CSE 566 Final

Olcay Taner YILDIZ

I. Question

Given the normalization condition for the binomial distribution,

$$\sum_{m=0}^{N} \binom{N}{m} \mu^m (1-\mu)^{N-m} = 1 \quad (1)$$

show that the mean of the binomial distribution is given by $N\mu$. To do this, differentiate both sides of the normalization condition with respect to μ and then rearrange to obtain an expression for the mean. Similarly, by differentiating normalization condition twice with respect to μ and making use of the result for the mean, prove that the variance of the binomial is $N\mu(1-\mu)$.

$$E[m] = \sum_{m=0}^{N} m \binom{N}{m} \mu^m (1-\mu)^{N-m}$$

$$var[m] = \sum_{m=0}^{N} (m - E[m])^2 \binom{N}{m} \mu^m (1-\mu)^{N-m}$$

II. Question

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

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

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.

III. Question

Given three normal distributions

$$p(x|C_1) \sim \mathcal{N}(1, 1)$$
$$p(x|C_2) \sim \mathcal{N}(2, 4)$$
$$p(x|C_3) \sim \mathcal{N}(3, 1)$$

and the priors

$$P(C_1) = \frac{1}{2}$$
$$P(C_2) = \frac{1}{4}$$
$$P(C_3) = \frac{1}{4}$$

calculate the Bayes' discriminant points analytically.

IV. Question

Given the following error rates for a four feature problem, what will be the result of the following attribute subset selection procedures?

- Stepwise forward selection
- Stepwise backward elimination

Error Rates: (F_1), 0.23; (F_2), 0.25; (F_3), 0.36; (F_4), 0.24; (F_1, F_2), 0.18; (F_1, F_3), 0.19; (F_1, F_4), 0.20; (F_2, F_3), 0.16; (F_2, F_4), 0.29; (F_3, F_4), 0.22; (F_1, F_2, F_3), 0.39; (F_1, F_2, F_4), 0.45; (F_1, F_3, F_4), 0.24; (F_2, F_3, F_4), 0.36; (F_1, F_2, F_3, F_4), 0.59.

V. Question

Let us say we have L classification algorithms. How can we order these L from best to worst? (Hint: Use tests comparing two classification algorithms)

VI. Question

Suppose a dishonest dealer has two coins, one fair and one biased; the biased coin has heads probability 1/4. Assume that the dealer never switches the coins. Which coin is more likely to have generated the sequence HTTTHHHTTTTHTHHTT? ($\log_2(3) = 1.585$).