ML Sheet 01

(1) $\cdot \| \alpha \cdot \vec{x} - \alpha \cdot \vec{y} \| = \underbrace{|\alpha|}_{>0} \cdot \| \vec{x} - \vec{y} \|$

$\cdot \| (\vec{x} + \vec{v}) - (\vec{y} + \vec{v}) \| = \| \vec{x} - \vec{y} \|$

$\cdot \| U\vec{x} - U\vec{y} \| = \langle U(\vec{x} - \vec{y}), U(\vec{x} - \vec{y}) \rangle^{1/2} = \langle \underbrace{U^T U}_{= Id} (\vec{x} - \vec{y}), \vec{x} - \vec{y} \rangle^{1/2} = \| \vec{x} - \vec{y} \|$ $(*)$

$\Rightarrow$ The sets $N_i := \{ j \in \{1, .., N\} \mid x_j$ is one of the K nearest neighbours of $x_i \}$

don't change under the affine-linear transformations in (a), (b), (c).

In particular, the sum $\sum_j$ is taken over the same indices.

$\cdot \sum_{i=1}^{N} \| \alpha \cdot \vec{x}_i - \sum_{j \in N_i} w_{ij} (\alpha \cdot \vec{x}_j) \|^2 = |\alpha| \cdot \sum_{i=1}^{N} \| \vec{x}_i - \sum_{j \in N_i} w_{ij} \vec{x}_j \|^2$

$= \underbrace{|\alpha|}_{>0} \cdot \mathcal{E}(w)$

$\cdot \sum_{i=1}^{N} \| (\vec{x}_i + \vec{v}) - \sum_{j \in N_i} w_{ij} (\vec{x}_j + \vec{v}) \|^2 = \sum_{i=1}^{N} \| \vec{x}_i - \sum_{j \in N_i} w_{ij} \vec{x}_j + \vec{v} - \vec{v} \underbrace{\sum_{j \in N_i} w_{ij}}_{=1} \|^2$

$= \mathcal{E}(w)$

$\cdot \sum_{i=1}^{N} \| U\vec{x}_i - \sum_{j \in N_i} w_{ij} (U\vec{x}_j) \|^2 = \sum_{i=1}^{N} \| U (\vec{x}_i - \sum_{j \in N_i} w_{ij} x_j) \|^2$

$\overset{(*)}{=} \mathcal{E}(w)$

Hence, the minimum is invariant in all cases.

(2) (a) $\mathcal{E}_i(w) = \| \vec{x} - \sum_{j=1}^{K} w_j \vec{\eta}_j \|^2 = \| \vec{x} \|^2 - 2 \langle \vec{x}, \sum_{j=1}^{K} w_j \vec{\eta}_j \rangle + \| \sum_{j=1}^{K} w_j \vec{\eta}_j \|^2$

$= \underbrace{w^T \mathbb{1}}_{=1} \| \vec{x} \|^2 \underbrace{\mathbb{1}^T w}_{=1} - 2 \langle \vec{x}, \eta^T w \rangle + \| \eta^T w \|^2$

$= w^T \mathbb{1} \vec{x}^T \vec{x} \mathbb{1}^T w - 2 \vec{x}^T \eta^T w + (\eta^T w)^T \eta^T w$

$= w^T \mathbb{1} \vec{x}^T \vec{x} \mathbb{1}^T w - \underbrace{w^T \mathbb{1}}_{=1} \vec{x}^T \eta^T w - (\underbrace{w^T \mathbb{1}}_{=1} \underbrace{\vec{x}^T \eta^T w}_{\text{scalar}})^T + w^T \eta \eta^T w$

$= w^T \mathbb{1} \vec{x}^T (\mathbb{1} \vec{x}^T)^T w - w^T \mathbb{1} \vec{x} \eta^T w - w^T \eta (\mathbb{1} \vec{x}^T)^T w + w^T \eta \eta^T w$

$= w^T (\mathbb{1} \vec{x}^T - \eta)(\mathbb{1} \vec{x}^T - \eta)^T w$

$= w^T C w$

Since $\sum_{j=1}^{k} w_j = w^T \mathbb{1}$, the problems are equivalent.

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

- $\frac{\partial}{\partial w} \mathcal{L}(w,\lambda) = Cw + (w^T C)^T + \lambda \mathbb{1}$

$$= 2Cw + \lambda \mathbb{1} \qquad (\text{since } C^T = C)$$

$$\overset{!}{=} 0$$

$$\Leftrightarrow \quad w = -\frac{\lambda}{2} C^{-1} \mathbb{1} \qquad (\mathrm{I})$$

- $\frac{\partial}{\partial \lambda} \mathcal{L}(w,\lambda) = w^T \mathbb{1} - 1 \overset{!}{=} 0 \qquad (\mathrm{II})$

Plugging (I) into (II) yields:

$$0 = -\frac{\lambda}{2} (C^{-1} \mathbb{1})^T \mathbb{1} - 1$$

$$= -\frac{\lambda}{2} \mathbb{1}^T \underbrace{C^{-T}}_{= C^{-1}} \mathbb{1} - 1$$

$$\Leftrightarrow \quad \lambda = \frac{-2}{\mathbb{1}^T C^{-1} \mathbb{1}}$$

Hence, by (I), we get

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

(c) Solving $C\tilde{w} = \mathbb{1}$ gives $\tilde{w} = C^{-1} \mathbb{1}$. Setting $w = \alpha \cdot \tilde{w}$ and requiring

$$1 \overset{!}{=} w^T \mathbb{1} = \alpha (C^{-1} \mathbb{1})^T \mathbb{1} = \alpha \cdot \mathbb{1}^T \underbrace{C^{-T}}_{= C^{-1}} \mathbb{1}$$

yields $\alpha = \frac{1}{\mathbb{1}^T C^{-1} \mathbb{1}}$. Hence we get the same solution as in (b).

③ (a) $$\frac{\partial C}{\partial q_{ij}} = p_{ij} \cdot \frac{1}{\left(\frac{p_{ij}}{q_{ij}}\right)} \cdot (-1)\,\frac{p_{ij}}{q_{ij}^2}$$

$$= -\frac{p_{ij}}{q_{ij}}$$

(b) $\cancel{\partial q_{ij}}$ $$C = \sum_{i,j=1}^{N} p_{ij}\left[\log(p_{ij}) - \log\left(\frac{\exp(z_{ij})}{\sum_{k,l=1}^{N}\exp(z_{kl})}\right)\right]$$

$$= \sum_{i,j=1}^{N} p_{ij}\left[\log(p_{ij}) - z_{ij} + \log\left(\sum_{k,l=1}^{N}\exp(z_{kl})\right)\right]$$

$$\Rightarrow \frac{\partial C}{\partial z_{ij}} = -p_{ij} + \sum_{k,l=1}^{N} p_{kl} \cdot \underbrace{\frac{1}{\left(\sum_{m,n=1}^{N}\exp(z_{mn})\right)} \cdot \exp(z_{ij})}_{= q_{ij}}$$

$$= -p_{ij} + q_{ij} \cdot \underbrace{\sum_{k,l=1}^{N} p_{kl}}_{=1}$$

$$= -p_{ij} + q_{ij}$$

(c) ~~The gradient in (1) is unstable since~~

$$\left|-\frac{p_{ij}}{q_{ij}}\right| \to \infty$$ ~~, for~~ $q_{ij} \searrow 0$. ~~The gradient in (2) on the other hand is bounded since~~

$$|-p_{ij} + q_{ij}| \leq \underbrace{|p_{ij}|}_{\leq 1} + \underbrace{|q_{ij}|}_{\leq 1} \leq 2.$$

(d) $$\frac{\partial C}{\partial \vec{y}_i} = \sum_{j=1}^{N} \frac{\partial C}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial \vec{y}_i}$$

$$= \sum_{j=1}^{N} (-p_{ij} + q_{ij}) \cdot (-2) \cdot (\vec{y}_i - \vec{y}_j)$$

$$= \sum_{j=1}^{N} 2(p_{ij} - q_{ij}) \cdot (\vec{y}_i - \vec{y}_j)$$

(c) • The gradient in (1) is unstable since $\left|\frac{-p_{ij}}{q_{ij}}\right| \to \infty$ for $q_{ij} \to 0$. The one in (2) on the other hand is bounded,

$$|-p_{ij} + q_{ij}| \leq |p_{ij}| + |q_{ij}| \leq 2$$ (as long as $q_{ij} \leq 1$)

• A gradient descent step in (1) yields

$$\tilde{q}_{ij} = q_{ij} - \eta \frac{\partial C}{\partial q_{ij}} = q_{ij} + \eta \frac{p_{ij}}{q_{ij}},$$

hence

$$\sum_{i,j=1}^{N} \tilde{q}_{ij} = \underbrace{\sum_{i,j=1}^{N} q_{ij}}_{=1} + \underbrace{\eta \sum_{i,j=1}^{N} \frac{p_{ij}}{q_{ij}}}_{>0} > 1,$$

showing that $\{\tilde{q}_{ij}\}_{i,j=1}^{N}$ does not define a probability distribution anymore. It is also possible that $\tilde{q}_{ij} > 1$ for some $i,j \in \{1,\dots,N\}$.

In (2) we always have

$$0 \leq \frac{\exp(z_{ij})}{\sum_{k,l=1}^{N} \exp(z_{k,l})} = q_{ij} \leq 1$$

and

$$\sum_{i,j=1}^{N} q_{ij} = \sum_{i,j=1}^{N} \frac{\exp(z_{ij})}{\sum_{k,l=1}^{N} \exp(z_{k,l})} = 1,$$

no matter what values the z_{ij} take, i.e., $\{q_{ij}\}_{i,j=1}^{N}$ always defines a probability distribution.