

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 $(\mu_{iY}, \sigma_{iY}^2)$ are needed to describe $P(X_i|Y)$ for $2 \leq i \leq d$.

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

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (1)$$

$$\propto P(X|Y) \cdot P(Y) \quad (2)$$

$$= P(X_1|Y) \cdot \prod_{i=2}^d N(x_i; \mu_{iY}, \sigma_{iY}^2) \cdot P(Y) \quad (3)$$

Where, $N(x_i; \mu_{iY}, \sigma_{iY}^2) = \frac{1}{\sqrt{2\pi\sigma_{iY}^2}} e^{-\frac{(x_i - \mu_{iY})^2}{2\sigma_{iY}^2}}$

2. (a) Naive Bayes decision rule is the function that maximizes the posterior probability however the denominator of the posterior probability is merely the normalization term to satisfy the property of probability and it does not depend on y thus it can be dropped.

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

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

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

$P(y)$ can also be dropped.

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

Therefore, predict Alice and Bob go hiking if $P(Sunny|Hike) \times P(Windy|Hike) > P(Sunny|\neg Hike) \times P(Windy|\neg Hike)$ and else otherwise.

(b) The joint probability of Alice and Bob go hiking and the weather is sunny and windy can be considered as the product of the likelihood of the weather given they go hiking $P(Sunny, Windy|Hike)$ and the prior probability of going hiking $P(Hike)$.

$$P(Sunny, Windy, Hike) = P(Sunny, Windy|Hike) \cdot P(Hike) \quad (6)$$

$$= P(Sunny|Hike) \cdot P(Windy|Hike) \cdot P(Hike) \quad (7)$$

$$= 0.8 \times 0.4 \times 0.5 \quad (8)$$

$$= 0.16 \quad (9)$$

Similarly,

$$P(Sunny, Windy, \neg Hike) = P(Sunny|\neg Hike) \cdot P(Windy|\neg Hike) \cdot P(\neg Hike) \quad (10)$$

$$= 0.7 \times 0.5 \times 0.5 \quad (11)$$

$$= 0.175 \quad (12)$$

And the probability of prediction error in case of Alice and Bob go hiking when the weather is sunny and windy is simply the probability that the weather is sunny and windy and Alice and Bob isn't go hiking.

$$P_e = P(Sunny, Windy, \neg Hike) \quad (13)$$

$$= 0.175 \quad (14)$$