Dhanashree Kharate
30871817

Excercise 3.3

a. $$g(\tilde{w}) = \sum_{p=1}^{P} (\tilde{x}_p^T \tilde{w} - y_p)^2$$

$$\because (a-b)^2 = a^2 - 2ab + b2$$

$$= \sum_{p=1}^{P} \left[(\tilde{x}_p^T \tilde{w})^2 - 2(\tilde{x}_p^T \tilde{w})(y_p) + y_p^2 \right]$$

$$= \sum_{p=1}^{P} \left[(\tilde{x}_p^T \tilde{w}) \cdot (\tilde{x}_p^T \tilde{w}) - 2(\tilde{x}_p^T \tilde{w})(y_p) + y_p^2 \right]$$

$$= \sum_{p=1}^{P} \left[(\tilde{x}_p^T \cdot \tilde{w}) \cdot (\tilde{x}_p \tilde{w}^T) - 2(\tilde{x}_p^T \cdot \tilde{w}) y_p + y_p^2 \right]$$

Dividing equation by 2

$$= \sum_{p=1}^{P} \left[\frac{1}{2} (\tilde{x}_p^T \tilde{w}) \cdot (\tilde{x}_p \tilde{w}^T) - 2(\tilde{x}_p^T \cdot \tilde{w}) \cdot y_p + \frac{y_p^2}{2} \right]$$

Comparing above equation with $g(\tilde{w}) = \frac{1}{2} \tilde{w}^T Q \tilde{w} + r^T \tilde{w} + d$

$$Q = \sum_{p=1}^{P} \tilde{x}_p \cdot \tilde{x}_p^T$$

$$r = -\tilde{x}_p^T \cdot y_p$$

$$d = \frac{y_p^2}{2}$$

b. if $N \times N$ square symmetric matrix Q satisfies

$z^T Q z \geq 0$ for all values of z then it must have all non negative eigen values

$$z^T Q z \geq 0 \qquad (1)$$

Substitute $Q = \sum_{p=1}^{P} \tilde{x}_p^T \cdot x_p$ in eq (1)

$$z^T \left(\sum_{p=1}^{P} \tilde{x}_p^T \cdot \tilde{x}_p^T \right) \cdot z \geq 0$$

$$\sum_{p=1}^{P} z^T \cdot \tilde{x}_p \cdot x_p^T \cdot z \geq 0$$

$$\alpha = z^T \cdot \tilde{x}_p$$

$$\alpha^T = \tilde{x}_p^T \cdot z$$

$$\sum_{p=1}^{P} \alpha \cdot \alpha^T \geq 0$$

$$\alpha \cdot \alpha^T = \|\alpha\|^2$$

magnitude of α is always positive
i.e. non negative

c. $\nabla^2 g(\tilde{w}) = Q$

$$g(\tilde{w}) = \sum_{p=1}^{P} (\tilde{x}_p^T \tilde{w} - y_p)^2$$

$$\nabla g(\tilde{w}) = \sum_{p=1}^{P} 2(\tilde{x}_p^T \tilde{w} - y_p)\tilde{x}_p$$

$$\nabla^2 g(\tilde{w}) = 2\sum_{p=1}^{P} (\tilde{x}_p \tilde{w} \cdot x_p^T - y_p \cdot \tilde{x}_p) \quad \therefore \nabla^2 g(\tilde{w}) \geq 0$$

$$\sum_{p=1}^{P} \tilde{x}_p \tilde{x}_p^T \geq 0$$

$$\nabla^2 g(\tilde{w}) = 2\sum_{p=1}^{P} \tilde{x}_p \cdot x_p^T = Q.$$

d. By Newton's method,

$$\nabla^2 g(w^{k-1}) w^k = \nabla^2 g(w^{k-1}) w^{k-1} - \nabla g(w^{k-1}) \quad ①$$

$$g(\tilde{w}) = \sum_{p=1}^{P} (\tilde{x}_p^T \tilde{w} - y_p)^2$$

$$\nabla g(\tilde{w}) = 2\sum_{p=1}^{P} (\tilde{x}_p^T \tilde{w} - y_p)\tilde{x}_p$$

$$\nabla^2 g(\tilde{w}) = 2\sum_{p=1}^{P} \tilde{x}_p \tilde{x}_p^T \quad \text{(from 3.c.)}$$

Substituting in eq ①

$$\nabla^2 g(w^{k-1}) w^k = \nabla^2 g(w^{k-1}) w^{k-1} - \nabla g(w^{k-1})$$

$$2\sum_{p=1}^{P} \tilde{x}_p x_p^T \tilde{w} = \cancel{2\sum_{p=1}^{P} \tilde{x}_p x_p^T w} - \cancel{2\sum_{p=1}^{P} \tilde{x}_p x_p^T w} + 2\sum_{p=1}^{P} y_p \tilde{x}_p$$

$$\boxed{\left(\sum_{p=1}^{P} \tilde{x}_p \cdot \tilde{x}_p^T\right)\tilde{w} = \sum_{p=1}^{P} x_p y_p}$$

3.10.

a. $\sigma^{-1}(\sigma(t))$

$$\sigma(t) = \frac{1}{1+e^{-t}}$$

$$\Rightarrow \sigma^{-1}\left(\cancel{\frac{1}{1+e^{-t}}}\right)$$

$\because f^{-1}(t) = \log(t)$

$\cancel{f^{-1}}(\cancel{\phi(t)}) =$

$f^{-1}(f(t)) = \log e^{t} = t$

$$= \cancel{\log}\quad t = \frac{1}{1+e^{-\sigma(t)}}$$

$$t(1+\cancel{\sigma}e^{-\sigma(t)}) = 1$$

$$t + t \cdot e^{-\sigma(t)} = 1$$

~~logat~~ + ~~log~~

$$e^{-\sigma(t)} = \frac{1-t}{t}$$

$$\log e^{-\sigma(t)} = \log \frac{1-t}{t}$$

$$-\sigma(t) \log e = \log \frac{1-t}{t}$$

$$\boxed{\sigma(t) = \log \frac{t}{1-t}} \quad 1$$

$$\sigma^{-1}(\sigma(t))$$

$$\sigma^{-1}\left(\frac{1}{1+e^{-t}}\right)$$

$$= \log \frac{\frac{1}{1+e^{-t}}}{1-\frac{1}{1+e^{-t}}}$$ from 3.10.a.1

$$= \log \frac{\frac{1}{1+e^{-t}}}{\frac{\cancel{1}+e^{-t}-\cancel{1}}{1+e^{-t}}}$$

$$= \log e^{+t}$$

$$\sigma^{-1}(\sigma(t)) = t \log e$$

$$\boxed{\sigma^{-1}(\sigma(t)) = t}$$

b. Ask

$$b + x_p w \approx \log\left(\frac{y_p}{1-y_p}\right)$$

$$\sigma^{-1}(t) = \log \frac{t}{1-t}$$

$$\cancel{\sigma^{-1}(b + x_p w) = \log \times b + x_p w}$$

$$\sigma(b + x_p w) \approx y_p$$

Taking σ^{-1} on both sides

$$\sigma^{-1}(\sigma(b + x_p w)) \approx \sigma^{-1} y_p$$

From 3. a

$$\sigma^{-1}(\sigma(t)) = 1$$

$$\sigma^{-1}(t) = \log \frac{t}{1-t}$$

$$b + x_p w \approx \log \frac{y_p}{1-y_p}$$

Exercise 3.11

$$\sigma(b + x_p^T w) \approx y_p$$

$$g(\tilde{w}) = \sum_{p=1}^{P} \left(\sigma(b + x_p^T \tilde{w}) - y_p\right)^2$$

$$\nabla g(\tilde{w}) = 2\sum_{p=1}^{P} (\sigma_p(b + x_p^T \tilde{w}) - y_p) \cdot \nabla \sigma(b + \tilde{x}_p^T w))$$

$$\nabla g(\tilde{w}) = 2 \sum_{p=1}^{P} (\sigma(b + x_p^T \tilde{w}) - y_p)\, \sigma(\cancel{b+} x_p^T \tilde{w})\, (1 - \sigma(\cancel{b+} x_p^T \tilde{w}))\tilde{x}_p$$

$\because \sigma'(t) = \sigma(t)(1 - \sigma(t))$ $\because \sigma'(b) = 0$ b is constant.

Excercise 3.13

$$g(\tilde{w}) = \sum_{p=1}^{P} (\sigma(b + \tilde{x}_p^T \tilde{w}) - y_p)^2 + \lambda \|w\|_2^2$$

$$\nabla g(\tilde{w}) = 2\sum_{p=1}^{P} (\sigma(b + \tilde{x}_p^T \tilde{w} - y_p)) \nabla \sigma((b + \tilde{x}_p^T \tilde{w}) - y_p) + \nabla \lambda \|w\|_2^2$$

$$\nabla g(\tilde{w}) =$$

$$= 2\sum_{p=1}^{P} (\sigma(\tilde{x}_p^T \tilde{w}) - y_p) \sigma(\tilde{x}_p^T \tilde{w})(1 - \sigma(\tilde{x}_p^T \tilde{w})) \tilde{x}_p + 2\lambda \tilde{w}$$

$$\tilde{w} \to \begin{bmatrix} 0 \\ w \end{bmatrix}$$

$\because \nabla \|w\|^2 = 2\tilde{w}$

$\because \sigma'(t) = \sigma(t)(1 - \sigma(t))$