Problem Set 3

June 22, 2022

DUE: Thursday June 30 @ 23:59 EST
Written Problems (100pts)

1. (20pts) Consider a perceptron that uses the perceptron learning rule:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta \nabla E_P(\mathbf{w}) = \mathbf{w}^{(t)} + \eta \sum_{i \in \mathcal{M}} \phi(\mathbf{x}^{(i)}) y_{gt}^{(i)}$$

   where $\mathcal{M}$ denotes all incorrectly classified points and t indexes the number of times the parameters $\mathbf{w}$ have been updated.

   (a) Show that the learned weights $\mathbf{w}$ can be written as a linear combination of the vectors $\phi(\mathbf{x}^{(i)}) y_{gt}^{(i)}$ where $y_{gt}^{(i)} \in \{-1, 1\}$. Denote the coefficient of the term $\phi(\mathbf{x}^{(i)}) y_{gt}^{(i)}$ in the linear combination with $\alpha^{(i)}$.

   (b) Using $\alpha^{(i)}$, express the perceptron learning rule as well as the predictive function

$$\hat{y}^{(i)} = sign(\mathbf{w}^T \phi(\mathbf{x}^{(i)}))$$

   (c) Show that the feature vector $\phi(\mathbf{x}^{(i)}$ only appears in the form of a kernel $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}^{(i)})^T \phi(\mathbf{x}^{(i)})$

2. (20 points) Show that from a dataset containing only two points, one from each class, that regardless of the dimensionality of the data, we can learn maximum-margin hyperplane.

3. (20 points) Show that the vlaue ρ of the margin for the maximum-margin hyperplane is given by the following equation:

$$\frac{1}{\rho^2} = \sum_i \alpha^{(i)}$$

   where $\alpha^{(i)}$ are given from maximizing

$$\tilde{L}(\alpha) = \sum_i \alpha^{(i)} - \frac{1}{2} \sum_{i,j} \alpha^{(i)} \alpha^{(j)} y_{gt}^{(i)} y_{gt}^{(j)} k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$$

subject to

$$\begin{aligned} \alpha^{(i)} &\geq 0 && i = 1, 2, \ldots, N \\ \sum_i \alpha^{(i)} y_{gt}^{(i)} &= 0 \end{aligned}$$

4. Using the setup from the previous problem, show that:

   (a) $\frac{1}{\rho^2} = 2\tilde{L}(\alpha)$

   (b) $\frac{1}{\rho^2} = ||\mathbf{w}||^2$