CSE 347/447 Data Mining: Homework 4 Solution

Q1: GINI, Entropy, and Misclassification Error

Answer:

Q2: Bayes Classifier

Answer:
The main difference in the first case is that $P(X|C=i)$ is proportional to the new value of the probability density in the Bayes rule. In the case where μ_i is unknown, maximum likelihood methods need to be applied to estimate μ_i for each class. In this case, it can be shown that the j-th component of μ_i is estimated to be the median of the j-th dimension of the training instances in class i. The reason for using the median instead of the mean is that the Manhattan distance is used instead of the Euclidean distance.

Q3: Cost Matrix

Answer:
(a) The total cost is $F = p(a+d) + q(b+c)$. Since $acc = a+d$, where $N = a+b+c+d$, we may write $F = N[acc(p-q)+q]$. Because $p-q$ is negative, minimizing the total cost is equivalent to maximizing accuracy.

(b) Following Equation (5.83) in TSK, one can derive the decision rule as:

$$
\begin{aligned}
p(+|t) &> \frac{\beta C(-,+) - \beta C(-,-)}{\beta C(-,+) - \beta C(-,-) + \beta C(+,-) - \beta C(+,+)} \\
&= \frac{C(-,+) - C(-,-)}{C(-,+) - C(-,-) + C(+,-) - C(+,+)}
\end{aligned}
$$

which is irrelevant to β.

(c) Following Equation (5.83) in TSK, one can derive the decision rule as:

$$
\begin{aligned}
p(+|t) &> \frac{\beta + C(-,+) - \beta - C(-,-)}{\beta + C(-,+) - \beta - C(-,-) + \beta + C(+,-) - \beta - C(+,+)} \\
&= \frac{C(-,+) - C(-,-)}{C(-,+) - C(-,-) + C(+,-) - C(+,+)}
\end{aligned}
$$

which is irrelevant to β.

Q4: SVM

Answer:

Given a data set of two data points, x_1 ($y_1 = +1$) and x_2 ($y_2 = -1$), the maximum margin hyperplane is determined by solving the SVM problem subject to the constraints $W^T x_1 + b = +1$ and $W^T x_2 + b = -1$. We do this by introducing Lagrange multipliers a_1, a_2, and solving $\arg\min_{w,b} \frac{1}{2} W^T W - a_1(w^T x_1 + b - 1) - a_2(w^T x_2 + b + 1)$.

Taking the derivative of this w.r.t. W and b and setting the results to zero, we obtain $W = a_1 x_1 + a_2 x_2$ and $-a_1 - a_2 = 0$. Let $a_1 = \lambda$, we have $W = \lambda x_1 - \lambda x_2$. Then, $b = -W^T(x_1 + x_2) = \lambda(x_2^T x_2 - x_1^T x_1)$.

Note that the Lagrange multiplier λ remains undetermined, which reflects the inherent indeterminacy in the magnitude of W and b.