CS 189: Introduction to Machine Learning - Discussion 7

1. Performance evaluation

   Suppose that we use some learning method to make a prediction y for a particular data point $\mathbf{x}$. Assume we are given a training set and validation set.

   1. Describe how we might estimate the standard deviation of our validation risk using bootstrap samples.
   2. Describe how we might estimate the standard deviation of our error rate by assuming errors are drawn from a binomial distribution.

2. Logistic Posterior with different variances

   We have seen that Gaussian class conditionals with the same variance lead to a linear decision boundary. Now we will consider the case where class conditionals are Gaussian but have different variances, i.e

$$\begin{aligned} X|Y = i &\sim \mathcal{N}(\mu_i, \sigma_i^2), \quad \text{where } i \in \{0, 1\} \\ Y &\sim \text{Bernoulli}(\pi) \end{aligned}$$

   Show that the posterior distribution of the class label given X is also a logistic function, however with a quadratic argument in X. Assuming 0-1 loss, what will the decision boundary look like (i.e., describe what the posterior probability plot looks like)?

3. Linear Regression with Laplace prior

   We saw in discussion 4 that there is a probabilistic interpretation of linear regression: $P(y|\mathbf{x}, \sigma^2) \sim \mathcal{N}(\mathbf{w^T x}, \sigma^2)$. We extend this by assuming some prior distribution on parameters $\mathbf{w}$. Let us assume the prior is a Laplace distribution, so we have:

$$w_j \sim Laplace(0, t), \text{ i.e. } P(w_j) = \frac{1}{2t} e^{-|w_j|/t} \text{ and } P(\mathbf{w}) = \prod_{j=1}^{D} P(w_j) = (\frac{1}{2t})^D \cdot e^{-\frac{\sum |w_j|}{t}}$$

   Show it is equivalent to minimizing the following risk function, and find the value of the constant λ:

$$R(\mathbf{w}) = \sum_{i=1}^{n} (y^{(i)} - \mathbf{w}^T \mathbf{x}^{(i)})^2 + \lambda \|\mathbf{w}\|_1, \text{ where } \|\mathbf{w}\|_1 = \sum_{j=1}^{D} |w_j|$$