CSE 566 Final

Olcay Taner YILDIZ

I. Question

Suppose that the data mining task is to cluster the following eight points (with (x, y) representing location) into three clusters. A_1(2, 10), A_2(2, 5), A_3(8,4), A_4(5, 8), A_5(7,5), A_6(6,4), A_7(1,2), A_8(4,9). The distance function is Euclidian distance. Suppose initially we assign A_1, A_4, and A_7 as the center of each cluster, respectively. Run the k-means algorithm one epoch on this data.

II. Question

Given two normal distributions $p(x|C_1) \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $p(x|C_2) \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and $P(C_1)$, $P(C_2)$, calculate the Bayes' discriminant points analytically.

III. Question

Let say we have two variables x_1 and x_2 and we want to make a quadratic fit using them, namely

$$f(x_1, x_2) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 (x_1)^2 + w_5 (x_2)^2$$

Given sample data X, how can we find optimum w_i's?

IV. Question

For a numeric input, instead of a binary split, one can use a ternary split with two thresholds and generate three possible intervals as

$$x_j < w_1, w_1 \leq x_j < w_2, x_j \geq w_2$$

Propose a modification of the rule induction algorithm to learn two thresholds, w_1, w_2.

V. Question

Given the normalization condition for the binomial distribution,

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

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 $N\mu(1-\mu)$.

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

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

VI. Question

$$\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u} du$$

Using integration by parts, prove the relation $\Gamma(x+1) = x\Gamma(x)$. Show also that $\Gamma(1) = 1$ and hence that $\Gamma(x+1) = x!$ when x is an integer.