Q1

prior prob of c — likelihood of c

a) $$p(c|d) = \frac{p(c)\,p(d|c)}{p(d)} = \frac{p(c)\,p(d|c)}{\sum_{c} p(d,c)}$$

normalising constant

b) $$p(d=6 \mid c=t) = 1/6 \qquad p(h) = p(c=t) = 1/2$$

$$p(d=6 \mid c=h) = 5/36 \quad [\,(1,5),\,(5,1),\,(3,3),\,(4,2),\,(2,4)]$$

$$\therefore\ p(c=h \mid d=6) = \frac{\cancel{p(c=h)}\,p(d=6|c=h)}{\cancel{p(c=h)}\,p(d=6|c=h) + \cancel{p(c=t)}\,p(d=6|c=t)}$$

$$= \frac{5/36}{5/36 + 1/6} = \frac{5}{11}$$

Q2

a) $\lambda_{ML} = \underset{\lambda}{\arg\max}\ p(\{t_n\}_{n=1}^N \mid \lambda) = \underset{\lambda}{\arg\max} \log p(\{t_n\}_{n=1}^N$

where

$$p(\{t_n\}_{n=1}^N \mid \lambda) = \prod_{n=1}^N p(t_n \mid \lambda) = \frac{1}{\lambda^N} \exp\left(-\sum_n t_n / \lambda\right)$$

b) $$\frac{d}{d\lambda}\left[-N \log \lambda - \frac{1}{\lambda} \sum_n t_n \right]\Bigg|_{\lambda = \lambda_{ML}} = 0$$

$$-\frac{N}{\lambda_{ML}} + \frac{1}{\lambda_{ML}^2} \sum_{n=1}^N t_n = 0$$

$$\Rightarrow \lambda_{ML} = \frac{1}{N} \sum_{n=1}^N t_n$$

c) use running averages

let $$\lambda_{ML}^{(N)} = \frac{1}{N} \sum_{n=1}^N t_n = \frac{1}{N} S_N = \frac{N-1}{N} \lambda_{ML}^{(N-1)} + \frac{1}{N} t_N$$

$\Rightarrow$ only need to store in memory $N-1$ & either S_{N-1} or $\lambda_{ML}^{(N-1)}$

# of data seen so far @ previous step

previous sum

or previous estimate of λ

Q3

a) $\mu_{MAP} = \underset{\mu}{\arg\max}\ p(\mu \mid y, \sigma_\mu^2) = \underset{\mu}{\arg\max}\ \log p(\mu, y \mid \sigma_\mu^2)$

b) $$\frac{d}{d\mu}\left[\log p(\mu \mid \sigma_\mu^2) + \log p(y \mid \mu)\right]\Bigg|_{\mu_{MAP}} = 0$$

$$\frac{d}{d\mu}\left[-\frac{1}{2\sigma_\mu^2}\mu^2 + -\frac{1}{2}(y-\mu)^2\right]\Bigg|_{\mu_{MAP}} = 0$$

$$-\frac{1}{\sigma_\mu^2}\mu_{MAP} - (\mu_{MAP} - y) = 0$$

$$\Rightarrow \mu_{MAP} = \frac{y}{1 + 1/\sigma_\mu^2}$$

c) $\mu_{MAP} \rightarrow \mu_{ML}$ as $\sigma_\mu^2 \rightarrow \infty$

Q4

a) $p(y_1|x,w) = \dfrac{1}{1+e^{-wx+b}}$

$b=0$ (centred)

$w=-10$

b) $y_2 = y_1(\varepsilon\sigma_1 + \mu_1 + c_1 x) + (1-y_1)(\varepsilon\sigma_2 + \mu_2 + c_2 x)$

$\sigma_1 = \frac{1}{2}$ $\mu_1 = 2$ $c_1 = 4$

$\sigma_2 = \frac{1}{10}$ $\mu_2 = -2$ $c_2 = -2$

approximate values we fine here

|||

$p(y_2|y_1,x) = y_1 N(y_2; c_1 x + \mu_1, \sigma_1^2) + (1-y_1) N(y_2; c_2 x + \mu_2, \sigma_2^2)$

Q5

a) Assignment step

$$s_n = \underset{k}{\arg\min} \; \| \underline{x}_n - \underline{m}_k \|$$

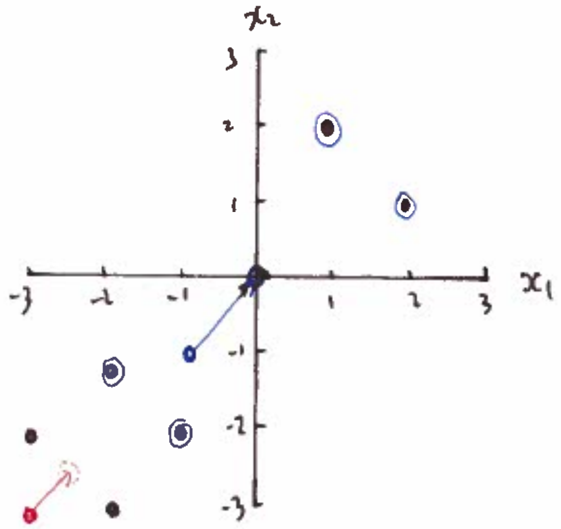
(assign each data point $\underline{x}_n$ to nearest cluster k^*)

Update step

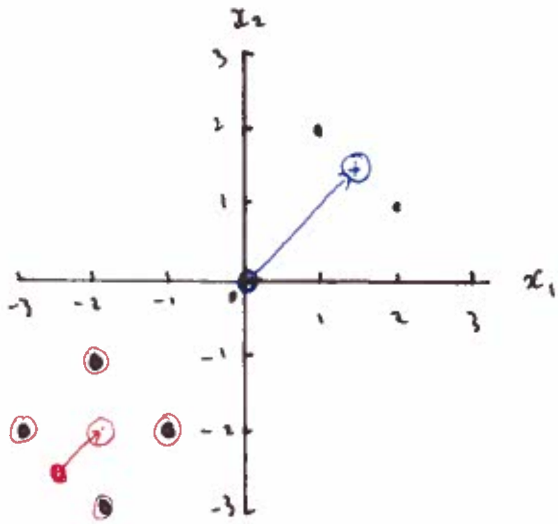
$$\underline{m}_k = \text{mean}(\underline{x}_n : s_n = k)$$

(update cluster centres $\underline{m}_k$ to be mean of data points assigned to that cluster)

b)



initial	(-1, -1)	(-3, -3)
update 1	(0, 0)	(-2.5, -2.5)
update 2	(1.5, 1.5)	(-2, -2)



Q6

a) $\alpha = 0.1$ as the KL divergence is minimized when $Q = P$

b) $$KL(P \| Q) = P_1 \log \frac{P_1}{\alpha} + P_2 \log \frac{P_2}{1-2\alpha} + P_3 \log \frac{P_3}{\alpha}$$

$$\left.\frac{dKL}{d\alpha}\right|_{\alpha=\alpha^*} = \left.\left(-\frac{P_1}{\alpha} + \frac{2P_2}{1-2\alpha} - \frac{P_3}{\alpha}\right)\right|_{\alpha=\alpha^*} = 0$$

$$\Rightarrow \quad (P_1+P_3)(1-2\alpha^*) = 2P_2\alpha^*$$

$$\frac{P_1+P_3}{2} = \alpha^*$$

$$\Rightarrow \quad \alpha^* = \frac{0.6+0.2}{2} = 0.4$$

$$\left.\frac{d^2}{d\alpha^2} KL\right|_{\alpha=\alpha^*} = \frac{0.6}{(\alpha^*)^2} + \frac{0.4}{(1-2\alpha^*)^2} + \frac{0.2}{(\alpha^*)^2} > 0 \Rightarrow \text{minim}$$

Q7

a) $p(x_1, x_2) = p(x_1)\, p(x_2 | x_1)$

$p(x_1 = 0, x_2 = 0) = 4/10 \;\; = 2/5$

$p(x_1 = 1, x_2 = 0) = 1/10$

$p(x_1 = 0, x_2 = 1) = 1/10$

$p(x_1 = 1, x_2 = 1) = 4/10 \;\; = 2/5$

b) $p(y_1, y_2) = \sum_{x_1, x_2} p(x_1, x_2)\, p(y_1 | x_1)\, p(y_2 | x_2)$

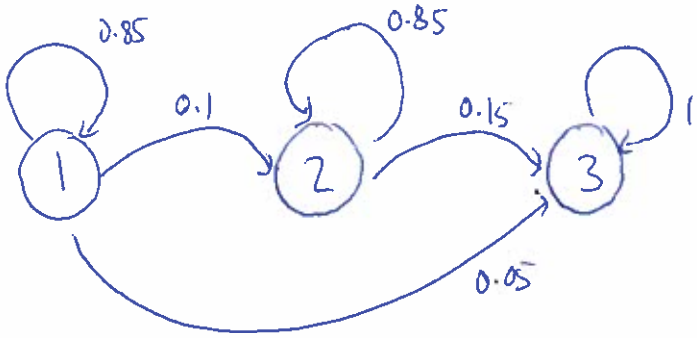
$$= \frac{2}{5} \cdot N\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} ; \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \underline{\underline{I}} \right) + \frac{1}{10} \cdot N\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} ; \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \underline{\underline{I}} \right)$$

$$+ \frac{1}{10} N\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} ; \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \underline{\underline{I}} \right) + \frac{2}{5} N\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} ; \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \underline{\underline{I}} \right)$$

c) $p(y_{1:T})$ is a mixture of Gaussians with 2^T components each weighted by the probability of the latent sequence $p(x_{1:T})$ that produced them.

Q8

a)



b) every bike ends up in state 3

$$\begin{bmatrix} p(S_\infty = 1) \\ p(S_\infty = 2) \\ p(S_\infty = 3) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

c)

$$\begin{bmatrix} 0.85 & 0 & a \\ 1 & 0.85 & b \\ 0.05 & 0.15 & 1-a-b \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \Rightarrow \begin{aligned} a &= 0.15 \\ b &= 0.05 \\ c &= 0.8 \end{aligned}$$

ie. $\underline{\underline{T}}\,\underline{p}_\infty = \underline{p}_\infty$