Northeastern University June 9, 2022
College of Engineering In Class Problem 9
Group member name(s): Yanming Liu
Group member UID(s): 002199402

IE 7374: Machine Learning

We have learned that for Gaussian Naive Bayes each feature is generated from $x_d \simeq N(\mu_{kj}, \sigma_{kj}^2)$, calculate μ_{kj}

Answer to Question

$D = \{(x_1, y_1), \dots, (x_n, y_n)\}$, iid.

$x_i \in \mathbb{R}^d$, $x^k \sim N(\mu_{kj}, \sigma_{kj}^2)$, then generative modelling

$$\begin{aligned}
P(D) &= P(y_1, \dots, y_n | x_1, \dots, x_n) \\
&= \prod_{i=1}^{n} P(y_i | x_1, \dots, x_n) \\
&= \prod_{i=1}^{n} P(x_1, \dots, x_n | y_i) P(y_i) \\
&= \prod_{i=1}^{n} P(x_i | y_i) P(y_i) \\
&= \prod_{i=1}^{n} \{ \prod_{j=0}^{L-1} [\theta_j^y]^{1:(y_i=j)} \cdot \prod_{j=0}^{L-1} \prod_{k=1}^{d} [\theta_{\bar{x}^k|j}^{x^k|y}]^{1:(y_i=j, x_i^k=\bar{x}^k)} \} \\
&\simeq \sum_{i=1}^{n} \sum_{j=0}^{L-1} [1:(y_i=j)] \theta_j^y + \sum_{i=1}^{n} \sum_{j=0}^{L-1} \sum_{k=1}^{d} [1:(y_i=j, x_i^k=\bar{x}^k)][\theta_{\bar{x}^k|j}^{x^k|y}] \\
&= \sum_{j=0}^{L-1} N_j \cdot log \theta_j^y + \sum_{j=0}^{L-1} \sum_{k=1}^{d} [1:(y_i=j, x_i^k=\bar{x}^k)] \cdot log[\theta_{\bar{x}^k|j}^{x^k|y}] \\
&= \sum_{j=0}^{L-1} N_j \cdot log \theta_j^y + \sum_{i=1}^{n} \sum_{j=0}^{L-1} \sum_{k=1}^{d} [1:(y_i=j, x_i^k=\bar{x}^k)] \cdot [-\log(\sqrt{2\pi\sigma_{kj}^2}) - \frac{(x_i^k - \mu_{kj})^2}{2\sigma_{kj}^2}]
\end{aligned}$$

To get the largest value of $P(D)$, we obtain derivative of it with respect to μ_{kj} or obtain that iff $\mu_{kj} = x_i^k$