subject

Year Month Date ( )

آرین تشکر - تکلیف چهارم درس شبکه‌های عصبی

۴۰۰۲۳۴۹۴

E 10.3.

Performance Index $(F(x)) = c - 2x^T h + x^T R x$; $C = E[t^2]$, $h = E[tz]$, $R = E[zz^T]$

$C: \; 0.25(-1)^2 + 0.75(1)^2 = 1$

$h: \; 0.25(-1)\begin{bmatrix}1\\-1\end{bmatrix} + 0.75(1)\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}0.5\\1\end{bmatrix}$

$R: \; 0.25\,p_2 p_2^T + 0.75\,p_1 p_1^T = 0.25\begin{bmatrix}1\\-1\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix} + 0.75\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} = \begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}$

$\rightarrow F(x) = 1 - 2x^T\begin{bmatrix}0.5\\1\end{bmatrix} + x^T\begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}x$

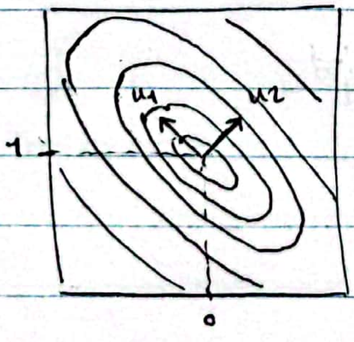
$\nabla^2 F(x) = 2R = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}$, $\lambda_1 = 1, u_1 = \begin{bmatrix}-0.7071\\0.7071\end{bmatrix}$

$\lambda_2 = 3, u_2 = \begin{bmatrix}0.7071\\0.7071\end{bmatrix}$

$\rightarrow$ both eigenvalues positive $\rightarrow$ global minimum

$x^* = R^{-1}h = \begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}^{-1}\begin{bmatrix}0.5\\1\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}$

Sketch of the contour:



Maximum stable Learning Rate: $\alpha < 2/\lambda_{max}(2R) = 2/3$

E10.4. i. $F(x) = c - 2x^T h + x^T R x$

$c = E[t^2] = 0.5(1)^2 + 0.5(-1)^2 = 1$

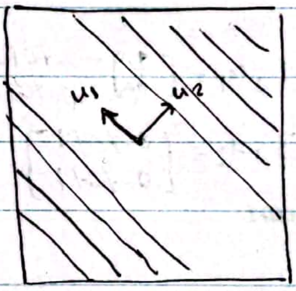
$h = E[tz] = 0.5(-1)\begin{bmatrix}-1\\-1\end{bmatrix} + 0.5(1)\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$

$R = E[zz^T] = 0.5\begin{bmatrix}1\\1\end{bmatrix}[1\ 1] + 0.5\begin{bmatrix}-1\\-1\end{bmatrix}[-1\ -1] = \begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$

$\rightarrow F(x) = 1 - 2x^T\begin{bmatrix}1\\1\end{bmatrix} + x^T\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}x$

$\nabla^2 F(x) = 2R = \begin{bmatrix}2 & 2\\2 & 2\end{bmatrix}$, $\begin{cases}\lambda_1 = 0, \ u_1 = \begin{bmatrix}-0.7071\\0.7071\end{bmatrix} \\ \lambda_2 = 4, \ u_2 = \begin{bmatrix}0.7071\\0.7071\end{bmatrix}\end{cases}$ $\rightarrow$ stationary valley

contour sketch.



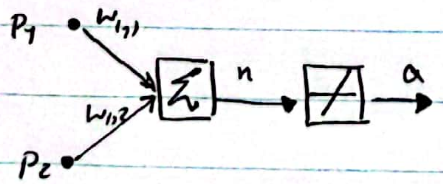
ii.

max stable learning rate: $\alpha < \dfrac{2}{\lambda_{max}(2R)} = \dfrac{2}{4} = 0.5$

iii - v $\rightarrow$ Refer to the Matlab script.

E10.6

i.



ii. ${}_1w(k+1) = {}_1w(k) + 2\alpha e(k) p(k)$, $e(k) = t(k) - {}_1w^T(k) p(k)$

(1). $t(0) = 1$, $a(0) = [0, 0]\begin{bmatrix}1\\1\end{bmatrix} = 0 \rightarrow e(0) = 1$

$\rightarrow {}_1w(1) = \begin{bmatrix}0\\0\end{bmatrix} + 2 \times 0.1 \times 1 \times \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}0.2\\0.2\end{bmatrix}$

②. $t(1) = 1,\ a(1) = [0.2\ \ 0.2]\begin{bmatrix}-1\\2\end{bmatrix} = 0.2 \rightarrow e(1) = 0.8$

$\rightarrow w(2) = \begin{bmatrix}0.2\\0.2\end{bmatrix} + 2\times0.1\times0.8\times\begin{bmatrix}-1\\2\end{bmatrix} = \begin{bmatrix}0.04\\0.52\end{bmatrix}$

③. $t(2) = -1,\ a(2) = [0.04\ \ 0.52]\begin{bmatrix}0\\-1\end{bmatrix} = -0.52 \rightarrow e(2) = -0.48$

$\rightarrow w(3) = \begin{bmatrix}0.04\\0.52\end{bmatrix} + 2\times0.1\times(-0.48)\times\begin{bmatrix}0\\-1\end{bmatrix} = \begin{bmatrix}0.04\\0.616\end{bmatrix}$

④. $t(3) = -1,\ a(3) = [0.04\ \ 0.616]\begin{bmatrix}-4\\1\end{bmatrix} = 0.456 \rightarrow e = -1.456$

$\rightarrow w(4) = \begin{bmatrix}0.04\\0.616\end{bmatrix} + 2\times0.1\times(-1.456)\times\begin{bmatrix}-4\\1\end{bmatrix} = \begin{bmatrix}1.2048\\0.3248\end{bmatrix}$

iii. $R = E[zz^T] = 0.25\times\begin{bmatrix}1\\1\end{bmatrix}\times[1\ \ 1] + 0.25\times\begin{bmatrix}-1\\2\end{bmatrix}\times[-1\ \ 2] + 0.25\times\begin{bmatrix}0\\-1\end{bmatrix}\times[0\ \ -1]$

$+ 0.25\times\begin{bmatrix}-4\\1\end{bmatrix}[-4\ \ 1] = \begin{bmatrix}4.5 & -1.25\\-1.25 & 1.75\end{bmatrix}$

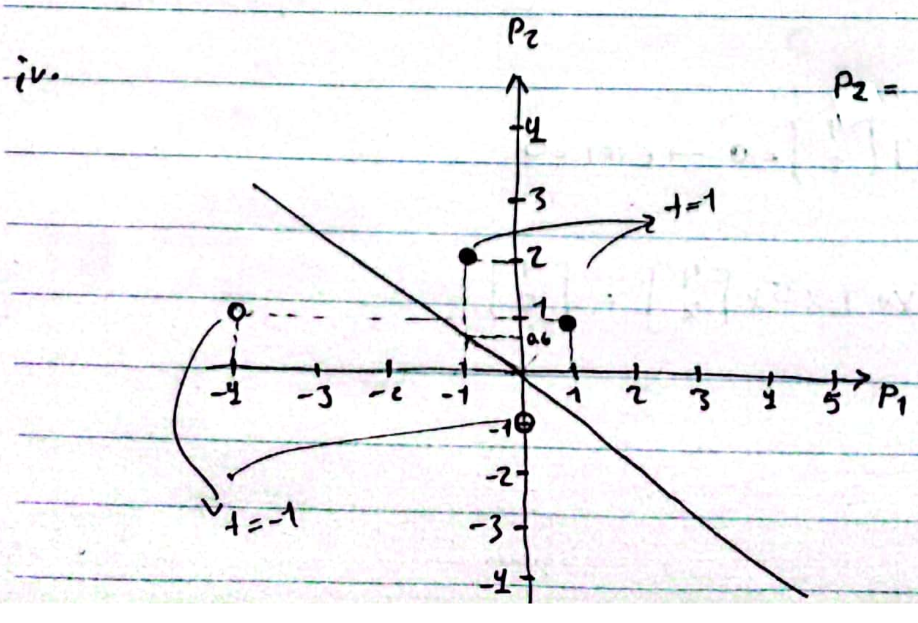
$h = E[tz] = 0.25\times1\times\begin{bmatrix}1\\1\end{bmatrix} + 0.25\times1\times\begin{bmatrix}-1\\2\end{bmatrix} + 0.25\times(-1)\times\begin{bmatrix}0\\-1\end{bmatrix}$

$+ 0.25\times(-1)\begin{bmatrix}-4\\1\end{bmatrix} = \begin{bmatrix}1\\0.75\end{bmatrix}$

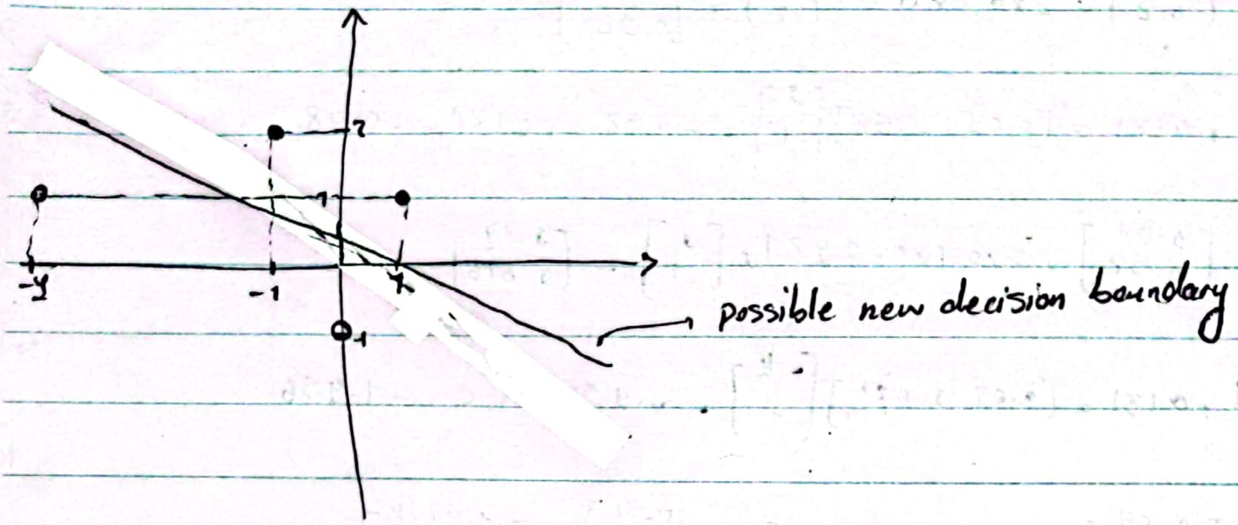
$x^* = R^{-1}h = \begin{bmatrix}0.4257\\0.7527\end{bmatrix} \rightarrow$ decision boundary: $[0.4257\ \ 0.7327]\begin{bmatrix}p_1\\p_2\end{bmatrix} = 0$

iv.

$p_2 = \dfrac{-0.4257}{0.7327}p_1 \approx -0.6\,p_1$

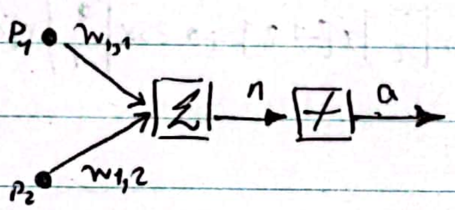


v. It would probably reposition slightly to lower the error on the further away points so as to minimize sum of squared errors.



E 10.9.

i.



ii. First we need to calculate $R = E[zz^t]$:

$$R = 0.25\begin{bmatrix}4\\2\end{bmatrix}[2\ 4] + 0.25\begin{bmatrix}2\\-4\end{bmatrix}[2\ -4] + 0.5\begin{bmatrix}-4\\4\end{bmatrix}[-4\ 4] = \begin{bmatrix}11 & -6\\ -9 & 14\end{bmatrix}$$

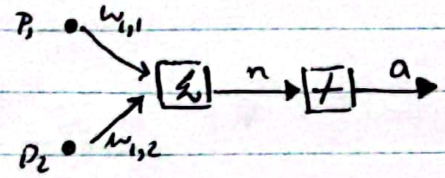
$$\therefore \alpha < \frac{1}{\lambda_{max}(R)} = \frac{1}{20} = 0.05$$

$\lambda_1 = 5$, $\lambda_2 = 20$

iii. $t(0) = 5$, $a(0) = [0\ 0]\begin{bmatrix}4\\2\end{bmatrix} = 0 \rightarrow e(0) = 5$

$$\rightarrow {}_1w(1) = \begin{bmatrix}0\\0\end{bmatrix} + 2 \times 0.1 \times 5 \times \begin{bmatrix}4\\2\end{bmatrix} = \begin{bmatrix}4\\2\end{bmatrix}$$

E10.11

i. 

ii. $t(0) = -1, \quad a(0) = \begin{bmatrix}0 & 0\end{bmatrix}\begin{bmatrix}-1\\2\end{bmatrix} = 0 \rightarrow e(0) = -1$

$\rightarrow {}_1w(1) = \begin{bmatrix}0\\0\end{bmatrix} + 2\times0.1\times(-1)\times\begin{bmatrix}-1\\2\end{bmatrix} = \begin{bmatrix}0.2\\-0.4\end{bmatrix}$

iii. we will calculate R, h and c:

$$c = E[t^2] = 0.25(-1)^2 + 0.25(-1)^2 + 0.25(1)^2 + 0.25(1)^2 = 1$$

$$h = E[tz] = 0.25(-1)\begin{bmatrix}-1\\2\end{bmatrix} + 0.25(-1)\begin{bmatrix}2\\-1\end{bmatrix} + 0.25(1)\begin{bmatrix}0\\-1\end{bmatrix} + 0.25(1)\begin{bmatrix}-1\\0\end{bmatrix}$$

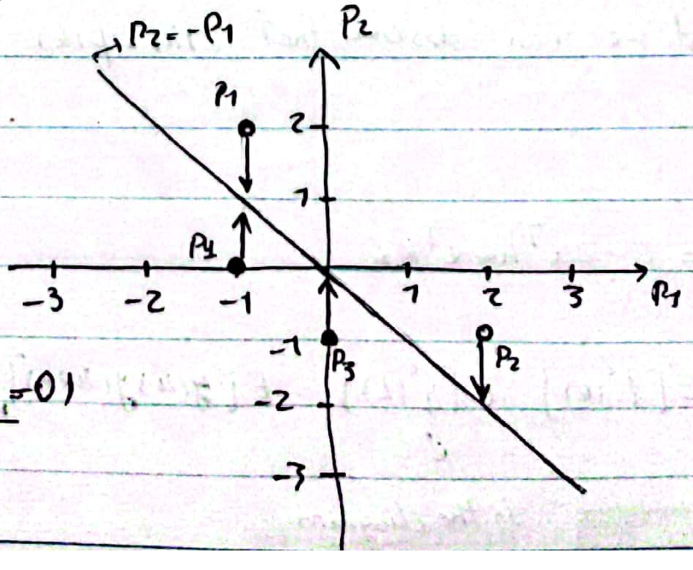
$$= \begin{bmatrix}-0.5\\-0.5\end{bmatrix}$$

$$R = E[zz^T] = 0.25\begin{bmatrix}-1\\2\end{bmatrix}\begin{bmatrix}-1 & 2\end{bmatrix} + 0.25\begin{bmatrix}2\\-1\end{bmatrix}\begin{bmatrix}2 & -1\end{bmatrix} + 0.25\begin{bmatrix}0\\-1\end{bmatrix}\begin{bmatrix}0 & -1\end{bmatrix} + 0.25\begin{bmatrix}-1\\0\end{bmatrix}\begin{bmatrix}-1 & 0\end{bmatrix}$$

$$= \begin{bmatrix}1.5 & -1\\-1 & 1.5\end{bmatrix}$$

optimal weights : $R^{-1}h = \begin{bmatrix}-1\\-1\end{bmatrix}$

iv & v.



The squared error is equal for all p_1 through p_4. So an optimal boundary with bias would be exactly the same. (optimal bias = 0) vi: $\alpha < \frac{1}{\lambda_{max}(R)} = \frac{1}{2.5} = 0.4$

$\lambda_1(R) = 0.5 \quad , \lambda_2(R) = 2.5$

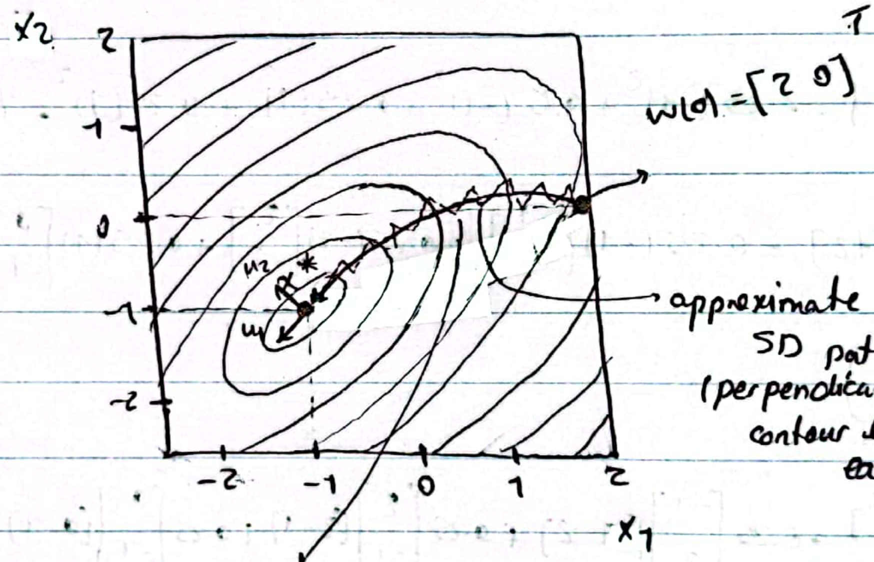
vii. $2R = \nabla^2 MSE(x) \qquad , x^* = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$\lambda_1(2R) = 1 \quad , u_1(2R) = \begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix}$

$\lambda_2(2R) = 5 \quad , u_2(2R) = \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix}$

global minimum

vii & viii.



approximate SD path (perpendicular to the contour lines at each step)

LMS path: a noisy version of SD due to approximation.

(if α too small it will look like SD)

E10.13.

i. first we can observe that $z(k) = p(k) = \begin{bmatrix} y(k-1) \\ y(k-2) \end{bmatrix}$ (first input delay once second input delayed twice)

$F(x) = c - 2x^T h + x^T R x$

$c = E[t^2(k)] = E[y^2(k)] = E[y(k)y(k+n)]_{n=0} = \underline{c_y(0)}$

in the diagram we have $t(k) = y(k)$

$$h = E[tz] = E\left[y(k)\begin{bmatrix} y(k-1) \\ y(k-2)\end{bmatrix}\right] = E\begin{bmatrix} y(k)y(k-1) \\ y(k)y(k-2)\end{bmatrix} = E\begin{bmatrix} y(k+1)\,y(k) \\ y(k+2)\,y(k)\end{bmatrix} = \begin{bmatrix} C_y(1) \\ C_y(2)\end{bmatrix}$$

$$R = E[zz^T] = E\begin{bmatrix} y^2(k-1) & y(k-1)y(k-2) \\ y(k-1)y(k-2) & y^2(k-2)\end{bmatrix} = E\begin{bmatrix} y^2(k) & y(k+1)y(k) \\ y(k+1)y(k) & y^2(k)\end{bmatrix}$$

$$= \begin{bmatrix} C_y(0) & C_y(1) \\ C_y(1) & C_y(0)\end{bmatrix}$$

ii. $y(k) = \sin\left(\frac{k\pi}{5}\right)$

$$C_y(n) = E[y(k)\,y(k+n)] = E\left[\sin\left(\frac{k\pi}{5}\right)\sin\left(\frac{(k+n)\pi}{5}\right)\right] =$$

$$\frac{1}{10}\sum_{k=1}^{10} \sin\frac{k\pi}{5}\sin\frac{(k+n)\pi}{5} = \frac{1}{10}\sum_{k=1}^{10} \frac{1}{2}\left[\cos\left(\frac{k\pi}{5} - \frac{(k+n)\pi}{5}\right) + \cos\left(\frac{k\pi}{5} + \frac{(k+n)\pi}{5}\right)\right]$$

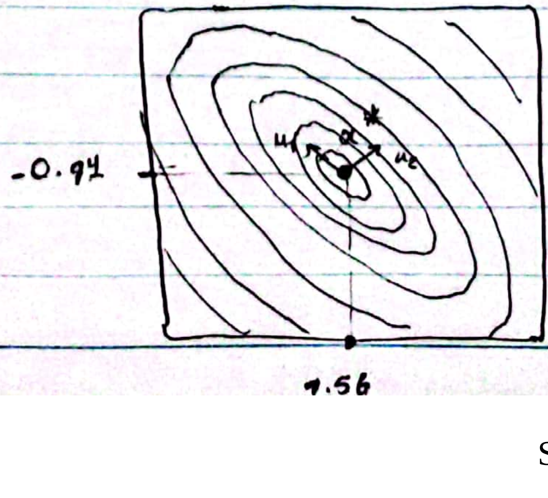
$$= \frac{1}{20}\left(10\cos\left(\frac{n\pi}{5}\right) + \cancel{\sum_{k=1}^{10}\cos\left(\frac{2k\pi+n\pi}{5}\right)}^{\,0}\right) = \frac{1}{2}\cos\frac{n\pi}{5}$$

$$\rightarrow C_y(0) = 0.5 \quad , \quad C_y(1) \approx 0.4 \quad , \quad C_y(2) \approx 0.15$$

$$\Rightarrow F(x) = c - 2x^T h + x^T R x = 0.5 - 2x^T\begin{bmatrix} 0.4 \\ 0.15\end{bmatrix} + x^T\begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.5\end{bmatrix}x$$

iii. $\nabla^2 F(x) = 2R = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1\end{bmatrix} \longrightarrow \begin{cases} \lambda_1 = 0.2 & u_1 = \begin{bmatrix} -0.7071 \\ 0.7071\end{bmatrix} \\ \lambda_2 = 1.8 & u_2 = \begin{bmatrix} 0.7071 \\ 0.7071\end{bmatrix}\end{cases}$

$$x^* = R^{-1}h \approx \begin{bmatrix} 1.56 \\ -0.94\end{bmatrix}$$



iv. $\alpha < \frac{2}{\lambda_{max}(2R)} = \frac{2}{1.8} \approx 1.11$

v. $\alpha = 1$, ${}_1w(0) = [0 \;\; 0]^T$

$p_0 = \begin{bmatrix} y(-1) \\ y(-2) \end{bmatrix} \approx \begin{bmatrix} -0.6 \\ -0.95 \end{bmatrix}$, $t(0) = y(0) = 0$

$a(0) = [0 \;\; 0]\begin{bmatrix} -0.6 \\ -0.95 \end{bmatrix} = 0 \rightarrow e(0) = 0 \Rightarrow {}_1w(1) = {}_1w(0) + 0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$p(1) = \begin{bmatrix} y(0) \\ y(-1) \end{bmatrix} \approx \begin{bmatrix} 0 \\ -0.6 \end{bmatrix}$, $t(1) = y(1) \approx +0.6$

$a(1) = [0 \;\; 0]\begin{bmatrix} 0 \\ -0.6 \end{bmatrix} = 0 \rightarrow e(1) = 0.6$

$\Rightarrow {}_1w(2) = {}_1w(1) + 2 \times 1 \times 0.6 \times \begin{bmatrix} 0 \\ -0.6 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.72 \end{bmatrix}$

$p(2) = \begin{bmatrix} y(1) \\ y(0) \end{bmatrix} \approx \begin{bmatrix} 0.6 \\ 0 \end{bmatrix}$, $t(2) = y(2) \approx 0.95$

$a(2) = [0 \;\; -0.72]\begin{bmatrix} 0.6 \\ 0 \end{bmatrix} = 0 \longrightarrow e(2) = 0.95$

$\Rightarrow {}_1w(3) = {}_1w(2) + 2 \times 1 \times 0.95 \times \begin{bmatrix} 0.6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.14 \\ -0.72 \end{bmatrix}$

vi & vii. $\longrightarrow$ Refer to the MATLAB script