3.10 (a) repeat the selection of the first point of a dataset.

$$bias = \mathcal{E}[\mu] - \mu = \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mu(k) - \mu = 0$$

Where $\mu(k)$ is the first point in dataset.
So this method is unbiased.

(b) this method is unbiased, it will generally have large variance, $\mathcal{E}[(x_i-\mu)^2] = \sigma^2$. the RMS error is independent of $n$.

If use $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ to estimate the mean,

$$\mathcal{E}[(\bar{x}-\mu)^2] = \mathcal{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}[\mathcal{E}[(x_i-\mu)^2]] = \frac{\sigma^2}{n}$$

the RMS error is $\frac{\sigma}{\sqrt{n}}$ which is much less than $\sigma$

3.47. (a) For the E step,

$$Q(\theta;\theta^0) = \mathcal{E}_{x_{32}}[\ln p(x_g, x_b;\theta)|\theta^0; D_g]$$

$$= \int_{-\infty}^{\infty} (\ln p(x_1|\theta) + \ln p(x_2|\theta) + \ln p(x_3|\theta))\, p(x_{32}|\theta^0, x_{31}=2)\, dx_{32}$$

$$= \ln p(x_1|\theta) + \ln p(x_2|\theta) + \int_{-\infty}^{\infty} \ln p(x_3|\theta)\, p(x_{32}|\theta^0; x_{31}=2)\, dx_{32}$$

$$= \ln p(x_1|\theta) + \ln p(x_2|\theta) + 2e\int_{-\infty}^{\infty} \ln p\left(\binom{2}{x_{32}}\Big|\theta\right)\cdot p\left(\binom{2}{x_{32}}\Big|\theta^0\right) dx_{32}$$

$$= \ln p(x_1|\theta) + \ln p(x_2|\theta) + k$$

For the K:

1. $3 \le \theta_2 \le 4$:

$$k = \frac{1}{4}\int_0^{\theta_2} \ln\left(\frac{1}{\theta_1} e^{-2\theta_1} \frac{1}{\theta_2}\right) dx_{32}$$

$$= \frac{1}{4}\theta_2 \ln\left(\frac{1}{\theta_1} e^{-2\theta_1} \frac{1}{\theta_2}\right)$$

2, $\theta_2 \geq 4$:

$$k = \frac{1}{4}\int_0^{4} \ln\left(\frac{1}{\theta_1} e^{-2\theta_1} \frac{1}{\theta_2}\right) dx_{32}$$

$$= \frac{1}{4} 4 \ln\left(\frac{1}{\theta_1} e^{-2\theta_1} \frac{1}{\theta_2}\right)$$

$$= \ln\left(\frac{1}{\theta_1} e^{-2\theta_1} \frac{1}{\theta_2}\right)$$

3, otherwise $k = 0$

$$\therefore Q(\theta, \theta^0) = \ln p(x_1|\theta) + \ln p(x_2|\theta) + k$$

$$= \ln\left(\frac{1}{\theta_1} e^{-\theta_1} \frac{1}{\theta_2}\right) + \ln\left(\frac{1}{\theta_1} e^{-3\theta_1} \frac{1}{\theta_2}\right) + k$$

$$= -\theta_1 - \ln(\theta_1\theta_2) - 3\theta_1 - \ln(\theta_1\theta_2) + k$$

$$= -4\theta_1 - 2\ln(\theta_1\theta_2) + k$$

consider the normalization condition $\int_{-\infty}^{\infty} p(x_1) dx_1 = 1$

$$\int_{-\infty}^{\infty} \frac{1}{\theta_1} e^{-\theta_1 x_1} dx_1 = 1 \qquad \therefore \theta_1 = 1$$

(b) 1, $3 \leq \theta_2 \leq 4$,

$$Q(\theta;\theta^0) = -4 - \left(2\ln\theta_2 + \frac{1}{4}\theta_2(2 + \ln\theta_2)\right)$$

$$\underset{\theta_2}{\arg\max}\, Q(\theta;\theta^0) = 3$$

2. $\theta_2 \geq 4$: $Q(\theta; \theta^0) = -6 - 3\ln\theta_2$,

$\mathrm{argmax}_{\theta_2} Q(\theta; \theta^0) = 4$

$\therefore \cancel{\theta =} \quad \theta = \begin{pmatrix}1\\3\end{pmatrix}$

(c) see attached figure.

5.4. (a) to minimize $\|x - x_a\|^2$ subject to $g(x) = 0$

the object function $f(x, \lambda) = \|x - x_a\|^2 + 2\lambda[g(x)]$.

$\lambda$ is a Lagrange undetermized multiplier. the point $x_a$ is fixed whil

$x$ varies on the hyperplane,

$$f(x, \lambda) = \|x - x_a\|^2 + 2\lambda[w^t x + w_0]$$

$$= (x - x_a)^t (x - x_a) + 2\lambda(w^t x + w_0)$$

$$= x^t x - 2x^t x_a + x_a^t x_a + 2\lambda(x^t w + w_0)$$

$$\frac{\partial f(x, \lambda)}{\partial x} = x - x_a + \lambda w = 0$$

$$\frac{\partial f(x,\lambda)}{\partial \lambda} = w^t x + w_0 = 0$$

$$x = x_a - \lambda w$$

$$w^t x + w_0 = w^t(x_a - \lambda w) + w_0$$

$$= w^t x_a + w_0 - \lambda w^t w$$

$$= 0$$

$$\therefore \lambda w^t w = \cancel{w^t w_0 +} w^t x_0 + w_0$$

$$\lambda = \frac{w^t x_0 + w_0}{w^t w}$$

$$x = x_a - \lambda w = \begin{cases} x_a - \left[\frac{w^t x_a + w_0}{w^t w}\right] w & \text{if } w \neq 0 \\ x_a & \text{if } w = 0 \end{cases}$$

$$\|x - x_a\| = \left\| x_a - \left[\frac{w^t x_a + w_0}{w^t w}\right] w - x_a \right\|$$

$$= \left\| \left(\frac{w^t x_a + w_0}{w^t w}\right) w \right\|$$

$$= \frac{|g(x_a)| \|w\|}{\|w\|^2} = \frac{|g(x_a)|}{\|w\|}$$

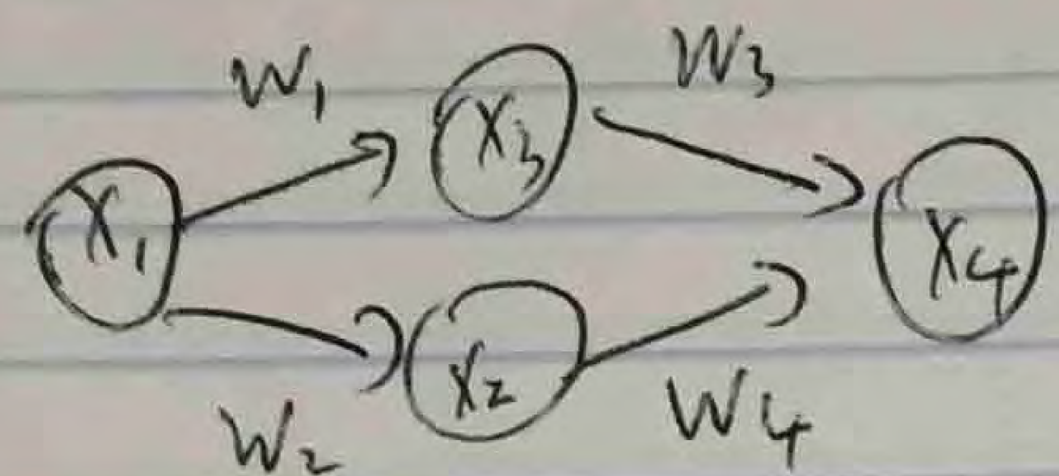
(b) the projection of $x_a$ onto the hyperplane is

$$x_0 = x_a - \lambda w$$

$$= x_a - \frac{g(x_a)}{\|w\|^2} w$$

$$\lambda = \frac{w^t x_a + w_0}{w^t w} = \frac{g(x_a)}{\|w\|^2}$$

$S(x) = \frac{1}{1+e^{-x}}$ $\quad S'(x) = S(x)(1-S(x)) =$




$\cancel{X_3 = S(\Sigma}$ $\quad X_3 = S(w_1 \cdot X_1) \quad X_2 = S(W_2 \cdot X_1)$

$X_4 = S(X_3 \cdot w_3 + X_2 \cdot w_4) =$

$\Delta W_1 = \frac{-\partial E}{\partial w_1}\eta \quad \Delta W_3 = \frac{-\partial E}{\partial w_3}\eta \quad \Delta W_4 = \frac{-\partial E}{\partial W_4}\eta \quad \Delta W_2 = \frac{-\partial E}{\partial w_2}\eta$

$\Delta W_3 = \frac{-\partial E}{\partial W_3} = \frac{-\partial E}{\partial net_3} \cdot \frac{\partial net_3}{\partial w_3} \pm \eta$

let $E = (d - X_4)^2$ $\quad d$ is the desired output.

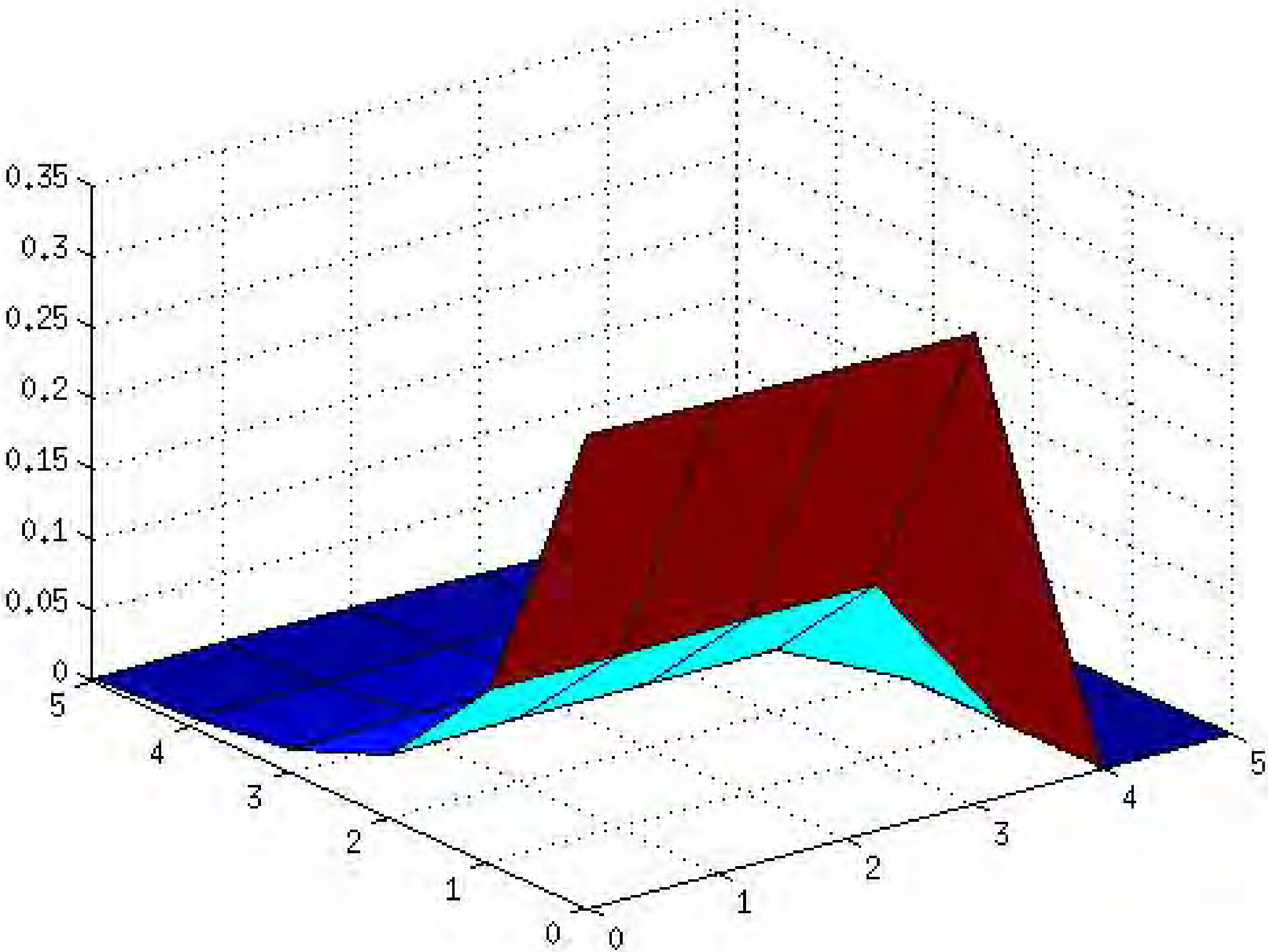
$\Delta W_3 = +2(d - X_4) s'(net_4) X_3 \eta = +2(d - X_4) X_4 (1 - X_4) X_3 \eta$

$\Delta W_4 = +2(d - X_4) s'(net_4) X_2 \eta = +2(d - X_4) X_4 (1 - X_4) X_2 \eta$

$\Delta W_1 = \frac{-\partial E}{\partial w_1}\eta = \frac{-\partial E}{\partial net_3} \cdot \frac{\partial net_3}{\partial net_1} \cdot \frac{\partial net_3}{\partial w_1}\eta =$

$\cancel{2(\ldots)}$ $(d - X_4) X_4 (1 - X_4) W_3 X_3 (1 - X_3) X_1 \eta$

$\Delta W_2 = (d - X_4) X_4 (1 - X_4) W_4 X_2 (1 - X_2) X_1 \eta$


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
5
4
3
2
1
0
0
1
2
3
4
5