ENGN 2520 practice problems

Problem 1

Given short answers for the questions below.

(a) Suppose we want to use regression to fit a polynomial function to some training data T. How can we pick the degree of the polynomial to use in the regression?

(b) What is an advantage of a multilayer neural network over a linear classifier?

(c) What is an advantage of a linear classifier over a multilayer neural network?

(d) Consider a binary classification problem with a set of training examples T. Let H_1 and H_2 be two sets of classifers and suppose the VC dimension of H_1 is smaller than the VC dimension of H_2. Suppose we find classifiers $c_1 \in H_1$ and $c_2 \in H_2$ that both correctly classify all examples in T. How should we pick between c_1 and c_2? Give a brief justification.

Problem 2

We want to classify fish into two types: bass and salmon. We have 2 real valued features: length and weight.

Suppose we assume the length and weight of a fish are independent conditional on the fish type, and each is distributed according to a Gaussian distribution with mean and variance that depends on the fish type.

Consider the distributions,

- $p(length|type = bass)$,
- $p(weight|type = bass)$,
- $p(length|type = salmon)$,
- $p(weight|type = salmon)$,
- $p(type = salmon)$,
- $p(type = bass)$,

(a) What is the classification rule defined by the Bayes optimal classifier in terms of the six distributions specified above. You should use a 0/1 loss to derive the classifier.

(b) Let T be a training set with n labeled examples. For $1 \leq i \leq n$ example i has length l_i, weight w_i and is of type y_i. What are the maximum likelihood estimates for the six distributions above?

(c) Suppose the resulting classifier has poor performance on a test set. Give 3 different reasons why this might happen in practice.

Problem 3

Consider a binary classification problem in which the training examples are points in $\mathbb{R}^2$. The positive examples are $(1, 1)$ and $(-1, -1)$. The negative examples are $(1, -1)$ and $(-1, 1)$.

(a) Are the training examples linearly separable in the original input space?

(b) Consider a feature map $\phi(x) = (1, x_1, x_2, x_1 x_2)$, where $x = (x_1, x_2)$ is an input in $\mathbb{R}^2$. Consider classifiers defined by thresholding $w^T \phi(x)$ at 0. Give a value for w that correctly classifies all training examples with no examples in the classification boundary.

(c) What is the kernel $K(x, z)$ that corresponds to the feature map ϕ defined above?

Problem 4

Suppose you have an instrument that measures the speed of a running animal. You install the instrument in a habitat that contains two kinds of animals: tigers and cheetahs. The speed of a tiger is a random variable with probability density $p_1(x)$. The speed of a cheetah is a random variable with probability density $p_2(x)$. Suppose p_1 and p_2 are known. These densities are different but they have significant overlap, so it is hard to decide if an animal is a tiger or a cheetah based on a single observation of its speed.

(a) Let x be the speed of a random animal in the habitat. Let w_1 be the fraction of animals in the habitat that are tigers, and w_2 be the fraction of animals in the habitat that are cheetahs (since there are no other animals in this habitat we have $w_1 + w_2 = 1$). What is the probability density of x?

(b) Suppose we record the speed of n animals $x_1, \ldots, x_n$. Assume each measurement comes from a random animal in the habitat. How can we estimate w_1 and w_2 from these measurements? Give an algorithm for estimating w_1 and w_2. Justify your answer.