5. perform multi linear regression.

to predict credit score.

credit score = $\theta_0 + \theta_1 \cdot \text{Age} + \theta_2 \cdot \text{education}$.

where

θ_0 : intercept

θ_1 : coefficient for Age

θ_2 : coefficient for education.

use normal equation

$$\theta = (X^T X)^{-1} X^T y$$

where

- X is the design matrix
  - column for intercept
  - Age
  - education
- y : vector of target values
- X^T is the transpose of X
- $(X^T X)^{-1}$ is the inverse of the matrix $X^T X$

Handling missing value.

education = 14.

* formulate the design matrix X,

Target vector y

$$X = \begin{vmatrix} 1 & 35 & 16 \\ 1 & 28 & 14 \\ 1 & 45 & 14 \\ 1 & 31 & 12 \\ 1 & 52 & 18 \\ 1 & 29 & 14 \\ 1 & 42 & 16 \\ 1 & 33 & 12 \end{vmatrix} \quad y = \begin{vmatrix} 720 \\ 650 \\ 750 \\ 600 \\ 780 \\ 630 \\ 710 \\ 640 \end{vmatrix}$$

LHP

LHP

* compute $X^T X$

$$X^T = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 35 & 28 & 45 & 31 & 52 & 29 & 42 & 33 \\ 16 & 14 & 14 & 12 & 18 & 14 & 16 & 12 \end{vmatrix}$$

$$X^T X = \begin{vmatrix} \Sigma 1 & \Sigma Age & \Sigma education \\ \Sigma Age & \Sigma Age^2 & \Sigma edu \times Age \\ \Sigma education & \Sigma Age . education & \Sigma edu^2 \end{vmatrix}$$

$$X^T X = \begin{vmatrix} 8 & 295 & 116 \\ 295 & 11393 & 4364 \\ 116 & 4364 & 1712 \end{vmatrix}$$

*

$$A = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$$

$$A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A)$$

$$A = \begin{matrix} 8 & 295 & 116 \\ 295 & 11393 & 4364 \\ 116 & 4364 & 1712 \end{matrix}.$$

$$\det(A) = 563712$$

$$\text{adj}(X^T X) = \begin{matrix} 460320 & -1184 & -34208 \\ 1184 & 240 & 672 \\ -34208 & 672 & 4119 \end{matrix}$$

LHP

$$(X^T X)^{-1} = \frac{1}{\det(X^T X)} \cdot \operatorname{adj}(X^T X)$$

$$= \frac{1}{563712} \begin{bmatrix} 460320 & -1184 & -34208 \\ 1184 & 240 & 672 \\ -34208 & 672 & 4119 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{14385}{17616} & \frac{-37}{17616} & \frac{-1069}{17616} \\ \frac{37}{17616} & \frac{5}{11748} & \frac{7}{732} \\ \frac{-1069}{17616} & \frac{7}{732} & \frac{1373}{187904} \end{bmatrix}$$

Compute $X^T y$

$$X^T y = \begin{vmatrix} \Sigma y_i \\ \Sigma(Age_i \cdot y_i) \\ \Sigma(education_i \cdot y_i) \end{vmatrix}$$

$$X^T y = \begin{vmatrix} 5280 \\ 205520 \\ 80220 \end{vmatrix}$$

LHP

6.

Credit Score =

208.5714 + 5.7143 · Age +

20000 · education

compute squared errors

ID 1 : 73.47

2 : 2.03

3 : 18.36

4 : 661.44

5 : 7347.06

6 : 589.81

7 : 3430.67

8 : 8.16

$$MSE = \frac{12131.0434}{8} = 1516.38$$

Calculate R^2

$SS_{res} = 12131$

$SS_{tot} = 33600$

$R^2 = 1 - \frac{12131.0434}{33600}$

$= 1 - 0.3610 = 0.6390$

7.

- formulate the logistic regression

$z = w_0 + w_1 \text{ age} + w_2 \text{ credit score}$

$\hat{y} = \frac{1}{1 + e^{-z}}$

given $w_0 = 0.5$, $w_1 = -0.02$,

$w_2 = 0.01$

$z = 0.5 - 0.02 \cdot \text{Age} + 0.01 \cdot CS$

T1 : Age = 40, credit score = 700

$z = 0.5 - 0.02 \cdot 40 + 0.01 \cdot 700$

$= 0.5 - 0.8 + 70.$

$= 6.7.$

Thứ ngày tháng năm

$$\hat{y} = \frac{1}{1 + e^{-67}} = 0.99877$$

cost function.

$$J(w) = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)]$$

$$= 3.516004.$$

Thứ ngày tháng năm

Olympic

δ. ≥ 0	$\hat{y}_i - y_i$
1	0.999088
2	−0.001595
3	0.999175
4	−0.002789
5	0.999297
6	−0.001987
7	0.998842
8	−0.001948

gra

$$\frac{\partial J}{\partial w_j} = \frac{1}{n} \sum (\hat{y}_i - y_i)\, x_{i,j} \quad , n = 8$$

$$\sum (\hat{y}_i - y_i) = 0.999088 - 0.001595 + \ldots +$$

$$\frac{\partial J}{\partial w_0} = \frac{3.988083}{8} = 0.498510.$$

Thứ ngày tháng năm

9. * Từ from Question 7

Age = 40

credit Score = 700

Age : 28 to 52 rang = 24

credit score : 600 to 780 range = 180

min max normalization.

(Scale to ___

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$$Age : Age_{norm} = \frac{40 - 28}{52 - 28} = 0.5$$

Thứ ngày tháng năm

$$credit\ score_{norm} = \frac{700 - 600}{780 - 600} = \frac{100}{180}$$

$$= 0.56$$

normalized Age, credit score

$$= 0.5,\ 0.56$$

* calculation

perceptron weight $= [0.3, 0.4]$

bias $= 0.1$

$z = w1 \cdot Age_{norm} + w2 \cdot credit\ score_{norm} + b$

$= (0.3 \times 0.5) + (0.4 \times 0.56) + 0.1$

$= 0.15 + 0.22 + 0.1$

$= 0.47$

Activation fu , threshold at 0.

$$\text{output} = \begin{cases} 1 \text{ if } z \geq 0 \\ 0 \text{ if } z < 0 \end{cases}$$

since $z = 0.42 > 0$

output = 1 (high risk)

10.

Input T_1 normalized

Age, credit score = 0.375, 0.583

Hidden layer

$W_1 = \begin{matrix} 0.3 & 0.5 \\ 0.4 & -0.2 \end{matrix}$ $b_1 = \begin{matrix} 0.1 \\ -0.1 \end{matrix}$

$z_1 = w_1 \cdot \text{Input} + b_1$

$z_{1_1} = 0.3 \times 0.375 + 0.4 \times 0.583 + 0.1$

$= 0.1125 + 0.2332 + 0.1$

$= 0.4457$

$z_{1_2} = 0.5 \times 0.35 + \_\_ \times 0.583 + (-0.1)$

$= 0.1875 - 0.1166 - 0.1$

$= -0.0291.$

Sigmoid activation

$$h_i = \sigma(z_{1i}) = \frac{1}{1+e^{-z_{1i}}}$$

$$= \sigma(0.4457) = 0.6$$

$$h_2 = \sigma(-0.0291) = 0.49$$

output layer

$$w_2 = \begin{matrix} 0.6 \\ -0.4 \end{matrix} \qquad b_2 = 0.2$$

$$z_2 = w_2 \cdot h + b_2$$

$$z_2 = 0.6 \times 0.6 + (-0.4 \cdot 0.4) + 0.2$$

$$= 0.3657$$

output $\hat{y} = \sigma(z_2) = 0.59$

11

$$P(x \mid y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for Age

Low ($y=0$): Ages: 35, 45, 52, 42

$\mu_{Age,0} = 43.5$

$\sigma^2_{Age,0} =$ $\sigma_{Age,0} = 6.103$

for Age $= 40$

$$P(Age = 40 \mid y = 0) = 0.0614$$

High ($y=1$) Ages: 28, 31, 29, 33

$\mu_{Age,1} = 30.25$

$\sigma^2_{Age,1} = 4.1875$ $\sigma_{Age,1} = 2046$

$\underline{1}$ (Age 40 | y = 1)

= 0.0017

for credit score

$\mu_{cs,0} = 740$

$\sigma^2_{cs,0} = 525$; $\sigma_{cs,0} = 22.913$

for credit score = 700

$\underline{1}$ (CS = 700 | y = 0) = 0.01

$\mu_{cs,1} = 630$

$\sigma^2_{cs,1} = 325$ $\sigma_{cs,1} = 18.028$

$\underline{1}$ (CS = 700 | y = 1) = 0.001.

$P(y \mid Age, CS) \propto P(y) \cdot P(Age \mid y) \cdot P(\quad \mid y)$

$P(y=0) \cdot P(Age = 40) \cdot P(CS = 700 \mid y = 0)$

$= 0.5 \times 0.0614 \times 0.0123$

$= 0.000377$

$P(y=1) \cdot P(Age = 40 \mid y = 1) \cdot$

$P(CS = 700 \mid y = 1)$

$= 0.5 \times 0.0017 \times 0.0019$

$= 0.0000016$

12.

Age distribution

Low mean = 43.5

high mean = 30.25

credit score distribution

Low mean = 740

high mean = 630.

Source of bias

Age (younger → high risk

mean 30.25, 43.5)

credit score (lower → high risk

mean 630 vs 740.