Bayes Decision Theory - Discrete Features

Bayes decision theory- discrete features

- Until now, we assumed that the feature vector X could be any point in a d-dimensional Euclidean space, $\mathbf{R}^d$.
- However, in many practical applications, the components of X (i.e., features) could be binary, ternary, etc. That is, the range of values a feature can assume is a discrete set.
- In these cases, $\int p(X|\omega_j)dx$ needs to be replaced with $\sum P(X|\omega_j)$.

Posteriori

- The posteriori probability when the given class is ω_j is:

$$P(\omega_j|X) = \frac{P(X|\omega_j)P(\omega_j)}{P(X)}$$

where

$$P(X) = \sum_{j=1}^{c} P(X|\omega_j)P(\omega_j)$$

Conditional Risk

- The definition of conditional risk $R(\alpha|X)$ is unchanged, and the fundamental Bayes decision rule remains the same.
- To minimize the overall risk, select the action α_i for which $R(\alpha_i|X)$ is minimum, i.e.,

$$\alpha^* = \text{argmin}_i \; R(\alpha_i|X)$$

Naive Bayes Classifier

- Assumption: features are independent of each other.
- Let $X = (x_1, \ldots, x_d)^t$, then $P(X) = P(x_1) \times \ldots \times P(x_d)$.
- Generally this is applied with discrete features (If it is not discrete, then it is discretized first).
- $P(x_i)$ can be found as a frequency ratio from the training set.