E1 213 Pattern Recognition and Neural networks

Problem Sheet 5

1. Given two sets of points in $\Re^d$, show that they are linearly separable if and only if their convex hulls do not intersect. (Given a set of points, $x_1, \cdots, x_n$, their convex hull is the set of all points z which can be written as $z = \sum_{j=1}^{n} \lambda_j x_j$ where $\lambda_j \geq 0$ and $\sum_j \lambda_j = 1$).

2. Consider the joint density of X, Y given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2\sigma^2(1-\rho^2)}(x^2 + y^2 - 2\rho xy)\right), \quad -\infty < x, y < \infty$$

   Here X, Y are jointly Gaussian with mean zero, variance σ^2 and correlation coefficient ρ. From the above, find the marginal density of X and show that it is Gaussian with mean zero and variance σ^2. Now show that the conditional density $f_{Y|X}$ is also Gaussian. From this, show that $E[Y \mid X]$ is a linear function of X.

3. Suppose we have $y = \mathbf{a}^T X + \xi$ where ξ is a zero-mean random variable with variance σ^2. Under this model we have calculated in the class the variance of the least squares solution, W^*. Calculate the expected value of the least squares solution.

4. We can pose the problem of learning a linear classifier as minimizing

$$J(W) = \sum_{i=1}^{n} L(W^T X_i, y_i)$$

   where L is a loss function. For least squares criterion, we take $L(a, b) = (a - b)^2$. If, instead we want to minimize absolute value of error, we can take $L(a, b) = |a - b|$. Show that logistic regression (in the 2-class case) can also be put in this framework with $L(W^T X, y) = \ln(1 + \exp(-yW^T X))$, where we assume that the class labels are $+1$ and -1. What would be the loss function corresponding to mult-class logistic regression?

5. Consider a classification problem with K classes: $C_1, \cdots, C_K$. We say that the training set is linearly separable if there are K functions: $g_j(X) = W_j^T X + w_{j0}$, $j = 1, \cdots, K$, such that we have $g_i(X) \geq$

$g_j(X), \forall j$, whenever $X \in C_i$. We say that a set of examples is totally linearly separable if given any C_i, there is a hyperplane that separates examples of C_i from the set of examples of all other claases. Show that totally linearly separable implies linearly separable but the converse need not be true.

6. Let $\mathcal{X} = \Re^2$, $\mathcal{Y} = \{0,\ 1\}$. Let c^r denote the circle (or circular disc) of radius r with center at origin. That is $c^r = \{(x, y) \in \Re^2 : x^2+y^2 \leq r^2\}$. Let $\mathcal{C}$ be family of such concentric circles with centre at origin. Show that this concept class is PAC learnable.