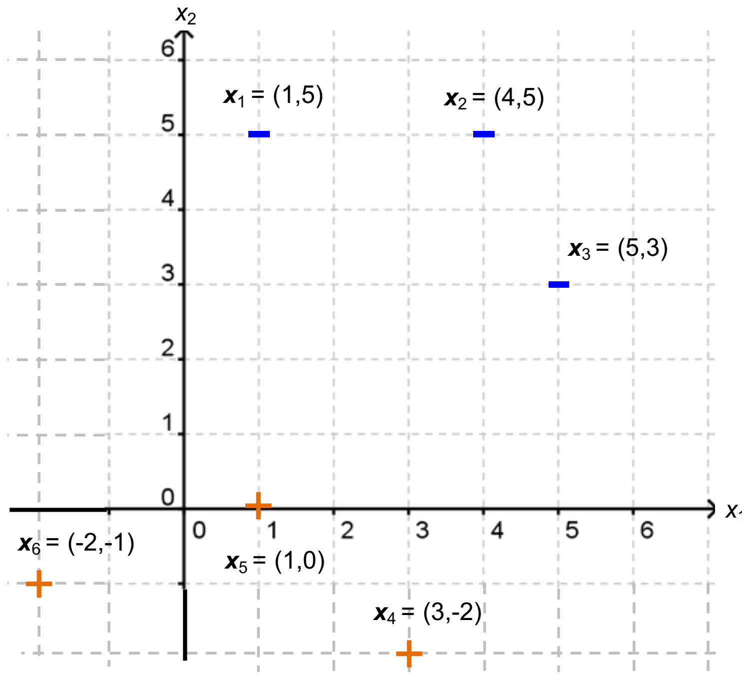
Hint: imagine a matrix of $f(x, y)$ values, with x denoting the row and y denoting the column. Think about first fixing the row and finding the min over all y values in that row. Then do that for each row and take the max. Then reverse the steps.

4. *Challenge* (not optional but worth few points): After finding the optimal alpha values α_i^* (for $i = 1, \cdots, n$), we can find the optimal weight vector $\vec{w}^*$. Demonstrate that the optimal bias is

$$b^* = -\frac{1}{2} \left(\max_{i:y_i=-1} \vec{w}^* \cdot \vec{x}_i + \min_{i:y_i=1} \vec{w}^* \cdot \vec{x}_i \right)$$

Hint: think about the constraints in our optimization problem after we forced $\hat{\gamma} = 1$. The bias only shifts our hyperplane up and down (doesn't change its direction), so we want to move it to the "middle" of the support vectors from each class.

5. The goal of this question is to gain some intuition about SVMs through a concrete example. Consider the 6 examples shown below from two classes (+1, −1).



(a) Using your knowledge of what the maximum margin hyperplane means, draw the "best" hyperplane on the figure above.

(b) What is an *equation* (in terms of the axes x_1 and x_2) that describes this optimal hyperplane?

(c) Which of the $\vec{x}_i$'s are support vectors?

(d) Given that the weight vector is orthogonal to the hyperplane, write $\vec{w}$ as

$$\vec{w} = a \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

for some concrete w_1 and w_2 and some unknown constant a. In other words, find the direction of the weight vector.

(e) Using the fact that the functional margin $\hat{\gamma} = 1$ (i.e. $\hat{\gamma}_i = 1$ for support vectors), solve for this constant a and the bias b. At this end of this part you should have numerical values for the optimal weight vector $\vec{w}^*$ and the optimal bias b^*.

(f) Finally, using the fact that $\vec{w} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$, find the optimal α_i^* for $i = 1, \cdots, 6$.