21S-COMSCIM146-1 Exam 1

NEVIN LIANG

TOTAL POINTS

94 / 100

QUESTION 1

25 pts

1.1 4 / 4

✓ **+ 4 pts** **Correct**

+ 2 pts One sub-question correct

+ 0 pts Incorrect

1.2 8 / 8

✓ **+ 8 pts** **All Correct**

+ 6 pts 3 Steps correct

+ 4 pts 2 Steps correct

+ 2 pts 1 Steps correct

+ 6 pts w1 and w2 correct

+ 0 pts Incorrect

1.3 4 / 4

✓ **+ 4 pts** **Correct**

+ 2 pts Correct with wrong justification

+ 0 pts Incorrect

1.4 4 / 4

✓ **+ 4 pts** **Correct**

+ 2 pts Correct without mentioning linearly seperable

+ 1 pts Incorrect with some reasonable justification

+ 0 pts Incorrect

- 1 pts Minor mistake

1.5 3 / 5

+ 5 pts Correct

✓ **+ 1 pts** **Equation (partially) Correct**

✓ **+ 2 pts** **Numerator correct**

+ 2 pts Denominator correct

+ 0 pts Incorrect

- 1 pts Numerical error

QUESTION 2

20 pts

2.1 6 / 6

✓ **+ 6 pts** **Correct**

+ 4 pts X correct

+ 2 pts y correct

+ 0 pts Incorrect

2.2 8 / 8

Using formula for vector case

+ 8 pts All correct

+ 4 pts Equation for closed form w correct

- 2 pts Minor calculation mistake

+ 0 pts Incorrect

- 2 pts Some steps are wrong

Using formula for scalar case

✓ **+ 8 pts** **Correct**

+ 4 pts w1 Correct

+ 4 pts w0 Correct

+ 2 pts w1 equation correct but calculation wrong

+ 2 pts w0 equation correct but calculation wrong

2.3 6 / 6

✓ **+ 6 pts** **Correct**

+ 3 pts Points are correct

+ 3 pts Line is correct

- 2 pts Line is off due to wrong answer in (b)

+ 0 pts Incorrect

QUESTION 3

25 pts

3.1 6 / 6

✓ **+ 6 pts** **Correct**

+ 3 pts Partially correct

+ **0 pts** Incorrect

3.2 7 / 7

✓ **+ 7 pts Correct**

+ **4 pts** Partially correct

+ **2 pts** Correct answer without derivation

+ **2 pts** Few steps are correct

+ **1 pts** Partially correct answer without derivation

- **2 pts** Answer Off by a factor

- **1 pts** Minor mistake

+ **0 pts** Incorrect

3.3 6 / 6

✓ **+ 6 pts Correct**

+ **3 pts** Partially correct

+ **0 pts** Incorrect

- **1 pts** Minor mistake

3.4 2 / 6

+ **6 pts** Correct

+ **4 pts** Incorrect due to wrong factor before regularization term

✓ **+ 2 pts Incorrect but shows understanding of SGD**

+ **0 pts** Incorrect

QUESTION 4

4 10 / 10

✓ **+ 10 pts Correct**

+ **0 pts** Incorrect

QUESTION 5

20 pts

5.1 2 / 2

✓ **+ 2 pts Correct**

+ **1 pts** Partially correct

+ **0 pts** Incorrect

5.2 8 / 8

✓ **+ 8 pts All Correct**

+ **6 pts** 3/4 conditional entropy correct

+ **4 pts** 2/4 condtional entropy correct

+ **2 pts** 1/4 condtional entropy correct

- **2 pts** Major calculation error(s)

- **1 pts** Minor calculation error(s)

+ **0 pts** Incorrect

5.3 4 / 4

✓ **+ 4 pts Correct**

+ **2 pts** Largely incorrect due to error(s) in (b)

+ **0 pts** Incorrect

- **2 pts** No numerical results

5.4 2 / 2

✓ **+ 2 pts Correct**

+ **0 pts** Incorrect

5.5 4 / 4

✓ **+ 4 pts Both Correct**

+ **0 pts** Incorrect

I Neem Liang with UID 705575353 have read and understood the policy on academic integrity

1. (a) i) $w^{t+1} = w^t + X$

ii) $w^{t+1} = w^t - X$

(b) Starting weights: $w = [0,1,1]$

Update based on $[10,10]$: $\begin{bmatrix} 1 \\ 10 \\ 10 \end{bmatrix}$

$[0 \ 1 \ 1]\begin{bmatrix} 1 \\ 10 \\ 10 \end{bmatrix} = 20$

$20 \cdot y = 20 \cdot 1 > 0$ no change

back on $[1,0]$: $[0 \ 1 \ 1]\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 1$

$1 \cdot -1 = -1 < 0$ change.

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

back on $[3,3]$: $[-1 \ 0 \ 1]\begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} = 2$

$2 \cdot -1 < 0$ change

$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \\ -2 \end{bmatrix}$

back on $[4,8]$: $[-2 \ -3 \ -2]\begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix} = -30$

$-30 \cdot 1 < 0$ change

$\begin{bmatrix} -2 \\ -3 \\ -2 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 6 \end{bmatrix}$

1.1 4 / 4

✓ **+ 4 pts** **Correct**

+ 2 pts One sub-question correct

+ 0 pts Incorrect

I Neem Liang with UID 705575353 have read and understood the policy on academic integrity

1. (a) i) $w^{t+1} = w^t + X$

ii) $w^{t+1} = w^t - X$

(b) Starting weights: $w = [0,1,1]$

Update based on $[10,10]$: $\begin{bmatrix} 1 \\ 10 \\ 10 \end{bmatrix}$

$[0 \ 1 \ 1]\begin{bmatrix} 1 \\ 10 \\ 10 \end{bmatrix} = 20$

$20 \cdot y = 20 \cdot 1 > 0$ no change

back on $[1,0]$: $[0 \ 1 \ 1]\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 1$

$1 \cdot -1 = -1 < 0$ change.

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

back on $[3,3]$: $[-1 \ 0 \ 1]\begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} = 2$

$2 \cdot -1 < 0$ change

$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \\ -2 \end{bmatrix}$

back on $[4,8]$: $[-2 \ -3 \ -2]\begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix} = -30$

$-30 \cdot 1 < 0$ change

$\begin{bmatrix} -2 \\ -3 \\ -2 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 6 \end{bmatrix}$

1.2 **8 / 8**

✓ **+ 8 pts** **All Correct**

+ 6 pts 3 Steps correct

+ 4 pts 2 Steps correct

+ 2 pts 1 Steps correct

+ 6 pts w1 and w2 correct

+ 0 pts Incorrect

(c) no. did not converge.

$w = \begin{bmatrix} -1 \\ 1 \\ 6 \end{bmatrix}$ does not satisfy sample #3.

$$\begin{bmatrix} -1 & 1 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = 20$$

$$20 \cdot -1 < 0$$

still have to keep changing

(d) yes. the perceptron algo converges if there is a line separating the data points and all values are bounded.

$$\text{Max iters} = R^2/\mu^2$$

(e) distance between $-10 + 2x_1 + x_2 = y$

$$\frac{\left| \begin{bmatrix} -10 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right|}{\sqrt{(-10)^2 + 2^2 + 1^2}} = \frac{|-11|}{\sqrt{105}} = \boxed{\frac{1}{\sqrt{105}}}$$

1.3 **4 / 4**

✓ **+ 4 pts** **Correct**

+ 2 pts Correct with wrong justification

+ 0 pts Incorrect

(c) no. did not converge.

$w = \begin{bmatrix} -1 \\ 1 \\ 6 \end{bmatrix}$ does not satisfy sample #3.

$$\begin{bmatrix} -1 & 1 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = 20$$

$$20 \cdot -1 < 0$$

still have to keep changing

(d) yes. the perceptron algo converges if there is a line separating the data points and all values are bounded.

$$\text{Max iters} = R^2/\mu^2$$

(e) distance between $-10 + 2x_1 + x_2 = y$

$$\frac{\left| \begin{bmatrix} -10 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right|}{\sqrt{(-10)^2 + 2^2 + 1^2}} = \frac{|-11|}{\sqrt{105}} = \boxed{\frac{1}{\sqrt{105}}}$$

1.4 4 / 4

✓ **+ 4 pts** **Correct**

+ 2 pts Correct without mentioning linearly seperable

+ 1 pts Incorrect with some reasonable justification

+ 0 pts Incorrect

- 1 pts Minor mistake

(c) no. did not converge.

$w = \begin{bmatrix} -1 \\ 1 \\ 6 \end{bmatrix}$ does not satisfy sample #3.

$$\begin{bmatrix} -1 & 1 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = 20$$

$$20 \cdot -1 < 0$$

still have to keep changing

(d) yes. the perceptron algo converges if there is a line separating the data points and all values are bounded.

$$\text{Max iters} = R^2/\mu^2$$

(e) distance between $-10 + 2x_1 + x_2 = y$

$$\frac{\left| \begin{bmatrix} -10 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right|}{\sqrt{(-10)^2 + 2^2 + 1^2}} = \frac{|-11|}{\sqrt{105}} = \boxed{\frac{1}{\sqrt{105}}}$$

1.5 **3 / 5**

+ **5 pts** Correct

✓ **+ 1 pts Equation (partially) Correct**

✓ **+ 2 pts Numerator correct**

+ **2 pts** Denominator correct

+ **0 pts** Incorrect

- **1 pts** Numerical error

(2)

(a)

X is $\begin{bmatrix} 0 & 2 & 4 \end{bmatrix}$

y is $\begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$

(b)

$w_1 = \dfrac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$

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

$= \left(\begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \begin{bmatrix} 0 & 2 & 4 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \cdot$

$\bar{x} = 2$

$\bar{y} = 4/3$

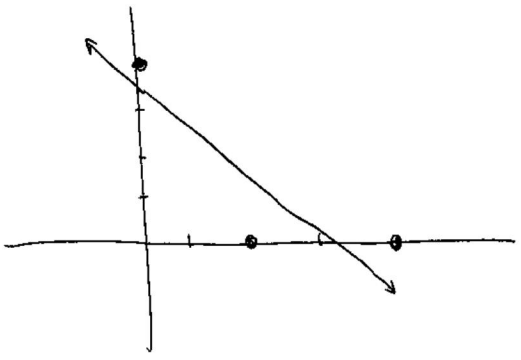
$w_1 = \dfrac{(-2)(8/3) + 0 \cdot 4 + 2 \cdot (-4/3)}{2^2 + 0^2 + 2^2} = \dfrac{-8}{8} = -1$

$w_0 = \bar{y} - w_1 \bar{x} = \frac{4}{3} - (-1) \cdot 2 = 10/3$

$w = \begin{bmatrix} 10/3 \\ -1 \end{bmatrix}$

$\hat{y} = -x + \frac{10}{3}$

(c)



2.1 **6 / 6**

✓ **+ 6 pts** **Correct**

+ 4 pts X correct

+ 2 pts y correct

+ 0 pts Incorrect

(2)

(a)

X is $\begin{bmatrix} 0 & 2 & 4 \end{bmatrix}$

y is $\begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$

(b)

$w_1 = \dfrac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$

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

$= \left(\begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \begin{bmatrix} 0 & 2 & 4 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \cdot$

$\bar{x} = 2$

$\bar{y} = 4/3$

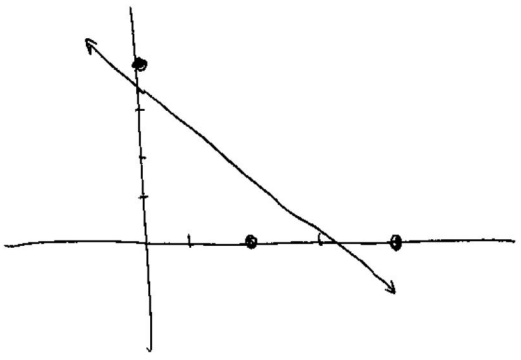
$w_1 = \dfrac{(-2)(8/3) + 0 \cdot 4 + 2 \cdot (-4/3)}{2^2 + 0^2 + 2^2} = \dfrac{-8}{8} = -1$

$w_0 = \bar{y} - w_1 \bar{x} = \frac{4}{3} - (-1) \cdot 2 = 10/3$

$w = \begin{bmatrix} 10/3 \\ -1 \end{bmatrix}$

$\hat{y} = -x + \frac{10}{3}$

(c)



2.2 8 / 8

Using formula for vector case

+ **8 pts** All correct
+ **4 pts** Equation for closed form w correct
- **2 pts** Minor calculation mistake
+ **0 pts** Incorrect
- **2 pts** Some steps are wrong

Using formula for scalar case

✓ + **8 pts** **Correct**
+ **4 pts** w1 Correct
+ **4 pts** w0 Correct
+ **2 pts** w1 equation correct but calculation wrong
+ **2 pts** w0 equation correct but calculation wrong

(2)

(a)

X is $\begin{bmatrix} 0 & 2 & 4 \end{bmatrix}$

y is $\begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$

(b)

$w_1 = \dfrac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$

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

$= \left(\begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \begin{bmatrix} 0 & 2 & 4 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \cdot$

$\bar{x} = 2$

$\bar{y} = 4/3$

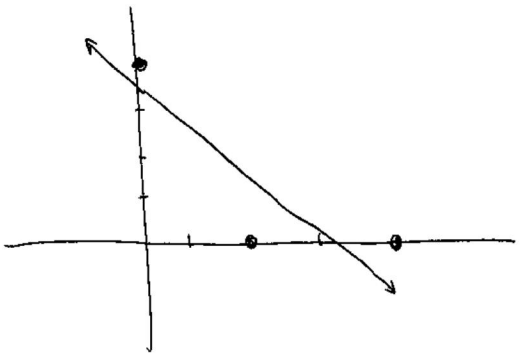
$w_1 = \dfrac{(-2)(8/3) + 0 \cdot 4 + 2 \cdot (-4/3)}{2^2 + 0^2 + 2^2} = \dfrac{-8}{8} = -1$

$w_0 = \bar{y} - w_1 \bar{x} = \frac{4}{3} - (-1) \cdot 2 = 10/3$

$w = \begin{bmatrix} 10/3 \\ -1 \end{bmatrix}$

$\hat{y} = -x + \frac{10}{3}$

(c)



2.3 **6 / 6**

✓ **+ 6 pts** **Correct**

+ 3 pts Points are correct

+ 3 pts Line is correct

- 2 pts Line is off due to wrong answer in (b)

+ 0 pts Incorrect

3)

(a) $J(w) = \frac{1}{2} \| Xw - y \|^2 + \frac{\lambda}{2} \cdot \|w\|^2$

(b) $\nabla_w (J(w)) = \nabla_w \left(\frac{1}{2} (Xw-y)^T (Xw-y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} \cdot (w^T X^T - y^T)(Xw-y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} (w^T X^T X w - w^T X^T y - y^T X w + y^T y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} \cdot w^T X^T X w - \frac{1}{2} w^T X^T y + \frac{1}{2} y^T y + \frac{\lambda}{2} w^T w \right)$

matrix rules on discussion 2 →

$= X^T X w - X^T y + \lambda w$

$X^T X w - X^T y + \lambda w = 0$

$(X^T X + \lambda I) w = X^T y$

$$\boxed{w = (X^T X + \lambda I)^{-1} \cdot X^T y}$$

(c) $\nabla_w (J(w)) = X^T X w - X^T y + \lambda w$

~~$\frac{\partial J}{\partial w_j} = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial w_j} (w^T x_i - y_i) + \frac{\lambda}{2} \cdot 2 w_j$~~

$X^T X w$ j^{th} row = j^{th} row of $X^T \cdot X \cdot w$

$$\left(\begin{bmatrix} x_{1,j} & x_{2,j} & x_{3,j} & \cdots & x_{n,j} \end{bmatrix} \cdot X \cdot w \right.$$

$$\left. - \begin{bmatrix} x_{1,j} & x_{2,j} & x_{3,j} & \cdots & x_{n,j} \end{bmatrix} \cdot y + \lambda w_j \right)$$

3.1 **6 / 6**

✓ **+ 6 pts** **Correct**

+ 3 pts Partially correct

+ 0 pts Incorrect

3)

(a) $J(w) = \frac{1}{2} \| Xw - y \|^2 + \frac{\lambda}{2} \cdot \|w\|^2$

(b) $\nabla_w (J(w)) = \nabla_w \left(\frac{1}{2} (Xw-y)^T (Xw-y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} \cdot (w^T X^T - y^T)(Xw-y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} (w^T X^T X w - w^T X^T y - y^T X w + y^T y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} \cdot w^T X^T X w - \frac{1}{2} w^T X^T y + \frac{1}{2} y^T y + \frac{\lambda}{2} w^T w \right)$

matrix rules on discussion 2 →

$= X^T X w - X^T y + \lambda w$

$X^T X w - X^T y + \lambda w = 0$

$(X^T X + \lambda I) w = X^T y$

$$\boxed{w = (X^T X + \lambda I)^{-1} \cdot X^T y}$$

(c) $\nabla_w (J(w)) = X^T X w - X^T y + \lambda w$

~~$\frac{\partial J}{\partial w_j} = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial w_j} (w^T x_i - y_i) + \frac{\lambda}{2} \cdot 2 w_j$~~

$X^T X w$ j^{th} row = j^{th} row of $X^T \cdot X \cdot w$

$$\left(\begin{bmatrix} x_{1,j} & x_{2,j} & x_{3,j} & \cdots & x_{n,j} \end{bmatrix} \cdot X \cdot w \right.$$

$$\left. - \begin{bmatrix} x_{1,j} & x_{2,j} & x_{3,j} & \cdots & x_{n,j} \end{bmatrix} \cdot y + \lambda w_j \right)$$

3.2 **7 / 7**

✓ **+ 7 pts** **Correct**

+ 4 pts Partially correct

+ 2 pts Correct answer without derivation

+ 2 pts Few steps are correct

+ 1 pts Partially correct answer without derivation

- 2 pts Answer Off by a factor

- 1 pts Minor mistake

+ 0 pts Incorrect

3)

(a) $J(w) = \frac{1}{2} \| Xw - y \|^2 + \frac{\lambda}{2} \cdot \|w\|^2$

(b) $\nabla_w (J(w)) = \nabla_w \left(\frac{1}{2} (Xw-y)^T (Xw-y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} \cdot (w^T X^T - y^T)(Xw-y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} (w^T X^T X w - w^T X^T y - y^T X w + y^T y) + \frac{\lambda}{2} w^T w \right)$

$= \frac{\partial}{\partial w} \left(\frac{1}{2} \cdot w^T X^T X w - \frac{1}{2} w^T X^T y + \frac{1}{2} y^T y + \frac{\lambda}{2} w^T w \right)$

matrix rules on discussion 2 →

$= X^T X w - X^T y + \lambda w$

$X^T X w - X^T y + \lambda w = 0$

$(X^T X + \lambda I) w = X^T y$

$$\boxed{w = (X^T X + \lambda I)^{-1} \cdot X^T y}$$

(c) $\nabla_w (J(w)) = X^T X w - X^T y + \lambda w$

~~$\frac{\partial J}{\partial w_j} = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial w_j} (w^T x_i - y_i) + \frac{\lambda}{2} \cdot 2 w_j$~~

$X^T X w$ j^{th} row = j^{th} row of $X^T \cdot X \cdot w$

$$\left(\begin{bmatrix} x_{1,j} & x_{2,j} & x_{3,j} & \cdots & x_{n,j} \end{bmatrix} \cdot X \cdot w \right.$$

$$\left. - \begin{bmatrix} x_{1,j} & x_{2,j} & x_{3,j} & \cdots & x_{n,j} \end{bmatrix} \cdot y + \lambda w_j \right)$$

Let A be the row vector

$$\begin{bmatrix} X_{1j} & X_{2j} & X_{3j} & \cdots & X_{nj} \end{bmatrix}$$

$$\nabla_{w_j}(J(w)) = AX_w - Ay + \lambda w_j = (w^T X^T - y^T)(A^T) + \lambda w_j$$

(d) ~~$w_{t+1} = w_t - \eta \cdot [AX_w - Ay + \lambda w_j]$~~

no need to divide by norm (problem said so)

~~$w_{t+1} = w_t - \eta \cdot$~~

$$\nabla_{w_j}(J(w)) = \left(w^T X^T - y^T\right) \cdot (j^{th} \text{ colum of } X) + \lambda w_j$$

$$w_{t+1} = w_t - \eta \left(\left(w_{t,j} \cdot X_{i,j} - y_i \right)\left(X_{i,j} \right) + \lambda w_j \right)$$

3.3 **6 / 6**

✓ **+ 6 pts** **Correct**

+ 3 pts Partially correct

+ 0 pts Incorrect

- 1 pts Minor mistake

Let A be the row vector

$$\begin{bmatrix} X_{1j} & X_{2j} & X_{3j} & \cdots & X_{nj} \end{bmatrix}$$

$$\nabla_{w_j}(J(w)) = AX_w - Ay + \lambda w_j = (w^T X^T - y^T)(A^T) + \lambda w_j$$

(d) ~~$w_{t+1} = w_t - \eta \cdot [AX_w - Ay + \lambda w_j]$~~

no need to divide by norm (problem said so)

~~$w_{t+1} = w_t - \eta \cdot$~~

$$\nabla_{w_j}(J(w)) = \left(w^T X^T - y^T\right) \cdot (j^{th} \text{ colum of } X) + \lambda w_j$$

$$w_{t+1} = w_t - \eta \left(\left(w_{t,j} \cdot X_{i,j} - y_i \right)\left(X_{i,j} \right) + \lambda w_j \right)$$

3.4 **2 / 6**

+ **6 pts** Correct

+ **4 pts** Incorrect due to wrong factor before regularization term

✓ + **2 pts** **Incorrect but shows understanding of SGD**

+ **0 pts** Incorrect

4.

$$\sigma'(x) = \frac{\partial}{\partial x}\left(\frac{1}{1+e^{-x}}\right) = -1(1+e^{-x})^{-2}\cdot(-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\sigma(x)(1-\sigma(x)) = \frac{1}{1+e^{-x}}\cdot\left(1-\frac{1}{1+e^{-x}}\right) = \frac{1\cdot e^{-x}}{(1+e^{-x})^2}$$ ✓

4 **10 / 10**

✓ **+ 10 pts** **Correct**

+ 0 pts Incorrect

5] a) $H(X\text{Disease}) = H\left(\frac{5}{8}\right) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8}$

$= \boxed{0.954}$

b) Fatisve: $\frac{4}{8}\cdot H\left(\frac{2}{4}\right) + \frac{4}{8}\cdot H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}\cdot 0.811 = \boxed{0.906}$

Fever: $\frac{4}{8}H\left(\frac{1}{4}\right) + \frac{4}{8}H(1) = \boxed{0.406}$

Cough: $\frac{5}{8}H\left(\frac{2}{5}\right) + \frac{3}{8}H(1) = \boxed{0.607}$

Headache: $\frac{3}{8}H\left(\frac{2}{3}\right) + \frac{5}{8}H\left(\frac{3}{5}\right) = \boxed{0.951}$

c) $I(X\text{Dixm}; X)$

If X=

Fatisve: $0.954 - 0.906 = 0.048$

Fever: $0.954 - 0.406 = 0.548$

Cough: $0.954 - 0.607 = 0.347$

Headache: $0.954 - 0.951 = 0.003$

d) Fever is max so FEVER

e) 9 has XDisease

10 does not.

5.1 2 / 2

✓ **+ 2 pts** **Correct**

+ 1 pts Partially correct

+ 0 pts Incorrect

5] a) $H(X\text{Disease}) = H\left(\frac{5}{8}\right) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8}$

$= \boxed{0.954}$

b) Fatisve: $\frac{4}{8}\cdot H\left(\frac{2}{4}\right) + \frac{4}{8}\cdot H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}\cdot 0.811 = \boxed{0.906}$

Fever: $\frac{4}{8}H\left(\frac{1}{4}\right) + \frac{4}{8}H(1) = \boxed{0.406}$

Cough: $\frac{5}{8}H\left(\frac{2}{5}\right) + \frac{3}{8}H(1) = \boxed{0.607}$

Headache: $\frac{3}{8}H\left(\frac{2}{3}\right) + \frac{5}{8}H\left(\frac{3}{5}\right) = \boxed{0.951}$

c) $I(X\text{Dixm}; X)$

If X=

Fatisve: $0.954 - 0.906 = 0.048$

Fever: $0.954 - 0.406 = 0.548$

Cough: $0.954 - 0.607 = 0.347$

Headache: $0.954 - 0.951 = 0.003$

d) Fever is max so FEVER

e) 9 has XDisease

10 does not.

5.2 **8 / 8**

✓ **+ 8 pts** **All Correct**

+ 6 pts 3/4 conditional entropy correct

+ 4 pts 2/4 condtional entropy correct

+ 2 pts 1/4 condtional entropy correct

- 2 pts Major calculation error(s)

- 1 pts Minor calculation error(s)

+ 0 pts Incorrect

5] a) $H(X\text{Disease}) = H\left(\frac{5}{8}\right) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8}$

$= \boxed{0.954}$

b) Fatisve: $\frac{4}{8}\cdot H\left(\frac{2}{4}\right) + \frac{4}{8}\cdot H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}\cdot 0.811 = \boxed{0.906}$

Fever: $\frac{4}{8}H\left(\frac{1}{4}\right) + \frac{4}{8}H(1) = \boxed{0.406}$

Cough: $\frac{5}{8}H\left(\frac{2}{5}\right) + \frac{3}{8}H(1) = \boxed{0.607}$

Headache: $\frac{3}{8}H\left(\frac{2}{3}\right) + \frac{5}{8}H\left(\frac{3}{5}\right) = \boxed{0.951}$

c) $I(X\text{Dixm}; X)$

If X=

Fatisve: $0.954 - 0.906 = 0.048$

Fever: $0.954 - 0.406 = 0.548$

Cough: $0.954 - 0.607 = 0.347$

Headache: $0.954 - 0.951 = 0.003$

d) Fever is max so FEVER

e) 9 has XDisease

10 does not.

5.3 **4 / 4**

✓ **+ 4 pts** **Correct**

+ 2 pts Largely incorrect due to error(s) in (b)

+ 0 pts Incorrect

- 2 pts No numerical results

5] a) $H(X\text{Disease}) = H\left(\frac{5}{8}\right) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8}$

$= \boxed{0.954}$

b) Fatisve: $\frac{4}{8}\cdot H\left(\frac{2}{4}\right) + \frac{4}{8}\cdot H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}\cdot 0.811 = \boxed{0.906}$

Fever: $\frac{4}{8}H\left(\frac{1}{4}\right) + \frac{4}{8}H(1) = \boxed{0.406}$

Cough: $\frac{5}{8}H\left(\frac{2}{5}\right) + \frac{3}{8}H(1) = \boxed{0.607}$

Headache: $\frac{3}{8}H\left(\frac{2}{3}\right) + \frac{5}{8}H\left(\frac{3}{5}\right) = \boxed{0.951}$

c) $I(X\text{Dixm}; X)$

If X=

Fatisve: $0.954 - 0.906 = 0.048$

Fever: $0.954 - 0.406 = 0.548$

Cough: $0.954 - 0.607 = 0.347$

Headache: $0.954 - 0.951 = 0.003$

d) Fever is max so FEVER

e) 9 has XDisease

10 does not.

5.4 2 / 2

✓ **+ 2 pts** Correct

+ 0 pts Incorrect

5] a) $H(X\text{Disease}) = H\left(\frac{5}{8}\right) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8}$

$= \boxed{0.954}$

b) Fatisve: $\frac{4}{8}\cdot H\left(\frac{2}{4}\right) + \frac{4}{8}\cdot H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}H\left(\frac{3}{4}\right) = \frac{4}{8} + \frac{4}{8}\cdot 0.811 = \boxed{0.906}$

Fever: $\frac{4}{8}H\left(\frac{1}{4}\right) + \frac{4}{8}H(1) = \boxed{0.406}$

Cough: $\frac{5}{8}H\left(\frac{2}{5}\right) + \frac{3}{8}H(1) = \boxed{0.607}$

Headache: $\frac{3}{8}H\left(\frac{2}{3}\right) + \frac{5}{8}H\left(\frac{3}{5}\right) = \boxed{0.951}$

c) $I(X\text{Dixm}; X)$

If X=

Fatisve: $0.954 - 0.906 = 0.048$

Fever: $0.954 - 0.406 = 0.548$

Cough: $0.954 - 0.607 = 0.347$

Headache: $0.954 - 0.951 = 0.003$

d) Fever is max so FEVER

e) 9 has XDisease

10 does not.

5.5 **4 / 4**

✓ **+ 4 pts** **Both Correct**

+ 0 pts Incorrect