ENGN 2520 Exam

Problem 1

Give short answers:

(a) What is the difference between the training error and test error?

(b) What is overfitting? How can we avoid it in practice?

(c) What is the margin in a support vector machine?

(d) What is a naive Bayes classifier?

(e) Give an example of a classification problem in $\mathbb{R}^2$ that *cannot* be solved by a linear-treshold function in the original input space.

Problem 2

Suppose we have a classification problem where the input space is $X = \mathbb{R}$ and the output space is $Y = \{0, 1\}$.

Let $p(x, y)$ be a joint density defined by,

$$p(y = 0) = \frac{1}{2} \qquad p(y = 1) = \frac{1}{2}$$

$p(x|y = 0)$ is a Gaussian with $\mu = 0$ and $\sigma = 1$.
$p(x|y = 1)$ is a Gaussian with $\mu = 0$ and $\sigma = 2$.

Derive the Bayes optimal classifier in this setting. You should use a 0/1 loss to derive the optimal classification rule. What are the decision regions and decision boundaries?

Problem 3

Let $\{(x_1, y_1), \ldots (x_n, y_n)\}$ be n labeled training examples with $y_i \in \mathbb{R}$. Let ϕ be a (fixed) feature map. Consider the regularized regression problem where we want to minimize

$$E(w) = \left(\sum_{i=1}^{n} (w^T \phi(x_i) - y_i)^2 \right) + \frac{\lambda}{2} ||w||^2$$

Here λ is a parameter that controls the regularization strength.

(a) Why is regularization useful and how does the value of λ affect the solution to the regression problem?

Although this problem can be solved using a linear system, it is also possible to minimize $E(w)$ using gradient descent.

(b) Give a formula for $\nabla E(w)$.

(c) Explain how gradient descent can be used to minimize $E(w)$. You should give a clear description of the resulting algorithm with pseudo-code.

Problem 4

There are two kinds of fish in a river: salmon and bass. Let f_S and f_B be the fraction of fish in the river that are salmon and bass respectively ($f_S + f_B = 1$). Our goal is to estimate f_S and f_B by measuring the lengths of random fish from the river.

You can assume that if we pick a random fish from the river it will be a salmon with probability f_S and a bass with probability f_B.

The length of a random salmon has probability density $p_S(x)$. The length of a random bass has probability density $p_B(x)$. For this problem you can assume p_S and p_B are known.

Note that p_S and p_B will have significant overlap so it is difficult to accurately classify an individual fish.

(a) Consider the length x of random fish from the river. What is the density of x?

(b) Suppose we record the length of n random fish from the river. How can we use the recorded lengths $x_1, \ldots, x_n$ to estimate f_S and f_B *without* classifying the individual fish?