CS 8725: Report for assignment 2

Chanmann Lim

September 9, 2015

1. For $Y \in \{T, F\}$, one parameter is needed to describe $P(Y)$, two parameters are needed to describe $P(X_1|Y)$ and four parameters (μ_{iY}, σ_{iY}) are needed to describe $P(X_i|Y)$ for $2 \leq i \leq d$.

$P(Y)$
$P(X_1|Y), \quad Y \in \{T, F\},\ X_1 \in \{T, F\}$
$P(X_i|Y) \sim N(\mu_{iY}, \sigma^2_{iY}), \quad Y \in \{T, F\},\ 2 \leq i \leq d$
The total number of parameters $= 1 + 2 + 4 \times (d - 1) = 4d - 1$.

$$\begin{aligned} P(Y|X) &= \frac{P(X|Y) \cdot P(Y)}{P(X)} && (1)\\ &\propto P(X|Y) \cdot P(Y) && (2)\\ &= P(X_1|Y) \cdot \prod_{i=2}^{d} N(\mu_{iY}, \sigma^2_{iY}) \cdot P(Y) && (3) \end{aligned}$$

2. (a) Naive Bayes decision rule for

$$f_{NB}(Sunny, Windy) = \underset{Y}{\operatorname{argmax}}\, P(Sunny|Y) \cdot P(Windy|Y) \cdot P(Y) \quad (4)$$

Where $Y \in \{Hike, \neg Hike\}$ and

$$P(Hike) = P(\neg Hike) = 0.5$$

$$f_{NB}(Sunny, Windy) = \underset{Y}{\operatorname{argmax}}\, P(Sunny|Y) \cdot P(Windy|Y) \quad (5)$$

(b)

$$\begin{aligned} P(Sunny, Windy, Hike) &= P(Sunny, Windy|Hike) \cdot P(Hike) && (6)\\ &= P(Sunny|Hike) \cdot P(Windy|Hike) \cdot P(Hike) && (7)\\ &= 0.8 \times 0.4 \times 0.5 && (8)\\ &= 0.16 && (9) \end{aligned}$$

Similarly,

$$\begin{aligned} P(Sunny, Windy, \neg Hike) &= P(Sunny|\neg Hike) \cdot P(Windy|\neg Hike) \cdot P(\neg Hike) && (10)\\ &= 0.7 \times 0.5 \times 0.5 && (11)\\ &= 0.175 && (12) \end{aligned}$$

And the probability of error:

$$\begin{aligned} P_e &= 1 - P(Correct) && (13)\\ &= 1 - P(Y|Sunny, Windy) && (14)\\ &= 1 - \frac{P(Sunny, Windy, Y) \cdot P(Y)}{P(Sunny, Windy)} && (15) \end{aligned}$$

For the case when the weather is sunny and windy the error probability:

$$
\begin{aligned}
P_e(Hike|Sunny, Windy) &= 1 - \frac{0.16}{0.16 + 0.175} && (16)\\
&= 1 - 0.48 && (17)\\
&= 0.52 && (18)
\end{aligned}
$$