University of Glasgow | School of Computing Science

# CSC3005 Laboratory/Tutorial 4 Solution: Classification Analysis II

## 6. Navies Bayes Theorem

a) Let the

$$P\left(\frac{Female}{Undergradaute}\right) = P\left(\frac{F}{UG}\right) = 0.15$$

$$P\left(\frac{Female}{Gradaute}\right) = P\left(\frac{F}{G}\right) = 0.23$$

$$P(Graduate) = P(G) = 0.2$$

$$P(Undergraduate) = P(UG) = 0.8$$

Find $P\left(\frac{Graduate}{Female}\right) = P\left(\frac{G}{F}\right)$?

Using Bayes Theorem,

$$P\left(\frac{G}{F}\right) = \frac{P\left(\frac{F}{G}\right) \times P(G)}{P\left(\frac{F}{UG}\right) \times P(UG) + P\left(\frac{F}{G}\right) \times P(G)} = \frac{0.23 \times 0.2}{0.15 \times 0.8 + 0.23 \times 0.2} = 0.277$$

b) As $P(UG) = 0.8 > P(G) = 0.2$ , the randomly chosen university student is more likely to an undergraduate student for part (a)

c) An undergraduate student is most likely to be chosen as

$$P\left(\frac{G}{F}\right) + P\left(\frac{UG}{F}\right) = 1$$

$$\rightarrow P\left(\frac{UG}{F}\right) = 1 - P\left(\frac{G}{F}\right) = 1 - 0.277 = 0.723$$

d) Find $P\left(\frac{Graduate}{Female,Hostel}\right) = P\left(\frac{G}{F,H}\right)$ and $P\left(\frac{undergraduate}{Female,Hostel}\right) = P\left(\frac{UG}{F,H}\right)$?

To find $P\left(\frac{G}{F,H}\right)$,

Using Baye theorem and assuming independent conditional probability

$$P\left(\frac{G}{F,H}\right) = \frac{P\left(\frac{F,H}{G}\right)P(G)}{P(FH)}$$
$$= \frac{P\left(\frac{F}{G}\right)P\left(\frac{H}{G}\right)P(G)}{P(FH)} = \frac{P\left(\frac{F}{G}\right)P\left(\frac{H}{G}\right)P(G)}{P(F)P(H)}$$

Similarly to find $P\left(\frac{UG}{F,H}\right)$

$$P\left(\frac{UG}{F,H}\right) = \frac{P\left(\frac{F,H}{UG}\right)P(UG)}{P(FH)}$$
$$= \frac{P\left(\frac{F}{UG}\right)P\left(\frac{H}{UG}\right)P(UG)}{P(FH)} = \frac{P\left(\frac{F}{UG}\right)P\left(\frac{H}{UG}\right)P(UG)}{P(F)P(H)}$$

Given that

$$P\left(\frac{H}{G}\right) = 0.3\,, P\left(\frac{H}{UG}\right) = 0.1$$

$$P(H) = P(H, UG, G) = P(H, UG) + P(H, G)$$
$$= P\left(\frac{H}{UG}\right)P(UG) + P\left(\frac{H}{G}\right)P(G)$$
$$= 0.1 \times 0.8 + 0.3 \times 0.2 = 0.14$$

$$P(F) = P(F, UG, G) = P(F, UG) + P(F, G)$$
$$= P\left(\frac{F}{UG}\right)P(UG) + P\left(\frac{F}{G}\right)P(G)$$
$$= 0.15 \times 0.8 + 0.23 \times 0.2 = 0.166$$

Therefore,

$$P\left(\frac{G}{F,H}\right) = \frac{P\left(\frac{F}{G}\right)P\left(\frac{H}{G}\right)P(G)}{P(F)P(H)} = \frac{0.23 \times 0.3 \times 0.2}{0.166 \times 0.14} = 0.59$$

$$P\left(\frac{UG}{F,H}\right) = \frac{P\left(\frac{F}{UG}\right)P\left(\frac{H}{UG}\right)P(UG)}{P(F)P(H)} = \frac{0.15 \times 0.1 \times 0.8}{0.166 \times 0.14} = 0.516$$

Since $P\left(\frac{G}{F,H}\right) > P\left(\frac{UG}{F,H}\right)$, the female student who stay in hostel is more likely a graduate student

## 7. K Nearest Neighbor Theory

| | categorical | categorical | continuous | class |
|---|---|---|---|---|
| ID | Home Owner | Marital Status | Annual Income | Defaulted Borrower |
| 1 | Yes | Single | 125k | No |
| 2 | No | Married | 100k | No |
| 3 | No | Single | 70k | No |
| 4 | Yes | Married | 120k | No |
| 5 | No | Divorced | 95k | Yes |
| 6 | No | Married | 60k | No |
| 7 | Yes | Divorced | 220k | No |
| 8 | No | Single | 85k | Yes |
| 9 | No | Married | 75k | No |
| 10 | No | Single | 90k | Yes |

The above training set can be firstly summarized for the Home owner and Marital Status

| Home Owner | | Class |
|---|---|---|
| Yes | No | |
| 0 | 3 | Yes |
| 3 | 4 | No |

| Marital Status | | | Class |
|---|---|---|---|
| Single | Married | Divorced | |
| 2 | 0 | 1 | Yes |
| 2 | 4 | 1 | No |

$$d(Home\ Owner = Yes\ , Home\ Owner = No) = \left|\frac{0}{3} - \frac{3}{7}\right| + \left|\frac{3}{3} - \frac{4}{7}\right| = \frac{3}{7} + \frac{3}{7} = \frac{6}{7}$$

$$d(Single, Married) = \left|\frac{2}{4} - \frac{0}{4}\right| + \left|\frac{2}{4} - \frac{4}{4}\right| = \frac{2}{4} + \frac{2}{4} = 1$$

$$d(Single, Divorced) = \left|\frac{2}{4} - \frac{1}{2}\right| + \left|\frac{2}{4} - \frac{1}{2}\right| = \frac{0}{4} + \frac{0}{4} = 0$$

$$d(Married, Divorced) = \left|\frac{0}{4} - \frac{1}{2}\right| + \left|\frac{4}{4} - \frac{1}{2}\right| = \frac{1}{2} + \frac{1}{2} = 1$$

## 8. Artificial Neural Network

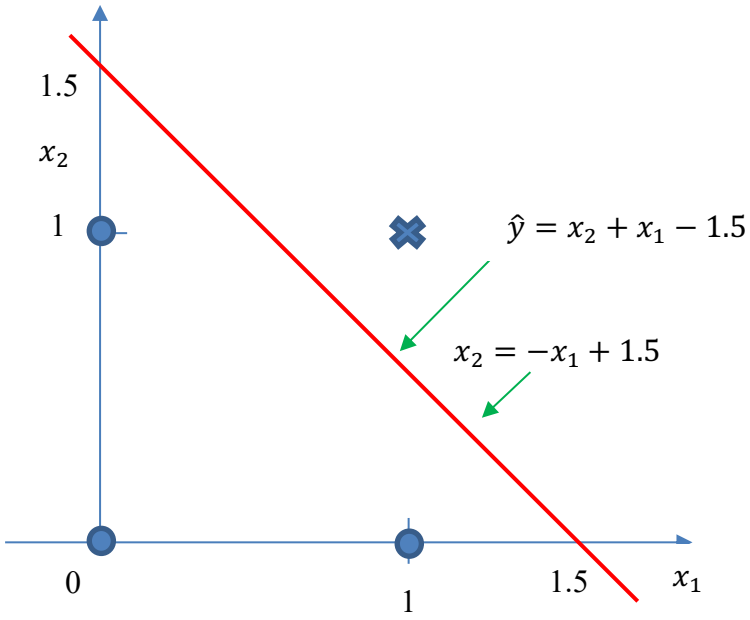
And Operation

| $x_1$ | $x_2$ | $y$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

Find the equation for the predicted y, $\hat{y}$

where

$$\hat{y} = w_1 x_1 + w_2 x_2 + w_0$$




From the above

the line $x_2 = -x_1 + 1.5$ can be transformed into

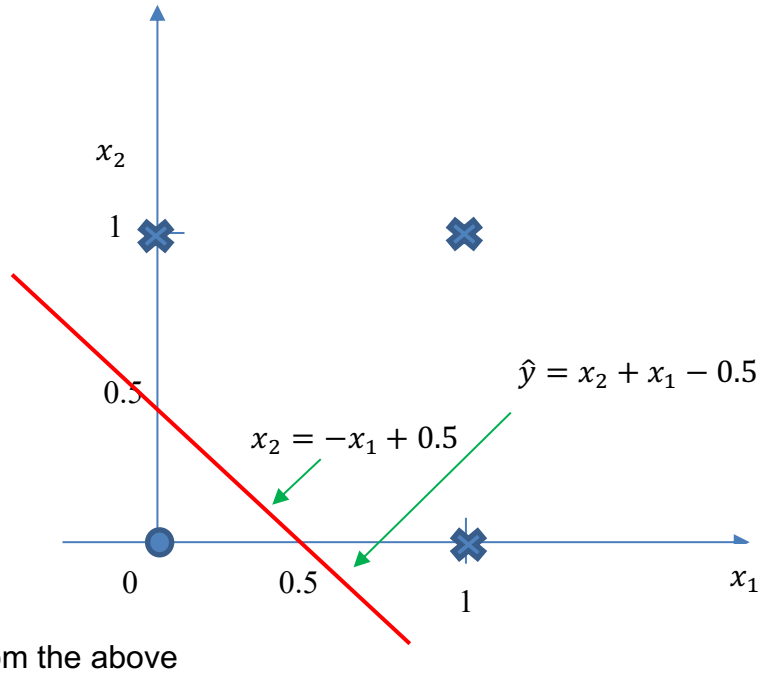
$$\begin{aligned}\hat{y} &= w_1x_1 + w_2x_2 + w_0 \\ &= -x_1 - x_2 + 1.5\end{aligned}$$

where $w_1 = -1, w_2 = -1$ and $w_0 = 1.5$. But to match the truth table, so that any value of $\hat{y} > 0$ will be class 1 and $\hat{y} < 0$ will be class 0 with any combination of $x_1$ and $x_2$, change the sign to

$$\begin{aligned}\hat{y} &= w_1x_1 + w_2x_2 + w_0 \\ &= x_1 + x_2 - 1.5\end{aligned}$$

OR Operation

| $x_1$ | $x_2$ | $y$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |




From the above

the line $x_2 = -x_1 + 0.5$ can be transformed into

$$\begin{aligned}\hat{y} &= w_1x_1 + w_2x_2 + w_0 \\ &= -x_1 - x_2 + 0.5\end{aligned}$$

where $w_1 = -1, w_2 = -1$ and $w_0 = 0.5$. But to match the truth table, so that any value of $\hat{y} > 0$ will be class 1 and $\hat{y} < 0$ will be class 0 with any combination of $x_1$ and $x_2$, change the sign to

$$\begin{aligned}\hat{y} &= w_1x_1 + w_2x_2 + w_0 \\ &= x_1 + x_2 - 0.5\end{aligned}$$