ML 2 Exercise sheet 1

Ex 1 (1) $\mathcal{E}(W) = \sum_i |\vec{X_i} - \sum_j W_{ij} \vec{X_j}|^2$

$\vec{X_i} \to \alpha \vec{X_i}$ $\quad \mathcal{E}_1(W) = \sum_i |\alpha \vec{X_i} - \sum_j W_{ij} \alpha \vec{X_j}|^2$

$= \sum_i |\alpha \vec{X_i} - \alpha \sum_j W_{ij} X_j|^2$

$= \alpha^2 \sum_i |\vec{X_i} - \sum_j W_{ij} \vec{X_j}|^2$

$= \alpha^2 \mathcal{E}(W)$

Since $\alpha \in \mathbb{R}^+ \setminus \{0\}$

$\min_W \mathcal{E}_1(W) = \min_W \alpha^2 \mathcal{E}(W)$

(2) $\vec{X_i} \to \vec{X_i} + \vec{V}$ $\quad \mathcal{E}_2(W) = \sum_i |(\vec{X_i} + \vec{V}) - \sum_j W_{ij}(X_j + \vec{V})|^2$

$= \sum_i |(\vec{X_i} + \vec{V}) - \sum_j W_{ij} X_j - \sum_j W_{ij} \vec{V}|^2$

$= \sum_i |\vec{X_i} - \sum_j W_{ij} X_i|$ because $\sum_j W_{ij} = 1$

$= \mathcal{E}(W)$

(3) $\vec{X_i} \to U \cdot \vec{X_i}$ $\quad \mathcal{E}_3(W) = \sum_i |U \cdot \vec{X_i} - \sum_j W_{ij} \cancel{X} U X_j|^2$

$= \sum_i |U \cdot (\vec{X_i} - \sum_j W_{ij} \vec{X_j})|^2$ $\qquad$ ~~$|\vec{X}|^2 = X^T \cdot X$~~

$= \sum_i (\vec{X_i} - \sum_j W_{ij} \vec{X_j})^T \cdot U^T \cdot U (\vec{X_i} - \sum_j W_{ij} \vec{X_j})$

as U is orthogonal matrix, $U^T U = \underline{1}$

so $\mathcal{E}_3(W) = \mathcal{E}(W)$

Exercise 2:

(i)
$$\begin{aligned}
\mathcal{E} &= w^T C w \\
&= w^T (\mathbb{1}\vec{x}^T - \eta)(\mathbb{1}\vec{x}^T - \eta)^T w \\
&= (w^T \mathbb{1}\vec{x}^T - w^T\eta)(w^T\mathbb{1}\vec{x}^T - w^T\eta)^T \\
&= (\vec{x}^T - w^T\eta)(\vec{x}^T - w^T\eta)^T \quad \text{given } w^T\mathbb{1} = 1 \\
&= |\vec{x}^T - w^T\eta|^2 \\
&= |\vec{x}^T - \sum_k w_k \eta_k|^2 \\
&= \left|\sum_j w_j(\vec{x} - \vec{\eta}_j)\right|^2 \\
&= \sum_{jk} w_j w_k C_{jk}
\end{aligned}$$

(ii) Given constraint optimization problem

$$\min_w \; w^T C w \quad \text{Subject to} \quad w^T\mathbb{1} = 1$$

Applying Lagrange's

$$L = w^T C w + \lambda(1 - \mathbb{1}^T w)$$

$$\frac{\partial L}{\partial w} = 2Cw - \lambda\mathbb{1} = 0.$$

$$w = \frac{\lambda C^{-1}\mathbb{1}}{2} \quad \ldots\ldots \text{(1)}.$$

we know $\mathbb{1}^T w = \dfrac{\lambda \mathbb{1}^T C^{-1}\mathbb{1}}{2} = 1$

$$\Rightarrow \lambda = \frac{2}{\mathbb{1}^T C^{-1}\mathbb{1}} \quad \ldots \text{(2)}$$

Substituting (2) in (1)

$$w = \frac{C^{-1}\mathbb{1}}{\mathbb{1}^T C^{-1}\mathbb{1}}$$

iii) $L = \frac{1}{2} w^T C w - \lambda (w^T \mathbb{1} - 1)$

$$\frac{\partial L}{\partial w} = C w - \mathbb{1} = 0$$

$$C w = \mathbb{1}$$

$$w = C^{-1} \mathbb{1} \quad \cdots \quad (1)$$

Rescaling $w^T \mathbb{1} = 1$

$$(C^{-1} \mathbb{1})^T \mathbb{1} = 1$$

$$\mathbb{1}^T C^{-1} \mathbb{1} = 1$$

From (1)

$w = \frac{C^{-1} \mathbb{1}}{1} = \frac{C^{-1} \mathbb{1}}{\mathbb{1}^T C^{-1} \mathbb{1}}$ which is same as obtained in ii)

Hence minimum w can be found from equation $Cw = \mathbb{1}$

ML Ex1 Ex3: i) $C = D_{KL}(P\|Q) = \sum_j p_j \log \frac{p_j}{q_j}$ $\qquad \sum_j p_j = 1 = \sum_j q_j$

$\Leftrightarrow \sum_j p_j (\log p_j - \log q_j)$

$\Leftrightarrow \sum_j p_j \log p_j - p_j \log q_j$

$$\frac{\partial C}{\partial q_i} = - \frac{p_i}{q_i}$$

ii) $q_i = \frac{e^{x_i}}{\sum_k e^{x_k}}$ $\quad \Rightarrow C = \sum_i p_i \log p_i - p_i \log \frac{e^{x_i}}{\sum_k e^{x_k}}$

$\Leftrightarrow \sum_i p_i \log p_i - p_i \log e^{x_i} + p_i \log \sum_k e^{x_k}$ $\quad$ with $i = k$

$\Leftrightarrow \sum_i p_i \log p_i - \sum_i p_i x_i + \underbrace{\sum_i p_i}_{=1} \log \sum_i e^{x_i}$

$\Leftrightarrow \sum_i p_i \log p_i - \sum_i p_i x_i + \log \sum_i e^{x_i}$

$$\frac{\partial C}{\partial x_i} = -p_i + \underbrace{\frac{1}{\sum_i e^{x_i}} \cdot e^{x_i}}_{= q_i} = -p_i + q_i$$

iii) 1) $\frac{\partial C}{\partial x_i}$ is more stable as $q_i = \frac{e^{x_i}}{\sum_i e^{x_i}}$ is normalized and thus between 0 and 1.

2) $\frac{\partial C}{\partial x_i}$ has a higher ability to produce a valid probability distribution as for $\frac{\partial C}{\partial q_i} = -\frac{p_i}{q_i}$ some of the divisions (for i) might not lie within 0 and 1.