Name: Chaoyang Zhu UIN: 930001840

1. [ZM] Exercises 18.5, Q2

Prior probabilities:

$$\hat{P}(Y) = \frac{n_1}{n} = \frac{4}{9} \qquad \hat{P}(N) = \frac{n_2}{n} = \frac{5}{9}$$

Likelihood:

Class Y:

$$\hat{\mu}_1 = \frac{5.0 + 7.0 + 3.0 + 6.0}{4} = 5.25$$

$$\hat{\sigma}_1^2 = \frac{(5 - 5.25)^2 + (7 - 5.25)^2 + (3 - 5.25)^2 + (6 - 5.25)^2}{4}$$
$$= 2.1875$$

$$\hat{\sigma}_1 = 1.479$$

$$\hat{f}(x \mid \hat{\mu}_1, \hat{\sigma}_1) = \hat{f}(1.0 \mid 5.25, 1.479) = 0.0043$$

Class N:

$$\hat{\mu}_2 = \frac{8.0 + 7.0 + 4.0 + 5.0 + 1.0}{5} = 5$$

$$\hat{\sigma}_2^2 = \frac{(8 - 5)^2 + (7 - 5)^2 + (4 - 5)^2 + (5 - 5)^2 + (1 - 5)^2}{5}$$
$$= 6$$

$$\hat{\sigma}_2 = 2.45$$

$$\hat{f}(x \mid \hat{\mu}_2, \hat{\sigma}_2) = \hat{f}(1.0 \mid 5, 2.45) = 0.0429$$

$$\hat{P}(x \mid Y) = \hat{P}(a_{1T} \mid Y) \cdot \hat{P}(a_{2F} \mid Y) \cdot \hat{f}(x \mid \hat{\mu}_1, \hat{\sigma}_1)$$
$$= \frac{3}{4} * \frac{1}{2} * 0.0043$$
$$= 1.6125 * 10^{-3}$$

$$\hat{P}(x \mid N) = \hat{P}(a_{1T} \mid N) \cdot \hat{P}(a_{2F} \mid N) \cdot \hat{f}(x \mid \hat{\mu}_2, \hat{\sigma}_2)$$
$$= \frac{1}{5} * \frac{2}{5} * 0.0429$$
$$= 3.432 * 10^{-3}$$

Posterior probabilities:

$$P(Y \mid x) \propto 1.6125 * 10^{-3} * \frac{4}{9} = 7.17 * 10^{-4}$$

$$P(N \mid x) \propto 3.432 * 10^{-3} * \frac{5}{9} = 1.91 * 10^{-3}$$

Thus, the predicted case is $\hat{y} = N$.

2. [ZM] Exercises 18.5, Q3

Prior probabilities:

$$P(c_1) = 0.5 \qquad P(c_2) = 0.5$$

Likelihood:

$$P(x \mid c_1) = f(x \mid \mu_1, \Sigma_1) = \frac{1}{2\pi\sqrt{|\Sigma_1|}} \exp\left\{-\frac{(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)}{2}\right\} = 0.0965$$

$$P(x \mid c_2) = f(x \mid \mu_2, \Sigma_2) = \frac{1}{2\pi\sqrt{|\Sigma_2|}} \exp\left\{-\frac{(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)}{2}\right\} = 0.0251$$

Posterior probabilities:

$$P(c_1 \mid x) \propto 0.0965 * 0.5 = 0.04825$$

$$P(c_2 \mid x) \propto 0.0251 * 0.5 = 0.01255$$

Thus, the predicted case is $\hat{y} = c_1$.

3. [ZM] Exercises 19.4, Q2

$$\hat{P}_H = \frac{4}{6} = \frac{2}{3} \qquad \hat{P}_L = \frac{2}{6} = \frac{1}{3}$$

$$H(D) = -(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}) = 0.918$$

$Age \leqslant 22.5$:

D_Y: 2H, 0L

P_H=2/2=1 P_L=0

D_N: 2H, 2L

P_H=2/4=1/2 P_L=2/4=1/2

$H(D_Y)$=0

$$H(D_N) = -(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}) = \log_2 2 = 1$$

$H(D_Y, D_N)$=2/6 $H(D_Y)$+ 4/6 $H(D_N)$=1/3*0+2/3*1=0.667

Gain=H(D)- $H(D_Y, D_N)$=0.918-0.667=0.251

$Age \leqslant 35$:

D_Y: 3H, 2L

P_H=3/5 P_L=2/5

D_N: 0H, 1L

P_H=1 P_L=0

$$H(D_Y) = -(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}) = 0.971$$

$H(D_N)$=0

$H(D_Y, D_N)$=5/6 $H(D_Y)$+ 1/6 $H(D_N)$=5/6*0.971+0=0.809

Gain=H(D)- $H(D_Y, D_N)$=0.918-0.809=0.109

$CarType \in Sports$

D_Y: 1H, 2L

P_H=1/3 P_L=2/3

D_N: 3H, 0L

P_H=1 P_L=0

$$H(D_Y) = -(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}) = 0.918$$

$H(D_N)$=0

$H(D_Y, D_N)$=3/6 $H(D_Y)$+ 3/6 $H(D_N)$=1/2*0.918=0.459

Gain=H(D)- H(D_Y, D_N)=0.918-0.459=0.459

$CarType \in Suv$

D_Y: 2H, 0L

P_H=1 P_L=0

D_N: 2H, 2L

P_H=1 /2 P_L=1/2

H(D_Y)=0

$$H(D_N) = -(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}) = \log_2 2 = 1$$

H(D_Y, D_N)=2/6 H(D_Y)+ 4/6 H(D_N)=1/3*0+2/3*1=0.667

Gain=H(D)- H(D_Y, D_N)=0.918-0.667=0.251

$CarType \in Vintage$

D_Y: 1H, 0L

P_H=1 P_L=0

D_N: 3H, 2L

P_H=3/5 P_L=2/5

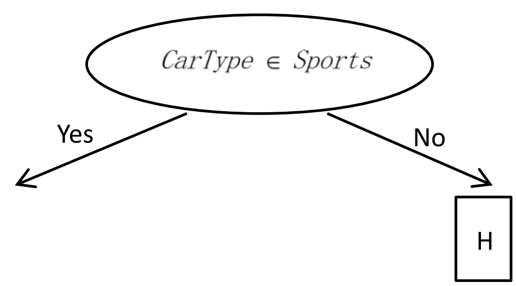
H(D_Y)=0

$$H(D_N) = -(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}) = 0.971$$

H(D_Y, D_N)=1/6 H(D_Y)+ 5/6 H(D_N)=1/6*0+5/6*0.971=0.809

Gain=H(D)- H(D_Y, D_N)=0.918-0.809=0.109

Thus $CarType \in Sports$ is chosen as the root of the decision tree.



Point	Age	Car	Risk
X1	25	Sports	L
X3	25	Sports	L
X5	20	Sports	H

Point	Age	Car	Risk
X2	20	Vintage	H
X4	45	Suv	H
X6	25	Suv	H

$$H(D) = -(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}) = 0.918$$ H(D)=0

$Age \leqslant 22.5$

D_Y: 1H, 0L

$P_H=1$ $P_L=0$

D_N: 0H, 2L

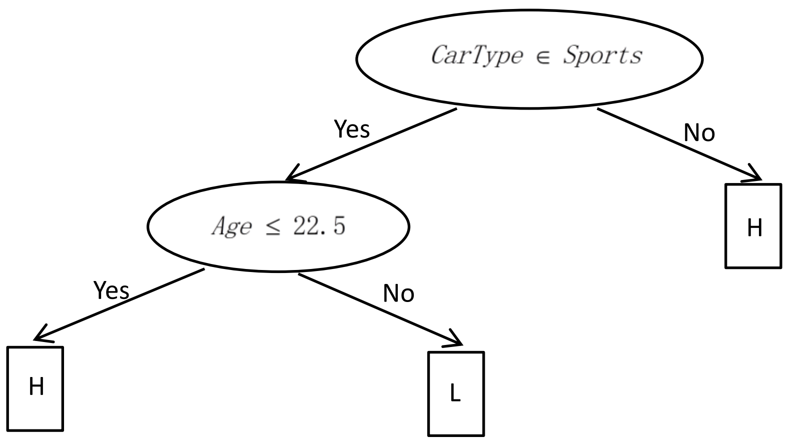
$P_H=0$ $P_L=1$

$H(D_Y)=0$

$H(D_N)=0$

$H(D_Y, D_N)=0$

Gain=H(D)- $H(D_Y, D_N)=0.918$



The point (Age=27,Car=Vintage) is classified as **H**.