CSE 566 Midterm

Olcay Taner YILDIZ

I. Question

In K-fold cross-validation, the dataset X is divided randomly into K equal sized parts, $X_i, i = 1, 2, \ldots, K$. To generate each train and test set pair, we keep one of the K parts out as the test set, and combine the remaining $K-1$ parts to form the training set. On the other hand, classes must be represented in the right proportions not to disturb the class prior probabilities (**stratification**). Propose an algorithm to do K-fold crossvalidation with stratification.

II. Question

Consider a learning problem where each instance is described by a conjunction of n boolean attributes $a_1, a_2, \ldots, a_n$. Thus, a typical instance would be

$$(a_1 = T) \wedge (a_2 = F) \wedge \ldots \wedge (a_n = T) \quad (1)$$

Now consider a hypothesis space H in which each hypothesis is a disjunction of constraints over these attributes. For example, a typical hypothesis would be

$$(a_1 = T) \vee (a_5 = F) \vee (a_7 = T) \quad (2)$$

Propose an algorithm that accepts a sequence of training examples and outputs a consistent hypothesis if one exists. Your algorithm should run in time that is polynomial in n and in the number of training examples.

III. Question

Show the steps taken by the decision tree algorithm according to the following training examples.

Classification	a_1	a_2
+	T	T
+	T	T
-	T	F
+	F	F
-	F	T
-	F	T

IV. Question

Consider the alternative error function for one hidden layer multliayer perceptron:

$$E = \frac{1}{2}\sum_{i=1}^{N}(r_i - o_i)^2 + \alpha\sum_{i,j} w_{ij}^2 + \beta\sum_i v_i^2 \quad (3)$$

where N represents the number of instances, r_i represents the actual output, o_i represents the calculated output, w_{ij} are the weights between the inputs and the hidden units, v_i are the weights between the hidden units and the output unit. Derive the gradient descent update rules for the weights.

V. Question

Consider applying Genetic Algorithm to the task of finding an appropriate set of weights for an artificial neural network with one hidden layer. Describe an encoding of network weights as a string, and describe an appropriate set of crossover and mutation operators.

VI. Question

Let x have an exponential density, for all x$\geq$0

$$p(x|\theta) = \theta e^{-\theta x}, \quad (4)$$

Suppose that n samples $x_1, x_2, \ldots, x_n$ are drawn independently according to $p(x|\theta)$. Show that the maximum likelihood estimate for θ is given by

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k} \quad (5)$$