Timothy Mah
8750634
txamah001@mymail.sim.edu.sg
CSIT375 Artificial Intelligence for Cybersecurity **quiz2 Q1**

FORMULA FOR LR

Predicted label = (the label when all features are 0)
+ [(change in label when feature +1) × Feature 1]
+ ....

FORMULA FOR MSE

$$MSE = \frac{(\text{predicted label} - \text{actual label})^2 + \dots}{2 \times \text{no. of datapoints}}$$

FORMULA FOR GRADIENT DESCENT

1. Find Gradients

gradient of feature = $-\frac{2}{\text{no. of points}}$ (label 1 + label 2 + ...)

gradient of label = $-\frac{2}{\text{no. of points}}$ [(label 1 × feature 1) + (label 2 × feature 2) + ...]

2. Find new data points (descent based on gradient)

new feature = feature 1 − (learning rate × feature gradient)
new label = label 1 − (learning rate × label gradient)

3. Update LR equation

predicted label 1 = new feature + (new label × old feature)

4. Repeat for each datapoint

5. Find new gradients again. Loop until gradient = 0

(a)(i)

if $\theta_0 = 0$ and $\theta_1 = 1$,

when feature = 1,
predicted label = 0 + 1 × 1 = 1
when feature = 2,
predicted label = 0 + 1 × 2 = 2
when feature = 3
predicted label = 0 + 1 × 3 = 3
when feature = 4
predicted label = 0 + 1 × 4 = 4

$$MSE = \frac{(1-2)^2 + (2-1)^2 + (3-4)^2 + (4-3)^2}{2 \times 4}$$
$$= \frac{1+1+1+1}{8}$$
$$= \frac{4}{8}$$
$$= 0.5 //$$

MSE IS 0.5

(a)(ii)

if $\theta_1 = 1$ and $\theta_1 = 3/5$

when feature = 1,
predicted label = $1 + 1 \times \frac{3}{5} = \frac{8}{5}$
when feature = 2,
predicted label = $1 + 2 \times \frac{3}{5} = \frac{11}{5}$
when feature = 3,
predicted label = $1 + 3 \times \frac{3}{5} = \frac{14}{5}$
when feature = 4,
predicted label = $1 + 4 \times \frac{3}{5} = \frac{17}{5}$

$$MSE = \frac{(\frac{8}{5} - 2)^2 + (\frac{11}{5} - 1)^2 + (\frac{14}{5} - 4)^2 + (\frac{17}{5} - 3)^2}{2 \times 4}$$
$$= \frac{\frac{4}{25} + \frac{36}{25} + \frac{36}{25} + \frac{4}{25}}{2 \times 4}$$
$$= \frac{\frac{80}{25}}{8}$$
$$= 0.4 //$$

MSE is 0.4

(b)

Round (1) of Gradient Descent

Gradient of feature = $-\frac{2}{4}(2 + 1 + 4 + 3)$
= −2

Gradient of label = $-\frac{2}{4}[(2 \times 1) + (1 \times 2) + (4 \times 3) + (3 \times 4)]$
= −14

New feature = 0 − (0.2 × −2)
= 0.4

New label = 0 − (0.2 × −14)
= 2.8

Round (2) of Gradient Descent

New predicted label 1 = 0.4 + (2.8 × 1)
= 3.2

New predicted label 2 = 0.4 + (2.8 × 2)
= 6

New predicted label 3 = 0.4 + (2.8 × 3)
= 8.8

New predicted label 4 = 0.4 + (2.8 × 3)
= 11.6

New gradient of features = $-\frac{2}{4}[(3.2-2) + (6-1) + (8.8-4) + (11.6-3)]$
= 9.8

New gradient of labels = $-\frac{2}{4}\{[(3.2-2) \times 1] + [(6-1) \times 2] + [(8.8-4) \times 3] + [(11.6-3) \times 4]\}$
= 30

New feature = 0.4 − (0.2 × 9.8)
= −1.56

New label = 2.8 − (0.2 × 30)
= −3.2

Round (3) of Gradient Descent

New predicted label 1 = −1.56 − (3.2 × 1)
= −4.76

New predicted label 2 = −1.56 − (3.2 × 2)
= −7.96

New predicted label 3 = −1.56 − (3.2 × 3)
= −11.16

New predicted label 4 = −1.56 − (3.2 × 3)
= −14.36

New gradient of features = $-\frac{2}{4}[(-4.76-2) + (-7.96-1) + (-11.6-4) + (-14.36-3)]$
= −24.12

New gradient of labels = $-\frac{2}{4}\{[(-4.76-2) \times 1] + [(-7.96-1) \times 2] + [(-11.16-4) \times 3] + [(-14.36-3) \times 4]\}$
= −69.8

New feature = −1.56 − (0.2 × −24.12)
= 3.264 //

New label = −3.2 − (0.2 × −69.8)
= 10.76 //

$\theta_1 = 3.264$ and $\theta_2 = 10.76$