CS 189: Introduction to Machine Learning - Discussion 8

1. Midterm Review

   a) We have trained an SVM with a Gaussian kernel:

   $$K(\mathbf{u}, \mathbf{v}) = e^{\frac{-(\mathbf{u}-\mathbf{v})^2}{2\sigma^2}}$$

   Now we have a set of n support vectors (the training points the SVM keeps) $\{\mathbf{x}^{(i)}\}$, the associated training labels $\{y^{(i)}\}$ and alpha weights $\{\alpha_i\}$.
   How do we classify a new test point $\mathbf{x}$?

   b) What's the difference between perceptron and Hebb's rule?

   c) What's the difference between the classifier given by an SVM or perceptron algorithm, and logistic regression?

   d) What's the difference between generative models and discriminative models?

   e) What is the difference between LDA and PCA?

2. *Optional: Extra for Experts!* Curse of Dimensionality

We have a training set: $(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(n)}, y^{(n)})$, $\mathbf{x}^{(1)} \in \mathbb{R}^d$. Our 1-nearest neighbor classifier is:

$$\text{class}(\mathbf{x}) = y^{(i^*)} \quad \text{where } \mathbf{x}^{(i^*)} \text{ is the nearest neighbor of } \mathbf{x}.$$

Assume any data point $\mathbf{x}$ is inside the Euclidean ball of radius 1, i.e. $\|\mathbf{x}\|_2 \leq 1$. To be confident in our prediction, we want the distance between $\mathbf{x}$ and its nearest neighbor to be small, within some positive ϵ:

$$\|\mathbf{x} - \mathbf{x}^{(i^*)}\|_2 \leq \epsilon \quad \text{for all } \|\mathbf{x}\|_2 \leq 1. \tag{1}$$

For this condition hold, at least how many data points should be in the training set? How does this lower bound depend on the dimension d?

Hint: Think about the volumes of the hyperspheres, and use the union bound:

$$\text{vol}(\cup_{j=1}^{k} S_j) \leq \sum_{j=1}^{k} \text{vol}(S_j), \text{ where } S_j \text{ is a hypersphere.}$$