Question 1

1a) $\mathcal{L}(\lambda) = \prod_{n=1}^{N} p(x_n|\lambda) = \prod_{n=1}^{N} \lambda e^{-\lambda x_n}$

$$\ell(\lambda) = \log \mathcal{L}(\lambda) = \sum_{n=1}^{N} \left[\log\lambda + \log e^{-\lambda x_n}\right] = N\log\lambda - \lambda \sum_{n=1}^{N} x_n$$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = 0 \Rightarrow \frac{N}{\lambda} - \sum_{n=1}^{N} x_n = 0 \Rightarrow \lambda = \frac{N}{\sum_{n=1}^{N} x_n} = \frac{1}{\bar{x}}$$

1b) $p(\lambda|\underline{X}) \propto p(\underline{X}|\lambda)\, p(\lambda)$

$$\propto \left(\prod_{n=1}^{N} \lambda e^{-\lambda x_n}\right) \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$\propto \lambda^N e^{-\lambda \sum x_n} \lambda^{a-1} e^{-b\lambda}$$

$$\propto e^{-\lambda(b + \sum_{n=1}^{N} x_n)} \lambda^{N+a-1} \sim \mathrm{Gamma}\left(N+a-1,\ b + \sum_{n=1}^{N} x_n\right)$$

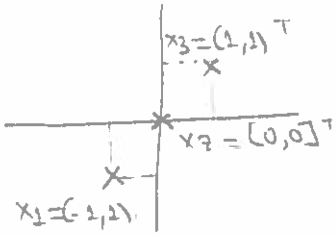
1c) $\log p(\lambda|x) = -\lambda\left(b + \sum x_n\right) + (N+a-1)\log\lambda$

$$\frac{\partial \log p(\lambda|x)}{\partial \lambda} = -\left(b + \sum x_n\right) + \frac{N+a-1}{\lambda} = 0 \Rightarrow \lambda = \frac{N+a-1}{b + \sum x_n}$$

$$\Rightarrow \overset{MAP}{\lambda} = \frac{1 + \frac{a-1}{N}}{\frac{b}{N} + \frac{1}{N}\sum x_n} = \frac{1 + \frac{a-1}{N}}{\frac{b}{N} + \bar{x}} \xrightarrow{N \to \infty} \frac{1}{\bar{x}} \quad (MAP \to MLE)$$

Question 2

2 a)



$$\vec{v} = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)^T$$

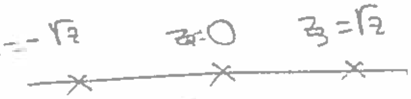
$\frac{\sqrt{2}}{2} \rightarrow \cos 45^\circ$, $\frac{\sqrt{2}}{2} \rightarrow \sin 45^\circ$

$\vec{x}_1 = [-1,-1]^T$

$\vec{x}_2 = [0,0]^T$

$\vec{x}_3 = [1,1]^T$

2 b). $z_1 = \vec{x}_1^T \vec{v} = -\sqrt{2}$



$z_2 = \vec{x}_2^T \vec{v} = 0$

$z_3 = \vec{x}_3^T \vec{v} = \sqrt{2}$

$\mathbb{E}\{z_1, z_2, z_3\} = 0$

$$\operatorname{Var}\{z_1, \dots, z_3\} = \frac{1}{3} \sum_{n=1}^{3} (z_n - \bar{z}) = \frac{1}{3}\left[(-\sqrt{2})^2 + 0^2 + (\sqrt{2})^2\right] = \frac{4}{3}$$

2 c) All points perfectly located on the direction of the first principal component

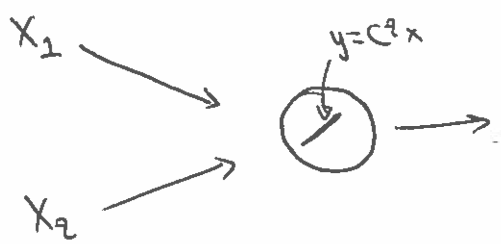
$\hat{\vec{x}}_1 = z_1 \cdot \vec{v} = -\sqrt{2}\left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T = [-1,-1]^T$

$\hat{\vec{x}}_2 = z_2 \vec{v} = [0,0]^T$

$\hat{\vec{x}}_3 = z_3 \vec{v} = \sqrt{2}\left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T = [1,1]^T$

Question 3

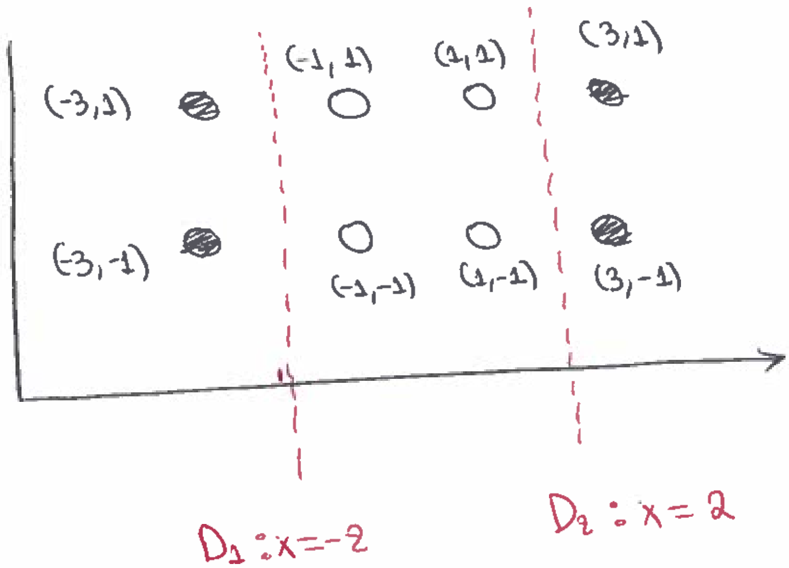
a) $Y = C^2 (w_5 w_1 + w_6 w_2) x_1 + C^2 (w_5 w_3 + w_6 w_4) x_2$



b) Yes, each layer can be thought as a linear transformation of the previous one, implemented by matrix multiplication. Therefore we can simply multiply matrices to find a representation with single layer

c) Hidden layer will contain linear activation functions.
Output layer will contain sigmoid activation.

Question 4a



$$\phi_1(x) = \begin{cases} 1, & x > -2 \\ -1, & \text{otherwise} \end{cases}$$

$$\phi_2(x) = \begin{cases} 1, & x < 2 \\ -1, & \text{otherwise} \end{cases}$$

We have built a classifier $f_{t-1}(\vec{x})$ and want to improve it by adding a new classifier $f(\vec{x}) = f_{t-1}(\vec{x}) + \beta_t h_t(\vec{x})$

Can we choose optimaly $h_t(\vec{x})$ and β_t ?

$\boxed{\text{Question 4b.i}}$

▷ Minimize exponential loss function:

$$(h_t^*(\vec{x}), \beta_t^*) = \arg\min \sum e^{-y_n f(\vec{x}_n)}$$

$$= \arg\min \sum e^{-y_n\left[f_{t-1}(\vec{x}_n) + \beta_t h_t(\vec{x}_n)\right]}$$

$$= \arg\min \sum \underbrace{e^{-y_n f_{t-1}(\vec{x}_n)}}_{w_t(n)} \cdot e^{-y_n \beta_t h_t(\vec{x}_n)}$$

$$= \arg\min \sum w_t(n)\, e^{-y_n \beta_t h_t(\vec{x}_n)}$$

$$y_n h_t(\vec{x}_n) = \begin{cases} 1, & \text{correctly classifed } h_t(\vec{x}_n) = y_n \\ -1, & \text{otherwuse} \end{cases} \quad = \arg\min \left[\sum_n w_t(n)\, e^{\beta_t}\, \mathbb{I}\left[y_n \neq h_t(\vec{x}_n)\right] + \sum_n w_t(n)\, e^{-\beta_t} \overbrace{\left(1 - \mathbb{I}\left[y_n \neq h_t(x_n)\right]\right)}^{\mathbb{I}(y_n = h_t(\vec{x}_n))}\right]$$

$$= \arg\min \left(e^{\beta_t} - e^{-\beta_t}\right) \sum_n w_t(n)\, \mathbb{I}\left[y_n \neq h_t(\vec{x}_n)\right] + e^{-\beta_t} \sum w_t(n)$$

$$E=\left(e^{\beta_t}-e^{-\beta_t}\right)\sum w_t(n)\,\mathbb{I}\left[y_n\neq h_t(\vec{x}_n)\right]+e^{-\beta_t}\sum w_t(n)$$

$$\frac{\partial E}{\partial \beta_t}=\left(e^{\beta_t}+e^{-\beta_t}\right)\sum w_t(n)\,\mathbb{I}\left[y_n\neq h_t(\vec{x}_n)\right]-e^{-\beta_t}\sum w_t(n)=0$$

Question 4b.ii

$$\left(e^{2\beta_t}+1\right)\varepsilon_t-1=0$$

$$\varepsilon_t e^{2\beta_t}+\varepsilon_t-1=0$$

$$\varepsilon_t e^{2\beta_t}=1-\varepsilon_t$$

$$e^{2\beta_t}=\frac{1-\varepsilon_t}{\varepsilon_t}$$

$$\boxed{\beta_t=\frac{1}{2}\log\frac{1-\varepsilon_t}{\varepsilon_t}}$$