Machine Learning

EX2-classification

Matthis Brocheton

Probleme 1:

(A)

X = $\begin{bmatrix} 0.5 & 0 & 1 \\ 0.75 & 0.25 & 1 \\ 1 & 0.5 & 1 \\ 0 & 0.5 & 1 \\ 0.25 & 0.75 & 1 \\ 0.5 & 1 & 1 \end{bmatrix}$

Y = $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

normal equation formula

$$W = (X^T X)^{-1} X^T Y$$

So, I use the normal equation to find the w1, w2, w3:

```python
import numpy as np


X = np.array([
    [0.5,  0,   1 ],
    [0.75, 0.25, 1 ],
    [1,    0.5,  1 ],
    [0,    0.5,  1 ],
    [0.25, 0.75, 1 ],
    [0.5,  1,    1 ]
])

# Vecteur Y
Y = np.array([0, 0, 0, 1, 1, 1])

print(np.linalg.inv(X.T @ X) @ X.T @ Y)
```

I found:

$w1 = -1$

$w2 = 1$

$w3 = 0.5$

Calculation

$$y = w_1 x_1 + w_2 x_2 + w_3$$

$$y_7 = (-1 \times 0.5) + (1 \times 0.25) + 0.5$$

$y_7 = 0.25$

$$y_8 = (-1 \times 0.5) + (1 \times 0.75) + 0.5$$

$$y_8 = 0.75$$

$$y_9 = (-1 \times 0.75) + (1 \times 0.5) + 0.5$$

$$y_9 = 0.25$$

(B)

$$X = \begin{bmatrix} 0.5 & 0 \\ 0.75 & 0.25 \\ 1 & 0.5 \\ 0 & 0.5 \\ 0.25 & 0.75 \\ 0.5 & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

normal equation formula

$$W = (X^T X)^{-1} X^T Y$$

So, I use the normal equation to find the w1, w2, w3:

```
import numpy as np

# Matrice X (sans colonne de biais)
X = np.array([
    [0.5,  0  ],
    [0.75, 0.25],
    [1,    0.5 ],
    [0,    0.5 ],
    [0.25, 0.75],
    [0.5,  1   ]
])

# Vecteur Y
Y = np.array([0, 0, 0, 1, 1, 1])

# Calcul de W
W = np.linalg.inv(X.T @ X) @ X.T @ Y

print(W)
```

I found:

$$w_1 = -0.57142857$$

$$w_2 = 1.42857143$$

Calculation

$y = w_1 x_1 + w_2 x_2$

$y_7 = (-1 \times 0.5) + (2 \times 0.25)$

$y_7 = 0$

$y_8 = (-1 \times 0.5) + (2 \times 0.75)$

$y_8 = 1$

$y_9 = (-1 \times 0.75) + (2 \times 0.5)$

$y_9 = 0.25$

Probleme 2:

Day	Outlook	Temperature	Humidity	Wind	Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Overcast	Cool	Normal	Weak	Yes
D7	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$\boldsymbol{P(Tennis = Yes) = \frac{9}{14}}$$

$$\boldsymbol{P(Tennis = No) = \frac{5}{14}}$$

$$P(\,Outlook \mid Tennis = Yes\,) = \frac{2}{9}$$

$$P(\,Outlook \mid Tennis = No\,) = \frac{3}{5}$$

$$P(\,Temperature \mid Tennis = Yes\,) = \frac{3}{9}$$

$$P(\,Temperature \mid Tennis = No\,) = \frac{1}{5}$$

$$P(\,HighHumidity \mid Tennis = Yes\,) = \frac{3}{9}$$

$$P(\,Humidity \mid Tennis = No\,) = \frac{4}{5}$$

$$P(\,Wind \mid Tennis = Yes\,) = \frac{3}{9}$$

$$P(\,Wind \mid Tennis = No\,) = \frac{3}{5}$$

$$P(Tennis = Yes|x) = \; P(Tennis = Yes) \times P(Outlook \mid Tennis = Yes) \times P(Temperature \mid Tennis = Yes) \times P(Humidity \mid Tennis = Yes) \times P(Wind \mid Tennis = Yes)$$

$$P(Tennis = Yes \mid x) \propto \frac{9}{14} \times \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9}$$

$$P(\,Tennis = No \mid x\,) \propto \frac{5}{14} \times \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{2}{5}$$

$$P(\,Tennis = Yes \mid x\,) = 0.00529100$$

$$P(\,Tennis = No \mid x\,) = 0.01371428$$

$$P(\,Tennis = Yes \mid x\,) < \; P(\,Tennis = No \mid x\,)$$

Probleme 3:

(A)

```
import numpy as np
CAT = 1
EGG = 0
AGE = (1 / 3) * np.log(40)

a0 = 4
a1 = 0.7
a2 = 0.03
a3 = 0.4

x1 = CAT
x2 = EGG
x3 = AGE

z = a0 + a1 * x1 + a2 * x2 + a3 * x3


print(z)

p_x = 1 / (1 + np.exp(-z))

print(f"Probability of heart disease : {p_x:.4f}")
```

$a_0 = 4$

$a_1 = 0.7$

$a_2 = 0.03$

$a_3 = 0.4$

$x_1 = 1$

$x_2 = 0$

$x_3 = \frac{1}{3} log(40)$

$z = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$

$z = 5.191850593881858$

$p(x) = \frac{1}{1 + e^{-z}}$

$p(x) = 0.994469056646$

We obtain a probability of 0.9944

(B)

$z = 5.191850593881858$

$$p(x) = \frac{1}{1 + e^{-z}}$$

$$\log \mathrm{p(x)} = \frac{p(x)}{1 - p(x)}$$

$\log \mathrm{p(x)} = 179.80098384$

There is 179.8 times more likely than otherwise

(C)

$z = 5.191850593881858$

$$\boldsymbol{p(x) = \frac{1}{1 + e^{-z}}}$$

$$\mathbf{ln}\boldsymbol{p(x)} = \mathbf{l\,n}\left(\frac{p(x)}{1 - p(x)}\right)$$

$\mathbf{ln}\boldsymbol{p(x)} = \mathbf{4}$

$\mathbf{ln}\boldsymbol{p(x)} = \boldsymbol{a_0}$

If all x is 0 $ln\ p(x)$ show the value a_0.

This means that the probability of having heart disease is determined **only** by the intercept a_0, which represents the baseline risk when none of the independent variables contribute to the prediction.