Tutorial & Practical 11: SVM and ANNs

Question 1

Assume a given data set of feature vectors $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, ..., N$ with corresponding label values $t \in \{-1, +1\}$. Within each class, we further assume that the density of the feature vector is modeled using a kernel density estimator with kernel $k(\mathbf{x}, \mathbf{x}')$.

1. Provide the form of the estimated probability density of the feature vector $\mathbf{x}$ in each class.
2. Derive the minimum misclassification-rate decision rule assuming the two classes have equal prior probability
3. Show that, if the kernel is chosen to be $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$, then the classification rule reduces to simply assigning a new input feature vector to the class having the largest mean
4. Show that, if the kernel takes the form $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^\top \phi(\mathbf{x}')$, that the classification is based on the closest mean in the feature space $\phi(\mathbf{x})$

Question 2

Show that the value $M = \frac{d}{2}$ of the margin for the maximum margin hyperplane is given by

$$\frac{1}{M^2} = \sum_{i=1}^{n} \mu_i$$

where $\{\mu_i\}$, $i = 1, ..., n$ are given maximizing

$$L_D = \sum_{i=1}^{n} \mu_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_i \mu_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j)$$

subject to

$$\mu_i \geq 0, \quad i = 1, ..., n, \quad \text{and} \quad \sum_{i=1}^{n} \mu_i y_i = 0.$$

Question 3

1. Find the relation between the hyperbolic tangent function

$$f(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$$

and the logistic or sigmoid function

$$g(u) = \frac{1}{1 + e^{-u}}$$

2. Consider a two layer neural network function in which $g(.)$ is used for the hidden unit nonlinear activation functions. Using the relation between $f(.)$ and $g(.)$ show that there exits an equivalent neural network which compute the same function but where $f(.)$ is used for hidden unit activation functions.

Question 4

Consider a single hidden layer neural network function for quantitative output with identity output function $g_k(t) = t$ and logistic sigmoid function for the hidden units. Assume the weights β_m from the input to the hidden units are nearly zero. Show that the resulting model is nearly linear in the inputs.