HW #5

1. $L(w) = w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4$

a. $dL/dw_1 = 2w_1 + 2$

$dL/dw_2 = 4w_2 - 4$

$dL/dw_3 = 2w_3 - 2w_4$

$dL/dw_4 = -2w_3 + 2w_4$

b. $\nabla L(w) = \begin{bmatrix} 2w_1+2 \\ 4w_2-4 \\ 2w_3-2w_4 \\ -2w_3+2w_4 \end{bmatrix}$ $\begin{bmatrix} dL/dw_1 \\ dL/dw_2 \\ dL/dw_3 \\ dL/dw_4 \end{bmatrix}$

c. $w_{t+1} = w_t - \eta_t \nabla L(w_t)$ $w = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$w_{t+1} = 0 - 0.1 \begin{bmatrix} 2(0)+2 \\ 4(0)-4 \\ 2(0)-2(0) \\ -2(0)+2(0) \end{bmatrix} = \begin{bmatrix} -0.2 \\ 0.4 \\ 0 \\ 0 \end{bmatrix}$

$= \begin{bmatrix} -0.2 \\ 0.4 \\ 0 \\ 0 \end{bmatrix}$

d. $\min(L(w))$

$\nabla L(w) = 0$ $\begin{bmatrix} 2w_1+2 \\ 4w_2-4 \\ 2w_3-2w_4 \\ -2w_3+2w_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$w_1 = -1$ $w_2 = 1$ $w_3 = w_4$ $w_3 = w_4 = x$

$L(w) = (-1)^2 + 2(1)^2 + x^2 - 2x^2 + x^2 + 2(-1) - 4(1)^2 + 4$

$= \boxed{1}$

e. No, there are infinitely many solutions with the form $\begin{bmatrix} -1 \\ 1 \\ x \\ x \end{bmatrix} \forall x \in \mathbb{R}$

when $w_3 = w_4$.

2. $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, find $w \in \mathbb{R}^4$ that minimizes $L(w) = \sum_{i=1}^{n} (w \cdot x^{(i)}) + \frac{1}{2c}\|w\|^2$
$c > 0$

a. $$dL/dw_j = \sum_{i=1}^{n} x_j^{(i)} + \frac{1}{2c}\frac{dL}{dw_j}\|w\|^2$$

$$= \cancel{cw_j} + \boxed{\sum_{i=1}^{n} x_j^{(i)} + cw_j}$$

b. $$\nabla L(w) = \begin{bmatrix} \sum_{i=1}^{n} x_1^{(i)} + cw_1 \\ \sum_{i=1}^{n} x_2^{(i)} + cw_2 \\ \vdots \\ \cancel{\sum_{i=1}^{n} x_j + cw_j} \\ \vdots \\ \sum_{i=1}^{n} x_d^{(i)} + cw_d \end{bmatrix}$$

$$= \sum_{i=1}^{n} x^{(i)} + cw$$

c. $$\nabla L(w) = 0$$

$$\sum_{i=1}^{n} x^{(i)} + cw = 0$$

$$w = -\frac{\sum_{i=1}^{n} x^{(i)}}{c} = \boxed{-\frac{1}{c}\sum_{i=1}^{n} x^{(i)}}$$

~~3. a.~~ $\cancel{L(w) = \sum_{i=1}^{n} (y^{(i)} - w \cdot x^{(i)})^2 + \lambda\|w\|^2}$

$\cancel{\nabla L(w)}$ $\cancel{\frac{dL}{dw_j} = \sum_{i=1}^{n} -2y^{(i)}x^{(i)} + 2(w_j \cdot x^{(i)})^2}$

$\cancel{= 2\sum_{i=1}^{n} w_j \cdot x^{(i)2} - y^{(i)}x^{(i)}}$

$\cancel{\nabla L(w) = \begin{bmatrix} 2\sum_{i=1}^{n} w_1 \cdot x_1^{(i)2} - y^{(i)}x \\ \vdots \end{bmatrix}}$

ignoring intercept form

$\cancel{\nabla L(w) = \begin{bmatrix} 2(\sum_{i=1}^{n} w_1 \cdot x^{(i)2} - y^{(i)}x^{(i)}) \\ \vdots \\ 2(\sum_{i=1}^{n} w_d \cdot x^{(i)2} - y^{(i)}x^{(i)}) \end{bmatrix}}$

$\cancel{= 2\sum_{i=1}^{n} w \cdot x^{(i)2} -}$

3. a. $L(w) = \sum_{i=1}^{n} (y^{(i)} - w \cdot x^{(i)})^2 + \lambda \|w\|^2$

$$\frac{dL}{dw_j} \sum_{i=1}^{n} (-2y^{(i)} + 2w \cdot x^{(i)}) x_j^{(i)} + 2\lambda w$$

$$\boxed{\nabla L(w) = 2 \sum_{i=1}^{n} (w \cdot x^{(i)} - y^{(i)}) x^{(i)} + 2\lambda w}$$

b. $w_{t+1} = w_t + \eta_t \nabla L(w_t)$

$$\boxed{w_{t+1} = w_t + 2\eta \sum_{i=1}^{n} (w_t \cdot x^{(i)} - y^{(i)}) x^{(i)} - 2\eta\lambda w_t}$$

c. - set $w_0 = 0$
- for $t = 0, 1, 2, \ldots$, until convergence or fixed iterations
- ~~$w_{t+1} = w_t + 2\eta(w_t \cdot x^{(i)}$~~
- randomly select a data point $(x^{(i)}, y^{(i)})$ from the dataset
- $w_{t+1} = w_t + 2\eta(w_t \cdot x^{(i)} - y^{(i)}) x^{(i)} - 2\eta\lambda w_t$
  - update w using gradient descent based off the one point
- get final vector w

4. a. $f''(x) = 2 > 0$ $\boxed{\text{convex}}$
b. $f''(x) = -2 < 0$ $\boxed{\text{concave}}$
c. $f''(x) = 2 > 0$ $\boxed{\text{convex}}$
d. $f''(x) = 0$ ~~neither~~ $\boxed{\text{both}}$
e. $f''(x) = 6x$ $\boxed{\text{neither}}$
f. $f''(x) = 12x^2 \geq 0$ $\boxed{\text{convex}}$
g. $f'' = -\frac{1}{x^2} < 0$ $\boxed{\text{concave}}$