SUBHODIP SAHA
5485317
HW4

1. $z^t = W x^t$
$v^t = W^T z^t$
So, $v^t = W^T W x^t$

a. For. $v^t = W^T W x^t$
When we are making the low-dimensional projection to d dim, then $W^T W \neq 1$
(In fact the opposite is true $W W^T = 1$).
As, $W^T W \neq 1$
then $\boxed{v^t \neq x^t}$
So, the claim is wrong

We can also prove that in another way,
when we are projecting to lower dimensional space we are removing some variance.

$$Var(v^t) = \cancel{\lambda_1 + \lambda_2 + \dots \lambda_d}$$

$\lambda_1, \lambda_2 \dots$ are eigenvalue after eigenvalue decomposition

whereas,
$$Var(x^t) = \lambda_1 + \lambda_2 + \dots \lambda_d + \dots \lambda_D.$$
$$Var(x^t) > Var(v^t)$$
The $Var(x^t)$ unequal then, they must be unequal if the variance are ~~as~~ so, $\boxed{x^t \neq v^t}$

Scanned by CamScanner

b So, $z^t = w x^t$

$v^t = w^T z^t$

$v^t = w^T w x^t$.

Now, $\|x^t - v^t\|^2 = (x^t - v^t)^T (x^t - v^t)$

$= (x^{tT} - v^{tT})(x^t - v^t)$

$= x^{tT} x^t - x^{tT} v^t - v^{tT} x^t + v^{tT} v^t$

$x^{tT} v^t = x^{tT} w^T w x^t$.

$v^{tT} x^t = (w^T w x^t)^T x^t$ ~~$\neq$~~ So, $v^{tT} x^t = x^{tT} x^t$

$= x^{tT} (w^T w)^T x^t$

$= x^{tT} w^T w x^t$

Now, $v^{tT} v^t = (w^T w x^t)^T (w^T w x^t)$

$= x^t (w^T w)^T (w^T w x^t)$

$= x^t w^T \underbrace{w w^T}_{1} w x^t$

$= x^t w^T w x^t$

$= v^{tT} x^t = x^{tT} x^t$

~~After~~ we know that after projection. $w w^T = 1$

$(w^T w \neq 1)$

We find,

$v^{t^T} v^t = v^{t^T} x^t = x^{t^T} v^t$.

$\|x^t - v^t\|^2 = x^{t^T} x^t - 2 x^{t^T} v^t + v^{t^T} v^t$

$= x^{t^T} x^t - v^{t^T} v^t$

$= \|x^t\|^2 - \|v^t\|^2$

$\therefore \sum_t \|x^t - v^t\|^2 = \sum_t \|x^t\|^2 - \sum_t \|v^t\|^2$

So, $\boxed{\sum_t \|x^t\|^2 - \sum_t \|v^t\|^2 = \sum_t \|x^t - v^t\|^2}$

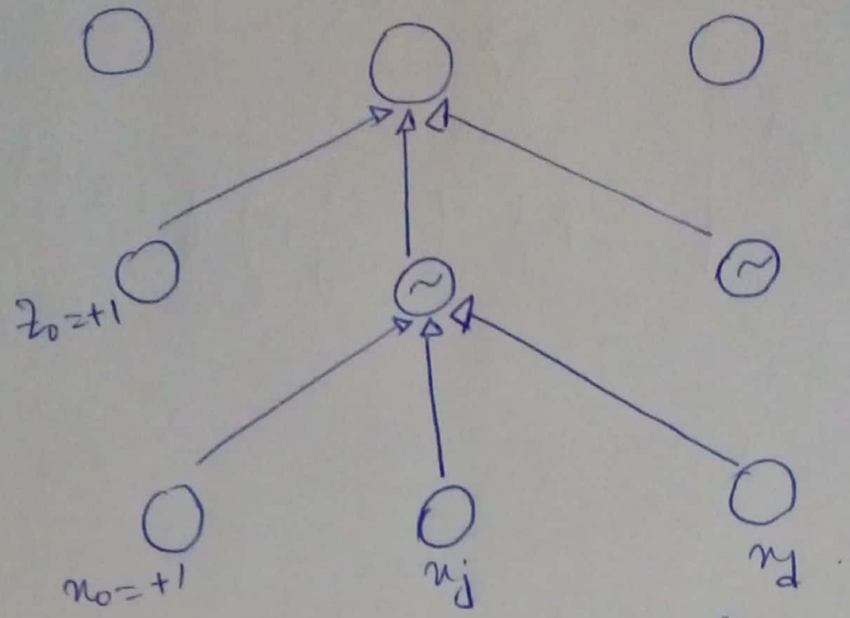
So the claim is correct.

1'

2

Problem 2

2°



$$z_h^t = g(a_h^t) = g\left(\sum_{j=1}^{d} w_{h,j} x_j^t + w_0\right)$$

$$y_i^t = g(a_i^t) = g\left(\sum_{h=1}^{H} v_{i,h} z_h^t + v_{i0}\right)$$

$$E(w, v \mid Z) = \sum_t \sum_i L(r_i^t, y_i^t)$$

$$\Delta v_{i,h} = -\eta \frac{\partial E}{\partial v_{i,h}}$$

$$\frac{\partial E}{\partial v_{i,h}} = \cancel{\sum_t} \frac{\partial L}{\partial y_i^t} \frac{\partial y_i^t}{\partial v_{i,h}} = \frac{\partial L}{\partial y_i^t} \frac{\partial y_i^t}{\partial v_{i,h}}$$

(As we are doing stochastic Gradient we choose only one t point for calculating gradient)

$$= \frac{\partial L}{\partial y_i^t} g'\left(\sum_{h'=1}^{H} v_{i,h'} z_{h'}^t + v_{i0}\right) z_{h'}^t \delta_{h'h}$$

$$\Delta v_{i,h} = -\eta \frac{\partial E}{\partial v_{i,h}} = \eta \underbrace{\left(-\frac{\partial L}{\partial y_i^t}\right) g'\left(\sum_{h=1}^{H} v_{i,h} z_h^t + v_{i0}\right)}_{\Delta_i^t} z_h^t$$

$$\Delta v_{i,h} = \eta \underbrace{\left(-\frac{\partial L}{\partial y_i^t}\right) g'\left(\sum_{h=1}^{H} v_{i,h} z_h^t + v_{i0}\right)}_{\Delta_i^t} z_h^t = \eta \Delta_i^t z_h^t$$

So

$$\boxed{\Delta v_{i,h} = \eta \Delta_i^t z_h^t}$$

where,

$$\Delta_i^t = \left(-\frac{\partial L}{\partial y_i^t}\right) g'(a_i^t)$$

b

$$\Delta w_{hj} = -\eta \frac{\partial E}{\partial w_{hj}}$$

$$\frac{\partial E}{\partial w_{hj}} = \left(\sum_{i=1}^{K} \frac{\partial L}{\partial y_i^t} \frac{\partial y_i^t}{\partial z_h^t}\right) \frac{\partial z_h^t}{\partial w_{hj}}$$

→ As we are doing SGD, we are considering only one t point

$$\frac{\partial y_i^t}{\partial z_h^t} = g'\left(\sum_{h'=1}^{H} v_{i,h'} z_{h'}^t + v_{i0}\right) v_{i,h'} \delta_{hh'}$$

$$= g'\left(\sum_{h=1}^{H} v_{i,h} z_h^t + v_{i0}\right) v_{i,h}$$

$$\frac{\partial z_h^t}{\partial w_{hj}} = g'\left(\sum_{j'=1}^{d} w_{hj'} x_{j'}^t + w_0\right) x_{j'}^t \delta_{j'j}$$

$$= g'\left(\sum_{j=1}^{d} w_{hj} x_j^t + w_0\right) x_j^t$$

$$\frac{\partial E}{\partial w_{hj}} = + \left[\left(\sum_{i=1}^{K} \frac{\partial E}{\partial y_i^t} \frac{\partial y_i^t}{\partial z_h^t} \right) \frac{\partial z_h^t}{\partial w_{hj}} \right.$$

2'

$$= - \left[\sum_{i=1}^{K} \underbrace{\left(- \frac{\partial L}{\partial y_i^t} \right) g'\left(\sum_{h=1}^{H} v_{i,h} z_h^t + v_{i0} \right)}_{\Delta_i^t} v_{ih} \right] g'\left(\sum_{j=1}^{d} w_{hj} x_j^t + w_0 \right) x_j^t$$

$$= - \underbrace{\sum_{i=1}^{K} \Delta_i^t v_{i,h} \, g'\left(\sum_{j=1}^{d} w_{hj} x_j^t + w_0 \right)}_{\Delta_h^t} x_j^t = - \Delta_h^t x_j^t$$

$$\Delta w_{hj} = - \eta \frac{\partial E}{\partial w_{hj}} = - \eta \left(- \Delta_h^t \right) x_j^t = \eta \Delta_h^t x_j^t$$

So, $\boxed{\Delta w_{hj} = \eta \Delta_h^t x_j^t}$

Problem EC1

EC1

EC1°

$L(p_i^t, y_i^t) = (p_i^t - y_i^t)^2$

a $g(u) = \max(0, u)$

~~are~~

$\cancel{\frac{\partial g}{\partial u}}$

$$g(u) = 0 \quad u < 0$$

$$= u \quad u > 0.$$

$$\text{So, } \frac{\partial g}{\partial u} = 0 \quad u < 0$$

$$= 1 \quad u > 0.$$

Now ~~cond~~ consider, $g\left(\sum_{a=1}^{H} v_{i,a} z_a^t + v_{i0}\right)$

So. $g\left(\sum_{a=1}^{H} v_{i,a} z_a^t + v_{i0}\right) = 0$ if $\sum_{a=1}^{H} v_{i,a} z_a^t + v_{i0} < 0$

$= \sum_{a=1}^{H} v_{i,a} z_a^t + v_{i0}$ otherwise

So. $\frac{\partial g}{\partial v_{ia}} = 0$ if $\sum_{a=1}^{H} v_{ia} z_a^t + v_{i0} < 0$

$= z_a^t$ otherwise.

$$-\frac{\partial L}{\partial y_i^t} = 2(p_i^t - y_i^t)$$

So, the update becomes,

if $\sum_{h} v_{ih} z_h^t + v_{i0} < 0$.

$\Delta v_{ih} = 0$.

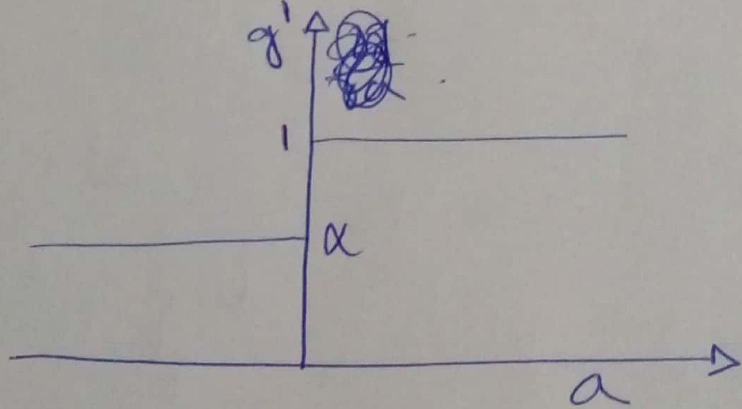
if $\sum_{h} v_{ih} z_h^t + v_{i0} > 0$

$\Delta v_{ih} = \eta\, 2 (p_i^t - y_i^t) z_h^t$.

b $g(a) = \max(0, a) + \alpha \min(0, a)$

If $a < 0$. $g(a) = 0 + \alpha a$, $\frac{\partial g}{\partial a} = \alpha = g'$

If $a > 0$ $g(a) = a$, $\frac{\partial g}{\partial a} = 1 = g'$



So, gradient $g'(a) = 1$ if $a > 0$

$= \alpha$ if $a < 0$.

EC 1'

C Let's take,

$y = g(vz + v_0)$ $\quad z = g(wx + w_0)$

So, $\boxed{y = g(v\, g(wx + w_0) + v_0)}$

(→ If we can find a value of α for which it becomes a linear fun' then it should be valid for any other case.

Now we try to find an α for which y becomes linear function of x.

* if $wx + w_0 < 0$

$z = \alpha(wx + w_0)$

* if $v\alpha(wx+w_0)+v_0<0$

$y = \alpha[v\alpha(wx+w_0) + v_0]$

if $\alpha = 1$

$y = v(wx + w_0) + v_0$

* if $v\alpha(wx+w_0)+v_0>0$

$y = v\alpha(wx+w_0)+v_0$

if $\alpha = 1$

$y = v(wx+w_0) + v_0$

* if $wx + w_0 > 0$

$z = wx + w_0$

* if $v(wx+w_0)+v_0<0$

$y = \alpha[v(wx+w_0) + v_0]$

if $\alpha = 1$

$y = v(wx+w_0) + v_0$

* if $v(wx+w_0)+v_0>0$

$y = v(wx+w_0)+v_0$

if $\alpha = 1$

$y = v(wx+w_0) + v_0$

All the y functional form becomes same if $\underline{\alpha = 1}$

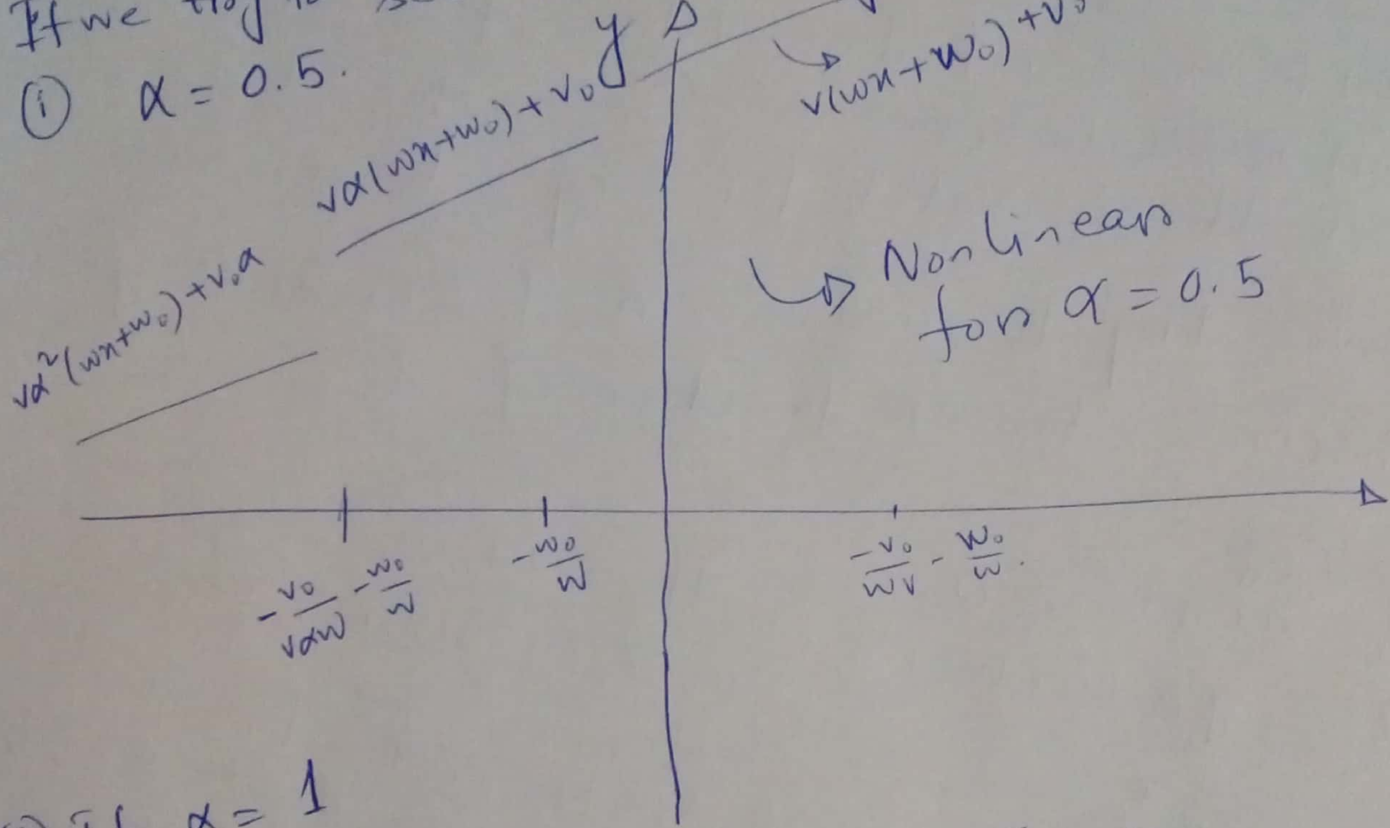
then $y(x)$ becomes linear on x.

So if $\alpha = 1$, then in this case.

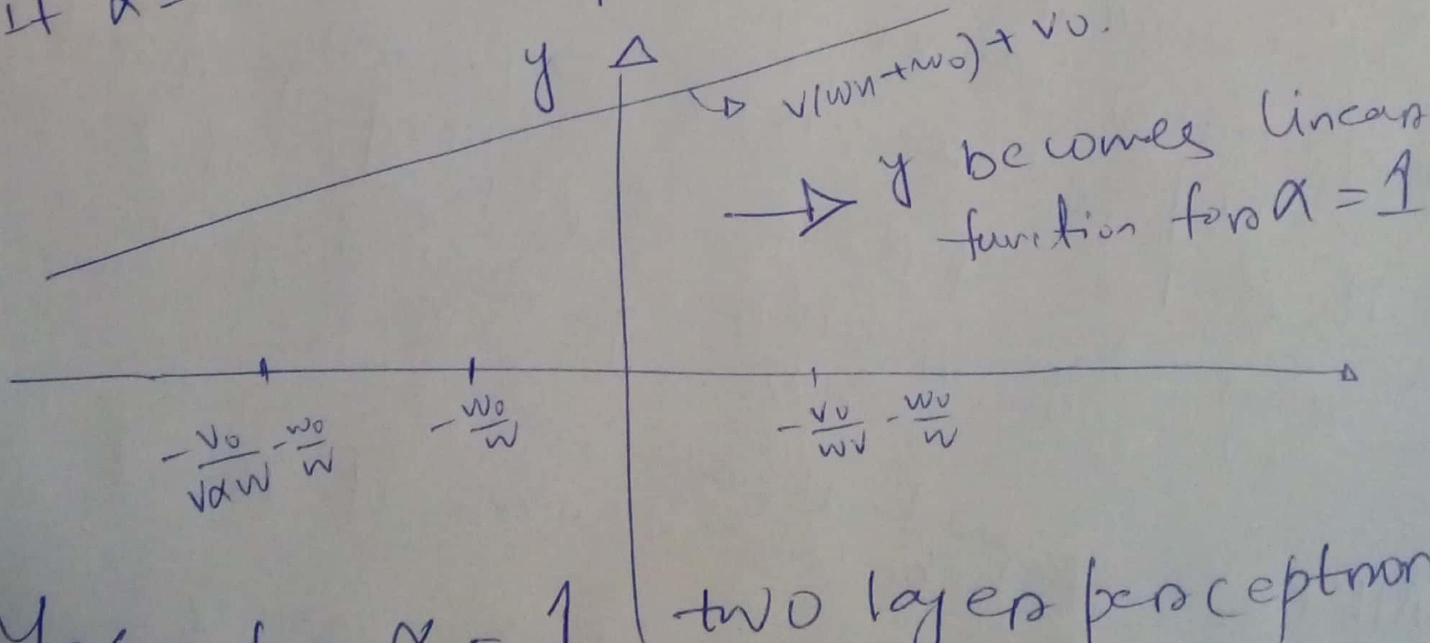
$y = v(wx + w_0) + v_0 \rightarrow$ linear fun' of x.
(for any value of x)

If we try to see it visually for two diff α.

(i) $\alpha = 0.5$.



(ii) If $\alpha = 1$



Yes, for $\alpha = 1$ two layer perceptron becomes a linear model.

Problem 3

1. Error rates for MySVM2 with m=40 for Boston 50

F1	F2	F3	F4	F5	Mean	SD
0.28	0.19	0.17	0.17	0.24	0.217	0.042

2. Error rates for MySVM2 with m=200 for Boston 50

F1	F2	F3	F4	F5	Mean	SD
0.21	0.25	0.21	0.16	0.28	0.229	0.040

3. Error rates for MySVM2 with m=n for Boston 50

F1	F2	F3	F4	F5	Mean	SD
0.21	0.25	0.20	0.19	0.25	0.227	0.025

4. Error rates for LogisticRegression for Boston 50

F1	F2	F3	F4	F5	Mean	SD
0.22	0.23	0.20	0.18	0.12	0.198	0.038

5. Error rates for MySVM2 with m=40 for Boston 75

F1	F2	F3	F4	F5	Mean	SD
0.21	0.33	0.28	0.29	0.16	0.261	0.060

6. Error rates for MySVM2 with m=200 for Boston 75

F1	F2	F3	F4	F5	Mean	SD
0.27	0.26	0.31	0.23	0.20	0.261	0.036

7. Error rates for MySVM2 with m=n for Boston 75

F1	F2	F3	F4	F5	Mean	SD
0.24	0.23	0.28	0.21	0.31	0.261	0.035

8. Error rates for LogisticRegression for Boston 75

F1	F2	F3	F4	F5	Mean	SD
0.16	0.06	0.07	0.07	0.11	0.102	0.036