Machine Learning 2 Homework 1

Group: DLBSP

April 24, 2019

1 Task 1

(i)

$$\begin{aligned}
\varepsilon(W) &= \textstyle\sum_i \left|\alpha\vec{X}_i - \sum_j W_{ij}\alpha\vec{X}_j\right|^2 \\
&= \textstyle\sum_i \left|\alpha\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right)\right|^2 \\
&\overset{\alpha\geq 0}{=} \textstyle\sum_i \sqrt{\alpha}\left|\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right)\right|^2 \\
&= \textstyle\sqrt{\alpha}\sum_i \left|\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right)\right|^2
\end{aligned}$$

The multiplication of ε with a scalar doesn't change the outcome of the minimization of ε.

(ii)

$$\begin{aligned}
\varepsilon(w) &= \sum_i \left|(\vec{X}_i + \vec{v}) - \sum_j W_{ij}(\vec{X}_i + \vec{v})\right|^2 \\
&= \sum_i \left|\vec{X}_i + \vec{v} - \sum_j W_{ij}\vec{X}_i - \sum_j W_{ij}\vec{v}\right|^2 \\
&= \sum_i \left|\vec{X}_i + \vec{v} - \sum_j W_{ij}\vec{X}_i - \vec{v}\sum_j W_{ij}\right|^2
\end{aligned}$$

Since the sum of all weights $\sum_j W_{ij} = 1$, we can conclude that:

$$\varepsilon(w) = \sum_i \left|\vec{X}_i + \vec{v} - \sum_j W_{ij}\vec{X}_i + \vec{v}\right|^2 = \sum_i \left|\vec{X}_i - \sum_j W_{ij}\vec{X}_i\right|^2$$

(iii)

$$\begin{aligned}
\varepsilon(W) &= \sum_i \left| U\vec{X}_i - \sum_j W_{ij} U\vec{X}_j \right|^2 \\
&= \sum_i \left| U\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right) \right|^2 \\
&= \sum_i \left\| U\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right) \right\|_2^2 \\
&= \sum_i \langle U\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right), U\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right)\rangle
\end{aligned}$$

Since U is an orthogonal matrix, it holds that

$$\langle Ux, Uy\rangle = \langle x, y\rangle$$

for arbitrary $x \in \mathbb{R}^D$ and $y \in \mathbb{R}^D$. So

$$\begin{aligned}
\varepsilon(W) &= \sum_i \langle \vec{X}_i - \sum_j W_{ij}\vec{X}_j, \vec{X}_i - \sum_j W_{ij}\vec{X}_j\rangle \\
&= \sum_i \left\| \vec{X}_i - \sum_j W_{ij}\vec{X}_j \right\|_2^2 \\
&= \sum_i \left| \vec{X}_i - \sum_j W_{ij}\vec{X}_j \right|^2
\end{aligned}$$

thus, the multiplication of the data with an orthogonal matrix $U \in \mathbb{R}^{D\times D}$ doesn't change the outcome of the minimization of ε.

The data can be transformed in such a way the data is centered around the arbitrary fixed, which is nothing but a linear transformation, which doesn't affect the minima of ε, proved by (ii). The transformed data then gets rotated, which is the multiplication of each data point with an orthogonal rotation matrix. In statement (iii) is was proved that the minima are invariant under such such an operation. The last step would be again a linear retransformation of the centered and rotated data into the primal reference system (statement (ii)). An arbitrary transformation also includes scaling, thus the rotation matrix "contains" a scalar values and can be expressed by $U = \beta\widetilde{U}$ with $\beta \in \mathbb{R}$. Statement (i) proves the invariance for this.

Task 2

(i)

$$\varepsilon = \left| \vec{x} - \sum_j w_j \vec{\eta}_j \right|^2$$

Since $\vec{x} = \sum_j w_j \vec{x}$ because $\sum_j w_j = 1 \iff w^\intercal \vec{1} = 1$ one can obtain

$$
\begin{aligned}
\varepsilon &= \left|\sum_j w_j\left(\vec{x}-\vec{\eta}_j\right)\right|^2 \\
&= \left(\sum_j w_j\left(\vec{x}-\vec{\eta}_j\right)\right)^{\intercal}\left(\sum_j w_j\left(\vec{x}-\vec{\eta}_j\right)\right) \\
&= \sum_{j,k} w_j\left(\vec{x}-\vec{\eta}_j\right)^{\intercal}\left(\vec{x}-\vec{\eta}_k\right) w_k \\
&= \sum_{j,k} w_j C_{jk} w_k \\
&= w^{\intercal} C w
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\min_w \varepsilon(w) = \min_w w^{\intercal} C w \\
&\text{s. t. } w^{\intercal}\vec{1} = 1
\end{aligned}
$$

(ii)

$$
\begin{aligned}
L(w,\lambda) &= \frac{1}{2} w^{\intercal} C w + \lambda(1 - w^{\intercal}\vec{1}) \\
\frac{\partial L}{\partial w} &= Cw - \lambda\vec{1} = 0 \\
\iff Cw &= \lambda\vec{1} \\
\iff w &= \lambda C^{-1}\vec{1}
\end{aligned}
$$

Choose λ in such a way, that the constraint is fulfilled

$$
\lambda = \frac{1}{\sum_j w_j} = \frac{1}{\vec{1}C\vec{1}}
$$

3.1

Thus

$$
w = \frac{C^{-1}\vec{1}}{\vec{1}C\vec{1}}
$$

(iii)

Task 3

$$
\begin{aligned}
\frac{\partial C}{\partial q_i} &= \frac{\partial}{\partial q_i}\left(\sum_j p_j \ln\left(\frac{p_j}{q_j}\right)\right) \\
&= \frac{\partial}{\partial q_i}\left(p_i \ln\left(\frac{p_i}{q_i}\right)\right) \\
&= \frac{\partial}{\partial q_i}\left(p_i\left(\ln\left(p_i\right) - \ln\left(q_i\right)\right)\right) \\
&= p_i \frac{\partial}{\partial q_i}\left(-\ln\left(q_i\right)\right) \\
&= -\frac{p_i}{q_i}
\end{aligned} \tag{1}
$$

3.2

$$
\begin{aligned}
\frac{\partial C(q_i(x_i))}{\partial x_i} &= \frac{\partial}{\partial x_i}\left(\sum_j p_j \ln\left(\frac{p_j \sum_k \exp(x_k)}{\exp(x_j)}\right)\right) \\
&= \frac{\partial}{\partial x_i}\left(\sum_j p_j \left(\ln p_j + \ln\left(\sum_k \exp(x_k)\right) - x_j\right)\right) \\
&= \frac{\partial}{\partial x_i}\left(\sum_j p_j \ln p_j + \sum_j p_j \ln\left(\sum_k \exp(x_k)\right) - \sum_j p_j x_j\right) \\
&= \frac{\partial}{\partial x_i} \sum_j p_j \ln\left(\sum_k \exp(x_k)\right) - p_i \\
&= \sum_j p_j \frac{\exp(x_i)}{\sum_k \exp(x_k)} - p_i \\
&= \sum_j p_j q_i - p_i
\end{aligned}
$$

Because $\sum_j p_j = 1$ it follows

$$
\frac{\partial C(q_i(x_i))}{\partial x_i} = q_i - p_i \tag{2}
$$

5.1

Task 3.2

$$q_{ij} = \frac{e^{z_{ij}}}{\sum_l \sum_m e^{z_{lm}}}$$

$$C = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} = \sum \sum \left[p_{ij} \log p_{ij} - p_{ij} \log q_{ij} \right]$$

$$= \sum_i \sum_j \left[p_{ij} \log p_{ij} - p_{ij} \left(\log e^{z_{ij}} - \log \sum \sum e^{z_{ij}} \right) \right.$$

$$\frac{\partial C}{\partial z_{ij}} = 0 - p_{ij} \frac{\partial}{\partial z_{ij}} \cdot z_{ij} + \sum_k \sum_s p_{ks} \cdot \underbrace{\frac{e^{z_{ij}}}{\sum \sum e^{z_{ij}}}}_{= q_{ij}}$$

$$= -p_{ij} + q_{ij} \cdot \sum_k \sum_s p_{ks} = -p_{ij} + q_{ij}$$

Task 3.4

$$z_{ij} = -\|y_i - y_j\|^2$$

$$C = \sum_l \sum_m p_{lm} \log p_{lm} - p_{lm} \left(\log e^{-\|y_l - y_m\|^2} - \log \sum \sum e^{-\|y_l - y_m\|^2} \right)$$

$$\frac{\partial C}{\partial y_i} = \frac{\partial}{\partial y_i} \left(\sum_l \sum_m -p_{lm} \left(-\|y_l - y_m\|^2 - \log \sum \sum e^{-\|y_l - y_m\|^2} \right) \right)$$

$$= \sum_j 4(y_i - y_j) p_{ij} - 4(y_i - y_j) q_{ij} \underbrace{\sum_k \sum_s p_{ks}}_{=1}$$

$$= \sum_j 4(y_i - y_j)(p_{ij} - q_{ij})$$

Index of comments

3.1 Explain, how you came up with this choice. Also, this equation is incorrect. sum(w_j)=1, so 1^T C 1=1 which means that w is always normalized anyways? Also, missing transposes

3.2 The solution is generally correct, but it is not the problem to the solution that was stated! C has double sums and q is indexed by two indices...

4.1 Why do you keep solving a different problem and use notation/variables that are different from what is stated in the task?

5.1 You already solved this