3 Bayes classifier

3.1

The likelihood for a single observation is

$$p(x_i, y_i|\theta) = p(x_i|y_i, \theta)p(y_i|\theta) \tag{1}$$

The parameters used in each case depend on the class label, as follows:

$$p(x_i, y_i|\theta) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_m)^2}{\sigma_m^2}\right)\pi_m & \text{if } y_i = m \\ \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_f)^2}{\sigma_f^2}\right)\pi_f & \text{if } y_i = f \end{cases} \tag{2}$$

This can be written as

$$p(x_i, y_i|\theta) = \left[\frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_m)^2}{\sigma_m^2}\right)\pi_m\right]^{\mathbb{I}(y_i=m)} \left[\frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_f)^2}{\sigma_f^2}\right)\pi_f\right]^{\mathbb{I}(y_i=f)} \tag{3}$$

The joint likelihood is the product of the above term for each data point,

$$p(x, y|\theta) = \prod_{i=1}^{N} p(x_i, y_i|\theta) \tag{4}$$

$$= \prod_{i=1}^{N} \left[\frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_m)^2}{\sigma_m^2}\right)\pi_m\right]^{\mathbb{I}(y_i=m)} \left[\frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_f)^2}{\sigma_f^2}\right)\pi_f\right]^{\mathbb{I}(y_i=f)} \tag{5}$$

The log-likelihood for this model is as follows. Note we make the substitution $\pi_f = 1 - \pi_m$, since we have the constraint $\pi_m + \pi_f = 1$.

$$\sum_{i=1}^{N} \log p(x_i, y_i) = \sum_{i=1}^{N} \mathbb{I}(y_i = m)\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_m^2 - \frac{1}{2}\frac{(x_i-\mu_m)^2}{\sigma_m^2} + \log\pi_m\right] \tag{6}$$

$$+ \sum_{i=1}^{N} \mathbb{I}(y_i = f)\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_f^2 - \frac{1}{2}\frac{(x_i-\mu_f)^2}{\sigma_f^2} + \log(1-\pi_m)\right] \tag{7}$$

In the following sections we find the ML estimates of the parameters by optimizing this log-likelihood.

3.1.1 π_m

$$\frac{\partial}{\partial \pi_m} \sum_{i=1}^{N} \log p(x_i, y_i) = \sum_{i=1}^{N} \left[\mathbb{I}(y_i = m) \frac{1}{\pi_m} - \mathbb{I}(y_i = f) \frac{1}{1 - \pi_m} \right] \tag{8}$$

$$= \frac{\sum_{i=1}^{N} \mathbb{I}(y_i = m)}{\pi_m} - \frac{\sum_{i=1}^{N} \mathbb{I}(y_i = f)}{1 - \pi_m} \tag{9}$$

Let $\sum_{i=1}^{N} \mathbb{I}(y_i = c) = N_c$. Setting to zero,

$$\frac{N_m}{\pi_m} - \frac{N_f}{1 - \pi_m} = 0 \tag{10}$$

$$(1 - \pi_m) N_m = \pi_m N_f \tag{11}$$

$$\pi_m = \frac{N_m}{N_m + N_f} \tag{12}$$

$$\pi_m = \frac{N_m}{N} \tag{13}$$

3.1.2 μ_m

$$\frac{\partial}{\partial \pi_m} \sum_{i=1}^{N} \log p(x_i, y_i) = \sum_{i=1}^{N} \mathbb{I}(y_i = m) \frac{x_i - \mu_m}{\sigma_m^2} \tag{14}$$

$$= \sum_{\{i:y_i=m\}} \frac{x_i - \mu_m}{\sigma_m^2} \tag{15}$$

Setting to zero,

$$\sum_{\{i:y_i=m\}} \frac{x_i - \mu_m}{\sigma_m^2} = 0 \tag{16}$$

$$\sum_{\{i:y_i=m\}} (x_i - \mu_m) = 0 \tag{17}$$

$$\mu_m \sum_{\{i:y_i=m\}} 1 = \sum_{\{i:y_i=m\}} x_i \tag{18}$$

$$\mu_m = \frac{\sum_{\{i:y_i=m\}} x_i}{N_m} \tag{19}$$

This is just the average of the x values for the males. The derivation for π_f is similar.

3.1.3 σ_m^2

$$\frac{\partial}{\partial \sigma_m^2} \sum_{i=1}^{N} \log p(x_i, y_i) = \sum_{i=1}^{N} \mathbb{I}(y_i = m) \left[-\frac{1}{2}\frac{1}{\sigma_m^2} + \frac{1}{2}\frac{(x_i - \mu_m)^2}{\sigma_m^4} \right] \tag{20}$$

$$= \sum_{\{i:y_i=m\}} \left[-\frac{1}{2}\frac{1}{\sigma_m^2} + \frac{1}{2}\frac{(x_i - \mu_m)^2}{\sigma_m^4} \right] \tag{21}$$

Setting to zero,

$$\sum_{\{i:y_i=m\}} \left[-\frac{1}{2}\frac{1}{\sigma_m^2} + \frac{1}{2}\frac{(x_i - \mu_m)^2}{\sigma_m^4} \right] = 0 \tag{22}$$

$$\sum_{\{i:y_i=m\}} (x_i - \mu_m)^2 = \sum_{\{i:y_i=m\}} \sigma_m^2 \tag{23}$$

$$\frac{\sum_{\{i:y_i=m\}} (x_i - \mu_m)^2}{N_m} = \sigma_m^2 \tag{24}$$

$$\tag{25}$$

3.1.4 Numerical results

$$\mu_m = \frac{67 + 79 + 71}{3} = 72.33 \tag{26}$$

$$\sigma_m^2 = \frac{(67 - 72.33)^2 + (79 - 72.33)^2 + (71 - 72.33)^2}{3} = 24.89 \tag{27}$$

$$\pi_m = 1/2 \tag{28}$$

$$\mu_f = \frac{68 + 67 + 70}{3} = 65 \tag{29}$$

$$\sigma_f^2 = \frac{(68 - 65)^2 + (67 - 65)^2 + (60 - 65)^2}{3} = 12.67 \tag{30}$$

$$\pi_f = 1/2 \tag{31}$$

$$\hat{\theta} = \big(\mu_m = 72.33, \sigma_m = 4.989, \pi_m = 1/2, \mu_f = 65, \sigma_f^2 = 3.559, \pi_f = 1/2\big) \tag{32}$$

3.2

$$
\begin{aligned}
p(y=m|x=72,\hat{\theta}) &= \frac{p(x=72|y=m,\hat{\theta})p(y=m|\hat{\theta})}{p(x=72|\hat{\theta})} && (33)\\
&= \frac{p(x=72|y=m,\hat{\theta})p(y=m|\hat{\theta})}{\sum_{c=\{m,f\}} p(x=72|y=c,\hat{\theta})p(y=c|\hat{\theta})} && (34)\\
&= \frac{\mathcal{N}(72|72.33,24.89)\times 0.5}{\mathcal{N}(72|72.33,24.89)\times 0.5+\mathcal{N}(72|65,12.67)\times 0.5} && (35)\\
&= \frac{0.07978\times 0.5}{0.07978\times 0.5+0.01620\times 0.5} && (36)\\
&= \frac{0.07978\times 0.5}{0.07978\times 0.5+0.01620\times 0.5} && (37)\\
&= 0.8312 && (38)
\end{aligned}
$$

3.3

A simple way to extend this model would be to use the Naive Bayes assumption; i.e. the attributes are independent conditional on the class.

$$
p(x_{1:d}|y) = \prod_{i=1}^{d} p(x_i|y) \quad (39)
$$

The class probabilities are then found by

$$
\begin{aligned}
p(y=c|x_{1:d}) &\propto p(x_{1:d}|y=c)p(y=c) && (40)\\
&\propto \left[\prod_{i=1}^{d} p(x_i|y=c)\right] p(y=c) && (41)
\end{aligned}
$$