HW 3 Part 2 6648-5170-90 Yu-Wei Chen

4.

(a) w_1 = evade taxes

w_2: do not evade taxes

(b)

Conditional Gaussanity

$$P(x|w_j) = N(\mu, \sigma^2)$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2}$$

$$\mu_1 = \frac{88+90+85}{3} = 87.67$$

$$\sigma_1^2 = \frac{\Sigma(x-\mu_i)^2}{n}$$
$$= \frac{(88-87.67)^2+(90-87.67)^2+(85-87.67)^2}{3}$$
$$= 4.22$$

$$\mu_2 = \frac{122+77+106+210+72+117+60}{7} = 109.14$$

$$\sigma_2^2 = \frac{12.86^2+(-32.14)^2+(-3.14)^2+(100.86)^2+(-37.14)^2+(7.86)^2+(-49.14)^2}{7}$$

$$= 2176.69$$

$$P(x|w_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = \frac{1}{5.15} e^{-\frac{(x-87.67)^2}{8.44}}$$

$$p(x \mid \omega_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} = \frac{1}{116.95} e^{-\frac{(x-109.14)^2}{4352.58}}$$

(c)

$$p(X_j = x_i \mid \omega_i) = \frac{n_{ji}}{n_i}$$

$X_1 = \{\text{Refund}, \text{No Refund}\}$

$p(X_1 = \text{Refund} \mid \omega_1) = 0$ $\qquad$ $p(X_1 = \text{No Refund} \mid \omega_1) = \frac{3}{3}$

$p(X_1 = \text{Refund} \mid \omega_2) = 1$ $\qquad$ $p(X_1\ \text{No Refund} \mid \omega_2) = \frac{4}{7}$

$X_2 = \{\text{Single}, \text{married}, \text{divorced}\}$

$p(X_2 = \text{single} \mid \omega_1) = \frac{2}{3}$ $\qquad$ $p(X_2 = \text{single} \mid \omega_2) = \frac{2}{7}$

$p(X_2 = \text{married} \mid \omega_1) = 0$ $\qquad$ $p(X_2 = \text{married} \mid \omega_2) = \frac{4}{7}$

$p(X_2 = \text{divorced} \mid \omega_1) = \frac{1}{3}$ $\qquad$ $p(X_2 = \text{divorced} \mid \omega_2) = \frac{1}{7}$

The estimate is valid because the probabilities sum to 1 for each feature given a class.

(d)

In the naive Bayes' Classifier, we assume that features are conditionally independent, meaning the probability of a feature set given a class is calculated as the product of individual conditional probabilities:

$$p(X \mid \omega_i) = p(X_1 \mid \omega_i) \cdot p(X_2 \mid \omega_i) \cdot \cdots \, p(X_k \mid \omega_i)\, p(\omega_i)$$

However, a problem arise when any of these conditional probab

is zero - this cause the entire expression to become zero.

To fix this issue, we apply laplace smoothing, which ensures

no probability is exact zero.

$$p(X = x_j \mid w_i) = \frac{n_{ji} + 1}{n_i + l}$$ ilities

where l represents the number of possible values that x can take. This adjustment guarantees that all features have a

non zero probability. that

(e)

Minimum Error Rate

$$p(w_1 \mid x) \underset{w_2}{\overset{w_1}{\gtrless}} p(w_2 \mid x)$$

$$\frac{\left[\prod_{i=1}^{k} p(x_i \mid w_1)\right] p(w_1)}{\cancel{p(x)}} \underset{w_2}{\overset{w_1}{\gtrless}} \frac{\left[\prod_{i=1}^{k} p(x_i \mid w_2)\right] p(w_2)}{\cancel{p(x)}}$$

$$p(w_1)\, p(x_1|w_1)\, p(x_2|w_1)\, p(x_3|w_1) \underset{w_2}{\overset{w_1}{\gtrless}} p(w_2)\, p(x_1|w_2)\, p(x_2|w_2)\, p(x_3|w_2)$$

x_1 = refund x_2 : Marital status x_3 = income

$p(x_1 = \text{Refund} | w_1) = \frac{0+1}{3+2} = \frac{1}{5}$ $p(x_1 = \text{Refund} | w_2) = \frac{3+1}{7+2} = \frac{4}{9}$

$p(x_1 = \text{No Refund} | w_1) = \frac{3+1}{3+2} = \frac{4}{5}$ $p(x_1 = \text{No Refund} | w_2) = \frac{4+1}{7+2} = \frac{5}{9}$

$p(x_2 = \text{Single} | w_1) = \frac{2+1}{3+3} = \frac{1}{2}$ $p(x_2 = \text{Single} | w_2) = \frac{2+1}{7+3} = \frac{3}{10}$

$p(x_2 = \text{Married} | w_1) = \frac{0+1}{3+3} = \frac{1}{6}$ $p(x_2 = \text{Married} | w_2) = \frac{4+1}{7+3} = \frac{1}{2}$

$p(x_2 = \text{Divorced} | w_1) = \frac{1+1}{3+3} = \frac{1}{3}$ $p(x_2 = \text{Divorced} | w_2) = \frac{1+1}{7+3} = \frac{1}{5}$

$$p(x_3|w_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = \frac{1}{5.15} e^{-\frac{(x-87.67)^2}{8.44}}$$

$$p(x_3|w_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} = \frac{1}{116.95} e^{-\frac{(x-109.14)^2}{4352.38}}$$