hw 4

1. (a) No. W^T is a $d \times D$ matrix where each row is a eigenvector of Σ. since for each w_i in W^T, $\|w_i\|=1$ and $w_i w_j = 0$ $(i \neq j)$ as stated in the PCA procedures. $W^T W = I$ in this way. However, W is only $D \times d$ where $d<D$. It is not a square matrix and thus is not invertible. And $W^T \neq W^{-1}$ since W^{-1} doesn't exist for rectangular matrices. Then, $WW^T \neq I$, so $V^t = Wz^t = WW^T x^t \neq x^t$

(b). No.

In PCA, $\|w_i\|=1$ for $i=1,\dots,d$, $\sum_{t=1}^{N} \|x^t\|_2^2 = N$

$$\sum_{t=1}^{N} \|V^t\|_2^2 = \sum_{t=1}^{N} \sum_{i=1}^{D} (V_i^t)^2 = \sum_{t=1}^{N} \sum_{i=1}^{D} (WW^T x_i^t)^2$$

$$\sum_{t=1}^{N} \|x^t - V^t\|_2^2 = \sum_{t=1}^{N} \|(1-WW^T)x^t\|_2^2 = \sum_{t=1}^{N} \sum_{i=1}^{D} [(1-WW^T)x_i^t]^2$$

$$= \sum_{t=1}^{N} \sum_{i=1}^{D} (x_i^t - WW^T x_i^t)^2 = N - 2\sum_{t=1}^{N} \sum_{i=1}^{D} WW^T (x_i^t)^2 + \sum_{t=1}^{N} \sum_{i=1}^{D} (WW^T x_i^t)^2$$

To prove $\sum_{t=1}^{N} \|x^t\|_2^2 - \sum_{t=1}^{N} \|V^t\|_2^2 = \sum_{t=1}^{N} \|x^t - V^t\|_2^2$

it should be $N - \sum_{t=1}^{N} \sum_{i=1}^{D} (WW^T x_i^t)^2 = N - 2\sum_{t=1}^{N} \sum_{i=1}^{D} WW^T (x_i^t)^2 + \sum_{t=1}^{N} \sum_{i=1}^{D} (WW^T x_i^t)^2$

which equals $\sum_{t=1}^{N} \sum_{i=1}^{D} WW^T (x_i^t)^2 = \sum_{t=1}^{N} \sum_{i=1}^{D} (WW^T x_i^t)^2$

which equals $\sum_{t=1}^{N} \sum_{i=1}^{D} (WW^T)(WW^T - I)(x_i^t)^2 = 0$

Because eigenvector cannot be zero and W is not invertible. $WW^T \neq 0$ and $WW^T \neq I$ so. $\sum_{t=1}^{N} \sum_{i=1}^{D} (WW^T)(WW^T - I)(x_i^t)^2 \neq 0$

Thus, $\sum_{t=1}^{N} \|x^t\|_2^2 - \sum_{t=1}^{N} \|V^t\|_2^2 = \sum_{t=1}^{N} \|x^t - V^t\|_2^2$ is not correct.

Since we deal with SGD in this case, we can only consider one data point at a time

2. (a)

$$\Delta v_{i,h} = -\eta \frac{\partial E}{\partial v} = -\eta \frac{\partial E}{\partial \eta}\cdot\frac{\partial \eta}{\partial v} = -\eta \frac{\partial L(r_i^t, \eta_i^t)}{\partial \eta_i^t}\cdot g'(a_i^t)\cdot z_h^t$$

$$= \eta\cdot\left(-g'(a_i^t)\frac{\partial L(r_i^t,\eta_i^t)}{\partial\eta_i^t}\right) z_h^t = \eta\Delta_i^t\cdot z_h^t$$

(b)

$$\Delta w_{h,j} = -\eta\sum_{i=1}^{K} \frac{\partial E}{\partial w} = -\eta\sum_{i=1}^{K} \frac{\partial E}{\partial\eta}\cdot\frac{\partial\eta}{\partial z}\cdot\frac{\partial z}{\partial w} = -\eta\sum_{i=1}^{K}\left(\frac{\partial L(r_i^t,\eta_i^t)}{\partial\eta_i^t}\right)\cdot\frac{\partial\eta}{\partial z}\cdot\frac{\partial z}{\partial w}$$

$$= -\eta\sum_{i=1}^{K}\left(\frac{\partial L(r_i^t,\eta_i^t)}{\partial\eta_i^t}\right)\left(g'(a_i^t)\cdot v_{i,h}\right)\cdot\left(g'(a_h^t)x_j^t\right)$$

$$= \eta\cdot g'(a_h^t)\left[\sum_{i=1}^{K}\left(-g'(a_i^t)\cdot\frac{\partial L(r_i^t,\eta_i^t)}{\partial\eta_i^t}\right)\cdot v_{i,h}\right]\cdot x_j^t \quad (1)$$

Q3

Summary

MyFLDA2 with Boston50

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.168316831683	0.0891089108911	0.108910891089	0.277227722772	0.0882352941176	0.146359930111	0.0716478439556

MyFLDA2 with Boston75

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.207920792079	0.0891089108911	0.247524752475	0.158415841584	0.0588235294118	0.152358765288	0.0706112973162

LogisticRegression with Boston50

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.128712871287	0.108910891089	0.0891089108911	0.267326732673	0.117647058824	0.142341292953	0.0638250381703

LogisticRegression with Boston75

K=0	K=1	K=2	K=3	K=4	Mean	Std
0.0891089108911	0.128712871287	0.138613861386	0.108910891089	0.0490196078431	0.102873228499	0.0318471419286

Yes.

extra

$$\max(0,a)=\begin{cases} a, & \text{if } a>0 \\ 0, & \text{if } a\le 0\end{cases} \qquad \frac{1}{2}\min(0,a)=\begin{cases} 0, & \text{if } a>0 \\ \frac{1}{2}a, & \text{if } a\le 0\end{cases}$$

$$f_{tlu}(a)=\max(0,a)+\frac{1}{2}\min(0,a) = \begin{cases} a, & \text{if } a>0 \\ \frac{1}{2}a, & \text{if } a\le 0\end{cases}$$

$$\frac{\partial f_{tlu}(a)}{\partial a} = \begin{cases} 1, & \text{if } a>0 \\ \frac{1}{2}, & \text{if } a\le 0\end{cases}$$

$$L_{sq}^{(tlu)}(w)=\sum_{\delta=1}^{N}\left(y^{\delta}-f_{tlu}(w^T x^{\delta})\right)^2$$

$$\frac{\partial L}{\partial w} = \begin{cases} -2\left(y^{\delta}-f_{tlu}(w^T x^{\delta})\right)\cdot x^{\delta}, & \text{if } w^T x^{\delta}>0 \\ -\left(y^{\delta}-f_{tlu}(w^T x^{\delta})\right)\cdot x^{\delta}, & \text{if } w^T x^{\delta}\le 0\end{cases}$$

$$\frac{\partial^2 L}{\partial w^2} = \begin{cases} 2x^{\delta}\cdot x^{\delta}, & \text{if } w^T x^{\delta}>0 \\ x^{\delta}\cdot \frac{1}{2}x^{\delta}, & \text{if } w^T x^{\delta}\le 0\end{cases} = \begin{cases} 2(x^{\delta})^2, & \text{if } w^T x^{\delta}>0 \\ \frac{1}{2}(x^{\delta})^2, & \text{if } w^T x^{\delta}\le 0\end{cases}$$

for $w^T x^{\delta} \in R$, $\frac{\partial^2 L}{\partial w^2}\ge 0$, Thus $L_{sq}^{(tlu)}(w)$ is a convex function of w