ELE 888 - Midterm Solutions - Version A  1 mark for version

Note Title  2/17/2016

Q1 2 - Regression and Classification are supervised learning algorithms [4]

- Regression predicts continuously valued outputs
- Classification predicts discrete valued outputs } 1

3 - Batch Gradient Descent performs a parameter update after all training examples have been processed

- Stochastic Gradient Descent performs a parameter update after 1 training example } 1

- Stochastic will be faster for large datasets because updates occur immediately instead of waiting for all examples } 1

6

(4) Setting the learning rate too large will make the algorithm oscillate in terms of the cost function or even diverge. 1 The optimal parameters can never be found. 1

(5) a) TRUE - 1 Adding new features and finding the optimal parameters using the training set will allow for higher accuracy though there is a chance for over fitting 1

b) FALSE - 1 If the regularization parameter is too large, there will be underfitting and the performance will be worse 1

1 5

c) FALSE. adding new features will allow accuracy to be better but there will be overfitting as a result 1

6 $\lambda = 0 \rightarrow \theta = \begin{bmatrix} 81.3562 \\ 12.65 \\ 9.54 \end{bmatrix}$ 1 $\lambda = 1 \rightarrow \theta = \begin{bmatrix} 13.4569 \\ 0.94 \\ 0.78 \end{bmatrix}$ 1

as $\lambda$ increases, the magnitudes of the parameters should decrease 1

$3.5$

Q2 ① 

| Example | $x_1$ ($x^2$) | $x_2$ ($\sqrt{x}$) |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 6.25 | 1.581 |
| 3 | 16 | 2 |
| 4 | 36 | 2.4495 |

(1) (1)

②

$$\mu_1 = \frac{1 + 6.25 + 16 + 36}{4}$$

$$= 14.8125 \quad (0.5)$$

$$\sigma_1 = \sqrt{\frac{1}{4-1}\left[(1-14.8125)^2 + (6.25-14.8125)^2 + (16-14.8125)^2 + (36-14.8125)^2\right]}$$

$$= \sqrt{\frac{1}{3}(190.785 + 73.3164 + 1.4102 + 448.910)}$$

$$= 15.4318 \quad (1)$$

$$\mu_2 = \frac{1 + 1.5811 + 2 + 2.4495}{4}$$

$$\mu_2 = 1.7577$$ 0.5

$$\sigma_2 = \sqrt{\frac{1}{4-1}\left[(1-1.7577)^2 + (1.581-1.7577)^2 + (2-1.7577)^2 + (2.4495-1.7577)^2\right]}$$ 1.5

$$\sigma_2 = 0.6171$$ 1

$x_1$

① $\frac{1-14.8125}{15.4318} = -0.8951$

② $\frac{6.25-14.8125}{15.4318} = -0.549$

③ $\frac{16-14.8125}{15.4318} = 0.077$ 2

④ $\frac{36-14.8125}{15.4318} = 1.373$

$x_2$ 2

① $\frac{1-1.7577}{0.6171} = -1.228$

② $\frac{1.5811-1.7577}{0.6171} = -0.286$

③ 0.3927

④ 1.121

(3) $\alpha = 0.75$ $\quad N = 1$

$\lambda = 0.5$ $\quad \theta_0 = \theta_1 = \theta_2 = 1$

$$X = \begin{bmatrix} 1 & -0.8951 & -1.228 \\ 1 & -0.5549 & -0.286 \\ 1 & 0.077 & 0.3927 \\ 1 & 1.373 & 1.121 \end{bmatrix} \qquad Y = \begin{bmatrix} 2 \\ 3.75 \\ 5.2 \\ 7.9 \end{bmatrix} \quad (6)$$

<u>Iteration #1</u> :

$$\theta \leftarrow \begin{bmatrix} \theta_0 \\ (1-\alpha\frac{\lambda}{m})\theta_1 \\ (1-\alpha\frac{\lambda}{m})\theta_2 \end{bmatrix} - \frac{\alpha}{m} X^T(X\theta - Y) \quad (1)$$

$$\theta \leftarrow \begin{bmatrix} 1 \\ (1 - 0.75(\frac{0.5}{4}))(1) \\ (1 - 0.75(\frac{0.5}{4}))(1) \end{bmatrix}_{(1)} - \frac{0.75}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -0.8951 & -0.5549 & 0.077 & 1.373 \\ -1.228 & -0.286 & 0.3927 & 1.121 \end{bmatrix}_{(2)} \left( \begin{bmatrix} 1 & -0.8951 & -1.228 \\ 1 & -0.5549 & -0.286 \\ 1 & 0.077 & 0.3927 \\ 1 & 1.373 & 1.121 \end{bmatrix}_{(2)} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 3.75 \\ 5.2 \\ 7.9 \end{bmatrix} \right)$$

3

$$X\theta - Y \Rightarrow \begin{bmatrix} -3.1228 \\ -3.5909 \\ -3.7303 \\ -4.406 \end{bmatrix}$$

1

$$\theta \leftarrow \begin{bmatrix} 1 \\ 0.90625 \\ 0.90625 \end{bmatrix} - \frac{0.75}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -0.8951 & -0.5549 & 0.077 & 1.373 \\ -1.228 & -0.286 & 0.3927 & 1.12 \end{bmatrix} \begin{bmatrix} -3.1228 \\ -3.5909 \\ -3.7303 \\ -4.406 \end{bmatrix}$$

1

$$\theta \leftarrow \begin{bmatrix} 1 \\ 0.90625 \\ 0.90625 \end{bmatrix} - \begin{bmatrix} -2.7846 \\ -0.2904 \\ -0.2894 \end{bmatrix} = \begin{bmatrix} 3.7848 \\ 1.1967 \\ 1.1957 \end{bmatrix}$$ 1

3

$$J(\theta) \Rightarrow \quad X\theta - Y = \begin{bmatrix} 1 & -0.6951 & -1.2 \\ 1 & -0.5549 & -0.2861 \\ 1 & 0.077 & 0.3927 \\ 1 & 1.371 & 1.121 \end{bmatrix} \begin{bmatrix} 3.7858 \\ 1.1967 \\ 1.1957 \end{bmatrix} - \begin{bmatrix} 2 \\ 3.75 \\ 5.2 \\ 7.9 \end{bmatrix}$$

$$= \begin{bmatrix} -0.7546 \\ -0.9716 \\ -0.8540 \\ -1.1323 \end{bmatrix}$$

2

$$J(\theta) = \frac{1}{2(4)}\left((0.7546)^2 + (0.9716)^2 + (0.8540)^2 + (1.1323)^2\right) +$$

$$\frac{0.5}{2(4)}\left((1.1967)^2 + (1.1957)^2\right) = 0.6194 \quad 1$$

(2)

(4) 2nA ⇒ Input becomes $x = [1 \quad 4 \quad \sqrt{2}]$

<u>Normalized:</u> $x = \left[1 \quad \dfrac{4-14.8125}{15.4318} \quad \dfrac{\sqrt{2}-1.7577}{0.6191}\right]$

(1) $x = [1 \quad -0.701 \quad 0.548]$

3nA ⇒ Input becomes $x = [1 \quad 9 \quad \sqrt{3}]$

<u>Normalized:</u> $x = \left[1 \quad \dfrac{9-14.8125}{15.4318} \quad \dfrac{\sqrt{3}-1.7577}{0.6191}\right]$

(1) $x = [1 \quad -0.3766 \quad -0.041]$

$(3.5)$

$\therefore$ Predictions are:

$$\begin{bmatrix} 1 & -0.701 & 65548 \\ 1 & -0.3766 & -0.041 \end{bmatrix} \begin{bmatrix} 3.7848 \\ 1.1967 \\ 1.1957 \end{bmatrix}$$

$(1)$

$$= \begin{bmatrix} 2.2806 \; \checkmark \\ 3.284 \; \checkmark \end{bmatrix} \quad (1)$$

---

(Q3) $$\mu_1 = \frac{37 + 37.2 + 36.8 + 37.3 + 38 + 38.5}{6} = 37.467 \quad (0.5)$$

$$\sigma_1 = \sqrt{\frac{1}{6-1}\Big((37-37.467)^2 + (37.2-37.467)^2 + (36.8-37.467)^2 + (37.3-37.467)^2 + (38-37.467)^2 + (38.5-37.467)^2} \quad (1)$$

1.5

$\sigma_1 = 0.6501$

$\mu_2 = \dfrac{3+2+1+1+1+0}{6} = 1.33$ 0.5

$\sigma_2 = \sqrt{\frac{1}{6-1}\left((3-1.33)^2 + (2-1.33)^2 - (1-1.33)^2 + (1-1.33)^2 + (1-1.33)^2 + (-1.33)^2\right.}$

$\sigma_2 = 1.0328$ 1

(6)

| Example | $x_1$ | $x_2$ |
| --- | --- | --- |
| 1 | $\frac{37 - 37.5167}{0.6501} = -0.7178$ | $\frac{3 - 1.33}{1.0328} = 1.6137$ |
| 2 | $\frac{37.2 - 37.5167}{0.6501} = -0.4102$ | $\frac{2 - 1.33}{1.0328} = 0.6455$ |
| 3 | $-1.0254$ (3) | $-0.3227$ (3) |
| 4 | $-0.2564$ | $-0.3227$ |
| 5 | $0.8204$ | $-0.3227$ |
| 6 | $1.5894$ | $-1.2910$ |

6

(2) $\alpha = 1$ $\quad N = 1$

$\theta_0 = \theta_1 = \theta_2 = 0.5$

$$X = \begin{bmatrix} 1 & -0.7178 & 1.6137 \\ 1 & -0.4102 & 0.6455 \\ 1 & -1.0254 & -0.3227 \\ 1 & -0.2564 & -0.3227 \\ 1 & 0.8204 & -0.3227 \\ 1 & 1.5894 & -1.2910 \end{bmatrix}$$

$$X\theta = \begin{bmatrix} 0.948 \\ 0.6177 \\ -0.1741 \\ 0.2104 \\ 0.7488 \\ 0.6492 \end{bmatrix}$$

$\rightarrow (0.5)(1) + (-0.7178)(0.5) + (1.6137)(0.5)$ (2)

$\rightarrow (0.5)(1) + (-0.4102)(0.5) + (0.6455)(0.5)$

Recall:

$$\theta \leftarrow \theta - \frac{\alpha}{m} X^T(g(X\theta) - y)$$ (2)

$$g(X\theta) = \begin{bmatrix} 0.7207 \\ 0.6497 \\ 0.4566 \\ 0.5524 \\ 0.6789 \\ 0.6568 \end{bmatrix}$$ (2)

8

For $\Theta_1$, $y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ 1

$$\Theta \leftarrow \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -0.7178 & -0.4102 & -1.0254 & -0.2564 & 0.8204 & 1.5894 \\ 1.6137 & 0.6455 & -0.3227 & -0.3227 & -0.3227 & -1.2910 \end{bmatrix} \begin{bmatrix} -0.2793 \\ -0.3503 \\ 0.456 \\ 0.524 \\ 0.6789 \\ 0.6568 \end{bmatrix}$$

1 2 1

$$\Theta \leftarrow \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - \begin{bmatrix} 0.2859 \\ 0.2228 \\ -0.3449 \end{bmatrix} = \begin{bmatrix} 0.2141 \\ 0.2775 \\ 0.8449 \end{bmatrix}$$

2 1

8

For $\Theta_2 \rightarrow Y = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} 1$

$$\Theta \leftarrow \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}_{1} - \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -0.7178 & -0.4102 & -1.0254 & -0.2564 & 0.8204 & 1.5894 \\ 1.6137 & 0.6455 & -0.3227 & -0.3227 & -0.3227 & -1.2910 \end{bmatrix}_{2} \begin{bmatrix} 0.7207 \\ 0.6497 \\ -0.5434 \\ -0.4476 \\ 0.6789 \\ 0.6568 \end{bmatrix}_{1}$$

$$\Theta \leftarrow \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - \begin{bmatrix} 0.2859 \\ 0.2482 \\ 0.1392 \end{bmatrix}_{2} = \begin{bmatrix} 0.2141 \\ 0.2518 \\ 0.3608 \end{bmatrix}_{1}$$

8

$$\text{for } \Theta_3 \rightarrow Y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad 1$$

$$\Theta \leftarrow \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} 1 - \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -0.7178 & -0.4102 & -1.0254 & -0.2564 & 0.8204 & 1.5894 \\ 1.6137 & 0.6455 & -0.3227 & -0.3227 & -0.3227 & -1.2910 \end{bmatrix} 2 \begin{bmatrix} 0.7207 \\ 0.6497 \\ 0.4566 \\ 0.5524 \\ -0.3211 \\ -0.3432 \end{bmatrix} 1$$

$$\Theta \leftarrow \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - \begin{bmatrix} 0.2859 \\ -0.3671 \\ 0.3006 \end{bmatrix} 2 = \begin{bmatrix} 0.2141 \\ 0.8671 \\ 0.1994 \end{bmatrix} 1$$

(B)

| $x_1$ | $x_2$ |
|---|---|
| 37.5 | 2 |
| 38 | 0 |

$$\Rightarrow \begin{bmatrix} 1 & \frac{37.5-37.4667}{0.6501} & \frac{2-1.33}{1.0328} \\ 1 & \frac{38-37.4667}{0.6501} & \frac{-1.33}{1.0328} \end{bmatrix} \quad (3)$$

Normalized $\Rightarrow \begin{bmatrix} 1 & 0.0513 & 0.6455 \\ & & \\ 1 & 0.8204 & -1.291 \end{bmatrix}$ (1) (1)

$$\therefore \quad X\,\Theta = \begin{bmatrix} 1 & 0.0513 & 0.6455 \\ 1 & 0.8204 & -1.291 \end{bmatrix} \begin{bmatrix} 0.2141 & 0.2141 & 0.2141 \\ 0.2775 & 0.2518 & 0.8671 \\ 0.8449 & 0.3608 & 0.1994 \end{bmatrix}$$

(1)

3

$$X0 = \begin{bmatrix} 0.7738 & 0.46 & 0.3873 \\ -0.6491 & -0.0451 & 0.668 \end{bmatrix}$$ 1

$$g(X0) = \begin{bmatrix} \boxed{0.6843} & 0.6130 & 0.5956 \\ 0.3432 & 0.4887 & \boxed{0.661} \end{bmatrix}$$

First input belongs to class 1 1

Second input belongs to class 3 1

$\text{Q4}$ $\quad$ (1) $\quad P(\omega_1) = P(\omega_2) = \frac{1}{2}$ $\quad$ (2) $\qquad$ (6)

(2) $\quad \mu_1 = \begin{bmatrix} \dfrac{-0.6757 + 1.5644 + 1.2643 + 1.168 - 0.535}{5} \\ \\ \dfrac{1.1287 + 0.0891 - 0.8235 - 0.7951 + 0.1214}{5} \end{bmatrix} = \begin{bmatrix} 0.5572 \\ \\ -0.0571 \end{bmatrix}$ (2)

$\mu_2 = \begin{bmatrix} \dfrac{0.8663 + 0.4452 + 0.7014 + 0.6513 - 0.105}{5} \\ \\ \dfrac{0.5371 + 1.8456 - 0.5329 + 1.09 + 0.8343}{5} \end{bmatrix} = \begin{bmatrix} 0.5118 \\ \\ 0.7548 \end{bmatrix}$ (2)

3

(3) $\Sigma = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 2 \end{bmatrix}$ 1

(4) 1
$$g_1(x) = -\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1) + \cancel{\log P(\omega_1)}$$
$$g_2(x) = -\frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2) + \cancel{\log P(\omega_2)}$$
cancels out when comparing

$$\Sigma^{-1} = \frac{1}{3.75}\begin{bmatrix} 2 & -0.5 \\ -0.5 & 2 \end{bmatrix}$$ 1

a) $x = [0.5, -0.5]$ $\Rightarrow$ $(x-\mu_1) = [(0.5,-0.5) - (0.5572, -0.0571)]$
$= [-0.0572, 0.4472]^T$
$(x-\mu_2) = [(0.5,-0.5), (0.5118, 0.7548)]$
$= [-0.0118, -1.2548]^T$

③

$$\therefore g_1(x) = \frac{-1}{2}\begin{bmatrix} -0.0572 & 0.4472 \end{bmatrix}\left(\frac{1}{3.75}\begin{bmatrix} 2 & -0.5 \\ -0.5 & 2 \end{bmatrix}\right)\begin{bmatrix} -0.0572 \\ 0.4472 \end{bmatrix}$$ ①

$$= \frac{-1}{2(3.75)}\begin{bmatrix} 0.0286 & -0.2286 \end{bmatrix}\begin{bmatrix} 0.0013 \\ -0.2608 \end{bmatrix} = -0.0448$$

$$g_2(x) = \frac{-1}{2}\begin{bmatrix} -0.0118 & -1.2548 \end{bmatrix}\left(\frac{1}{3.75}\begin{bmatrix} 2 & -0.5 \\ -0.5 & 2 \end{bmatrix}\right)\begin{bmatrix} -0.0118 \\ -1.2548 \end{bmatrix}$$ ①

$$g_2(x) = -0.4179$$ ①

$g_1(x) > g_2(x) \quad \therefore \omega_1$ is the class

(3)

b) $x = [1, 1]$ $\quad (X - \mu_1) = ([1,1) - (0.5572, -0.0571))$

$= (0.4426, 1.0571)$

$(X - \mu_2) = ([1,1) - (0.5572, -0.0571))$

$= (0.4882, 0.2452)$

$$g_1(x) = \frac{-1}{2}[0.4428 \quad 1.0571] \left[\frac{1}{3.75}\begin{pmatrix} 2 & -0.5 \\ -0.5 & 2 \end{pmatrix}\right]\begin{bmatrix} 0.4428 \\ 1.0571 \end{bmatrix}$$ (1)

$= -0.2878$

$$g_2(x) = [0.4297 \quad 0.0631] \left[\frac{1}{3.75}\begin{pmatrix} 2 & -0.5 \\ -0.5 & 2 \end{pmatrix}\right]\begin{bmatrix} 0.4297 \\ 0.0631 \end{bmatrix}$$ (1)

$= -0.0636$ $\quad g_1(x) < g_2(x) \quad \therefore \omega_2$ is the class (1)