CSC321 Homework 4

Q1. Solution:

(a) $\theta_i^{(t+1)} = \theta_i^{(t)} - \alpha \cdot \frac{\partial C(\theta^{(t)})}{\partial \theta_i^{(t)}}$

$\theta_i^{(t+1)} = \theta_i^{(t)} - \alpha \cdot a_i(\theta_i^{(t)} - r_i)$

$= (1-\alpha a_i)\theta_i^{(t)} + \alpha a_i r_i$

(b) Since $\theta_i^{(t+1)} = (1-\alpha a_i)\theta_i^{(t)} + \alpha a_i r_i$

then $\theta_i^{(t+1)} - r_i = (1-\alpha a_i)\theta_i^{(t)} + (\alpha a_i - 1) r_i$

then $\theta_i^{(t+1)} - r_i = (1-\alpha a_i)(\theta_i^{(t)} - r_i)$

hence $e_i^{(t+1)} = (1-\alpha a_i) e_i^{(t)}$

(c) Since $e_i^{(t+1)} = (1-\alpha a_i) e_i^{(t)}$ $\forall t$

hence $e_i^{(t)} = (1-\alpha a_i)^t e_i^{(0)}$

In this case,

when $0 < 1-\alpha a_i < 1$ i.e. $0 < \alpha < \frac{1}{a_i}$, the procedure is stable

when $\alpha > \frac{1}{a_i}$, the procedure is unstable

This indicates when α is large, α is easier to be bigger than $\frac{1}{a_i}$, which leads to the unstability. Also, high curvature dimensions i.e. a_i is large in this case, make $\frac{1}{a_i}$ very small. Hence even though we begin with a relatively small α, α is still easy to be larger than $\frac{1}{a_i}$ at those high curvature dimensions, which results in the unstability also.

(d) Since $e_i^{(t)} = \theta_i^{(t)} - r_i = (1-\alpha a_i)^t e_i^{(0)}$

then $C(\theta^{(t)}) = \frac{1}{2}\sum_{i=1}^{N}(\theta_i^{(t)} - r_i)^2$

$= \frac{1}{2}\sum_{i=1}^{N}\left[(1-\alpha a_i)^t e_i^{(0)}\right]^2$

$= \frac{1}{2}\sum_{i=1}^{N}(1-\alpha a_i)^{2t}(\theta_i^{(0)} - r_i)^2$

Consider the jth term of the summation $\frac{1}{2}(1-\alpha a_j)^{2t}(\theta_j^{(0)}-r_j)^2$

then $\lim_{t\to\infty} \dfrac{\frac{1}{2}(1-\alpha a_j)^{2t}(\theta_j^{(0)}-r_j)^2}{\frac{1}{2}\sum_{i=1}^{N}(1-\alpha a_i)^{2t}(\theta_i^{(0)}-r_i)^2}$

$$= \lim_{t\to\infty} \frac{(\theta_j^{(0)}-r_j)^2}{\sum_{i=1}^{N}\left(\frac{1-\alpha a_i}{1-\alpha a_j}\right)^{2t}(\theta_i^{(0)}-r_i)^2}$$

$$= \begin{cases} (\theta_j^{(0)}-r_j)^2 & \text{if } (1-\alpha a_j)^2 = \max\{(1-\alpha a_i)^2 \mid \forall i\} \\ 0 & \text{if } (1-\alpha a_j)^2 \neq \max\{(1-\alpha a_i)^2 \mid \forall i\} \end{cases}$$

hence $\frac{1}{2}(1-\alpha a_j)^{2t}(\theta_j^{(0)}-r_j)^2$ comes to dominate whose $(1-\alpha a_j)^{2t}$ is the largest over all $(1-\alpha a_i)^{2t}$

2. Solution:

(a) $E[y] = E[\sum_j m_j w_j x_j]$
$= \sum_j E[m_j w_j x_j]$
$= \sum_j x_j w_j E[m_j]$
$= \sum_j x_j w_j \cdot \frac{1}{2}$
$= \frac{1}{2} W^T X$

$Var[y] = Var[\sum_j m_j w_j x_j]$
$= \sum_j Var[m_j w_j x_j]$
$= \sum_j w_j^2 x_j^2 Var[m_j]$

Since each m_j is a Bernoulli random variable

then $Var[m_j] = \frac{1}{2}((1-\frac{1}{2})^2 + (0-\frac{1}{2})^2) = \frac{1}{4}$

hence $Var[y] = \frac{1}{4}\sum_j w_j^2 x_j^2$
$= \frac{1}{4}\sum_j (w_j x_j)^2$
$= \frac{1}{4}(W^T W)^2$

(b) Since $E[y] = \sum_j w_j x_j \cdot \frac{1}{2}$

$$= \sum_j \left(\tfrac{1}{2} w_j\right) x_j$$

hence $\tilde{w}_j = \frac{1}{2} w_j$ for each j i.e. $\tilde{w} = \frac{1}{2} w$

(c) $\mathcal{E} = \frac{1}{2N} \sum_{i=1}^{N} E[(y^{(i)} - t^{(i)})]^2$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(E[(y^{(i)})^2 - 2y^{(i)} t^{(i)} + (t^{(i)})^2\right]$$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(E[(y^{(i)})^2] - 2\tilde{y}^{(i)} t^{(i)} + (t^{(i)})^2\right]$$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(E[(y^{(i)})]^2 + \mathrm{Var}(y^{(i)}) - 2\tilde{y}^{(i)} t^{(i)} + (t^{(i)})^2\right]$$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left[\left((\tilde{y}^{(i)})^2 - 2\tilde{y}^{(i)} t^{(i)} + (t^{(i)})^2\right) + \mathrm{Var}(y^{(i)})\right]$$

$$= \frac{1}{2N} \sum_{i=1}^{N} (\tilde{y}^{(i)} - t^{(i)})^2 + \frac{1}{2N} \sum_{i=1}^{N} \mathrm{Var}(y^{(i)})$$

$$= \frac{1}{2N} \sum_{i=1}^{N} (\tilde{y}^{(i)} - t^{(i)})^2 + \frac{1}{2N} \sum_{i=1}^{N} \frac{1}{4} (w^T x)^2$$

$$= \frac{1}{2N} \sum_{i=1}^{N} (\tilde{y}^{(i)} - t^{(i)})^2 + \frac{1}{2N} \sum_{i=1}^{N} (\tilde{w}^T x)^2$$

$$= \frac{1}{2N} \sum_{i=1}^{N} (\tilde{y}^{(i)} - t^{(i)})^2 + \frac{1}{2} (\tilde{w}^T x)^2$$

Hence $R(\tilde{w}_1, \cdots, \tilde{w}_D) = \frac{1}{2}(\tilde{w}^T x)^2$

$$= \frac{1}{2} \sum_{i=1}^{D} (\tilde{w}_i x_i)^2$$