

- ①
- $\| \alpha \cdot \vec{x} - \alpha \cdot \vec{y} \| = |\alpha| \cdot \| \vec{x} - \vec{y} \|$
 - $\| (\vec{x} + \vec{v}) - (\vec{y} + \vec{v}) \| = \| \vec{x} - \vec{y} \|$
 - $\| u\vec{x} - u\vec{y} \| = \langle u(\vec{x} - \vec{y}), u(\vec{x} - \vec{y}) \rangle = \langle \underbrace{u^T u}_{=I} (\vec{x} - \vec{y}), \vec{x} - \vec{y} \rangle = \| \vec{x} - \vec{y} \| \quad (*)$
- \Rightarrow The sets $N_i = \{j \in \{1, \dots, N\} \mid x_j \text{ is one of the } K \text{ nearest neighbours of } x_i\}$

don't change under the affine-linear transformations in (a), (b), (c).
In particular, the sum \sum is taken over the same indices.

$$\begin{aligned} \sum_{i=1}^N \| \alpha \cdot \vec{x}_i - \sum_{j \in N_i} w_{ij} (\alpha \cdot \vec{x}_j) \|^2 &= |\alpha|^2 \sum_{i=1}^N \| \vec{x}_i - \sum_{j \in N_i} w_{ij} \vec{x}_j \|^2 \\ &= |\alpha|^2 \cdot \mathcal{E}(w) \\ \sum_{i=1}^N \| (\vec{x}_i + \vec{v}) - \sum_{j \in N_i} w_{ij} (\vec{x}_j + \vec{v}) \|^2 &= \sum_{i=1}^N \| \vec{x}_i - \sum_{j \in N_i} w_{ij} \vec{x}_j + \vec{v} - \vec{v} \sum_{j \in N_i} w_{ij} \|^2 \\ &= \mathcal{E}(w) \\ \sum_{i=1}^N \| u\vec{x}_i - \sum_{j \in N_i} w_{ij} (u\vec{x}_j) \|^2 &= \sum_{i=1}^N \| u(\vec{x}_i - \sum_{j \in N_i} w_{ij} \vec{x}_j) \|^2 \\ &\stackrel{(*)}{=} \mathcal{E}(w) \end{aligned}$$

Hence, the minimum is invariant in all cases.

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

$$\begin{aligned} &= \underbrace{\vec{w}^T \mathbb{1}}_{=1} \| \vec{x} \|^2 \underbrace{\mathbb{1}^T \vec{w}}_{=1} - 2 \langle \vec{x}, \vec{q}^T w \rangle + \| \vec{q}^T w \|^2 \\ &= \vec{w}^T \mathbb{1} \vec{x}^T \vec{x} \mathbb{1}^T w - 2 \vec{x}^T \vec{q}^T w + (\vec{q}^T w)^T \vec{q}^T w \\ &= \vec{w}^T \mathbb{1} \vec{x}^T \vec{x} \mathbb{1}^T w - \underbrace{\vec{w}^T \mathbb{1} \vec{x}^T \vec{q}^T w}_{=1} - \underbrace{\left(\vec{w}^T \mathbb{1} \vec{x}^T \vec{q}^T w \right)^T}_{\text{Scalar}} + \vec{w}^T \vec{q} \vec{q}^T w \\ &= \vec{w}^T \mathbb{1} \vec{x}^T (\mathbb{1} \vec{x}^T w - \vec{q}^T w) + \vec{w}^T \vec{q} \vec{q}^T w \\ &= \vec{w}^T (\mathbb{1} \vec{x}^T - \vec{q}^T) (\mathbb{1} \vec{x}^T - \vec{q}^T)^T w \\ &= \vec{w}^T C w \end{aligned}$$

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

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

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

$$= 2C \mathbf{w} + \lambda \mathbf{1}$$

(since $C^T = C$)

$$\stackrel{!}{=} 0$$

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

$$\frac{\partial}{\partial \lambda} \mathcal{L}(w, \lambda) = \mathbf{w}^T \mathbf{1} - 1 \stackrel{!}{=} 0 \quad (II)$$

Plugging (I) into (II) yields:

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

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

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

Hence, by (I) we get

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

(c) Solving $C \tilde{\mathbf{w}} = \mathbf{1}$ gives $\tilde{\mathbf{w}} = C^{-1} \mathbf{1}$. Setting

$\mathbf{w} = \alpha \tilde{\mathbf{w}}$ and requiring

$$1 \stackrel{!}{=} \mathbf{w}^T \mathbf{1} = \alpha (\mathbf{1}^T C^{-1} \mathbf{1}) = \alpha \underbrace{\mathbf{1}^T C^{-1} \mathbf{1}}_{= C^{-1}}$$

yields $\alpha = \frac{1}{\mathbf{1}^T C^{-1} \mathbf{1}}$. Hence we get the

same solution as in (b).

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

$$(b) \cancel{24} C = \sum_{i,j=1}^N p_{ij} \left[\log(p_{ij}) - \log \left(\frac{\exp(z_{ij})}{\sum_{k,l=1}^N \exp(z_{kl})} \right) \right] \\ = \sum_{i,j=1}^N p_{ij} \left[\log(p_{ij}) - z_{ij} + \log \left(\sum_{k,l=1}^N \exp(z_{kl}) \right) \right] \\ \Rightarrow \frac{\partial C}{\partial z_{ij}} = -p_{ij} + \sum_{k,l=1}^N p_{kl} \cdot \frac{1}{\left(\sum_{m,n=1}^N \exp(z_{mn}) \right)} \cdot \exp(z_{ij}) \\ = -p_{ij} + q_{ij} \cdot \underbrace{\sum_{k,l=1}^N p_{kl}}_{=1} = -p_{ij} + q_{ij}$$

(c) The gradient in (a) is unstable since $\left| \frac{-p_{ij}}{q_{ij}} \right| \rightarrow \infty$ for $q_{ij} \rightarrow 0$. The gradient in (2) on the other hand is bounded since $|-p_{ij} + q_{ij}| \leq \underbrace{p_{ij}}_{\leq 1} + \underbrace{q_{ij}}_{\leq 1} \leq 2$.

$$(d) \frac{\partial C}{\partial \vec{y}_i} = \cancel{24} \sum_{j=1}^N \frac{\partial C}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial \vec{y}_i} \\ = \sum_{j=1}^N (-p_{ij} + q_{ij}) \cdot (-2) \cdot (\vec{y}_i - \vec{y}_j) \\ = \sum_{j=1}^N 2(p_{ij} - q_{ij}) \cdot (\vec{y}_i - \vec{y}_j)$$