CSCC11 FINAL EXAM QUESTION 2 WRITTEN

Zhe Fan Li

Dec 9, 2020

1 Problem 1

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{w_j}} &= -\Sigma_{i=1}^{N} \frac{1}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot p(y_i|x_i, \boldsymbol{w}_j, \sigma^2) + \lambda && \text{differentiate using log and chain rule} \quad (1)\\
\implies 0 &= -\Sigma_{i=1}^{N} \frac{1}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot p(y_i|x_i, \boldsymbol{w}_j, \sigma^2) + \lambda && \text{Set to 0} \quad (2)\\
\lambda &= \Sigma_{i=1}^{N} \frac{1}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot p(y_i|x_i, \boldsymbol{w}_j, \sigma^2) && (3)\\
1 &= \frac{1}{\lambda} \cdot \Sigma_{i=1}^{N} \frac{1}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot p(y_i|x_i, \boldsymbol{w}_j, \sigma^2) && (4)\\
1 &= \frac{w_j}{w_j} \frac{1}{\lambda} \cdot \Sigma_{i=1}^{N} \frac{1}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot p(y_i|x_i, \boldsymbol{w}_j, \sigma^2) && \text{Multiply by 1} \quad (5)\\
w_j &= \frac{1}{\lambda} \cdot \Sigma_{i=1}^{N} \frac{w_j}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot p(y_i|x_i, \boldsymbol{w}_j, \sigma^2) && (6)\\
w_j &= \frac{1}{\lambda} \cdot \Sigma_{i=1}^{N} q_{j,i} && \text{From Definition of } q_{j,i} \quad (7)
\end{aligned}
$$

■

2 Problem 2

$$
\begin{aligned}
\frac{\partial L}{\partial \lambda} &= (\lambda(\Sigma_{j=1}^{K} m_j - 1))' && \text{Left term doesn't have } \lambda \text{ in it at all.} \quad (8)\\
&= \Sigma_{j=1}^{K} (m_j - 1) && (9)\\
\implies 1 &= \Sigma_{j=1}^{K} m_j && (10)
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{1}{\lambda} \Sigma_{i=1}^{N} q_{j,i} - w_j && \text{From question 1} \quad (11)\\
\implies 0 + 0 + ... + 0 &= (\frac{1}{\lambda} \Sigma_{i=1}^{N} q_{1,i} - ... - \frac{1}{\lambda} \Sigma_{i=1}^{N} q_{K,i}) + (-w_1 - ... - w_K) && (12)\\
&= \frac{1}{\lambda} \Sigma_{i=1}^{N} \Sigma_{t=1}^{K} q_{t,i} - \Sigma_{t=1}^{K} w_t && (13)\\
&= \frac{1}{\lambda} \Sigma_{i=1}^{N} 1 - \Sigma_{t=1}^{K} w_t && \text{Since } \Sigma_{t=1}^{K} q_{t,i} = 1 \quad (14)\\
&= \frac{1}{\lambda} \Sigma_{i=1}^{N} 1 - 1 && \text{Since } \Sigma_{t=1}^{K} w_t = 1 \quad (15)\\
\implies \lambda &= N && (16)
\end{aligned}
$$

And hence,

$$w_j = \frac{1}{N} \cdot \Sigma_{i=1}^{N} q_{j,i}$$

■

3 Problem 3

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{w_j}} &= -\Sigma_{i=1}^{N} \frac{1}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot m_j \frac{\partial p(y_i|x_i, \boldsymbol{w}_j, \sigma^2)}{\partial w_j} && \text{differentiate using log and chain rule} && (17)\\
&= -\Sigma_{i=1}^{N} \frac{m_j p(y_i|x_i, \boldsymbol{w}_j, \sigma^2)}{\Sigma_{t=1}^{K} m_t p(y_i|x_i, \boldsymbol{w}_t, \sigma^2)} \cdot \frac{\partial logp(y_i|x_i, \boldsymbol{w}_j, \sigma^2)}{\partial w_j} && \text{use identity in hint} && (18)\\
&= -\Sigma_{i=1}^{N} q_{j,i} \cdot \frac{\partial logp(y_i|x_i, \boldsymbol{w}_j, \sigma^2)}{\partial w_j} && \text{From Definition of } q_{j,i} && (19)\\
& && && (20)
\end{aligned}$$

■

4 Problem 4

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{w_j}} &= -\Sigma_{i=1}^{N} q_{j,i} \cdot \frac{\partial logp(y_i|x_i, \boldsymbol{w}_j, \sigma^2)}{\partial w_j} && \text{From Question 3} && (21)\\
&= -\Sigma_{i=1}^{N} q_{j,i} \cdot \frac{\partial log(\frac{1}{\sqrt{2\pi\sigma^2}} exp(-(y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i)^2))/2\sigma^2}{\partial w_j} && && (22)\\
&= -\Sigma_{i=1}^{N} q_{j,i} \cdot \frac{\partial log(\frac{1}{\sqrt{2\pi\sigma^2}}) + log(exp(-(y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i)^2))/2\sigma^2)}{\partial w_j} && && (23)\\
&= -\Sigma_{i=1}^{N} q_{j,i} \cdot \frac{\partial \quad -(y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i)^2/2\sigma^2}{\partial w_j} && && (24)\\
&= -\Sigma_{i=1}^{N} q_{j,i} \cdot \frac{-2(y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i) \cdot -\boldsymbol{x}_i}{2\sigma^2} && && (25)\\
&= -\Sigma_{i=1}^{N} q_{j,i} \cdot \frac{(y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i) \cdot \boldsymbol{x}_i}{\sigma^2} && && (26)\\
\implies 0 &= \Sigma_{i=1}^{N} q_{j,i} \cdot (y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i) \cdot \boldsymbol{x}_i && && (27)
\end{aligned}$$

Note that this is equivalent to minimizing the weighted least squares problem, as

$$\frac{\partial}{\partial \boldsymbol{w}_j} \Sigma_{i=1}^{N} q_{j,i} \cdot (y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i)^2 = \Sigma_{i=1}^{N} q_{j,i} \cdot (y_i - \boldsymbol{w}_j^T \boldsymbol{x}_i) \cdot \boldsymbol{x}_i \tag{28}$$

This is the same as minimizing the expression in question 3. ■

5 Problem 5

$$\text{let } Q_j = \begin{bmatrix} q_{j,1} \\ q_{j,2} \\ \vdots \\ q_{j,N} \end{bmatrix} \quad \text{let } \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_N \end{bmatrix} \quad \text{let } \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad \text{let } \boldsymbol{w_j} \text{ be the } jth \text{ weight column vector.} \tag{29}$$

lets first vectorize

$$\Sigma_{i=1}^{N}(y_i - \boldsymbol{w}_j^T \boldsymbol{x_i})^2$$

this can be vectorized as

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j)^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j)$$

we can attach a q_{ji} coefficient to each row by element wise multiplying Q_j with one of the factors from the expression above.

Therefor

$$\Sigma_{i=1}^{N} q_{j,i}(y_i - \boldsymbol{w}_j^T \boldsymbol{x_i})^2$$

we can vectorized as

$$(Q_j \circ (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j))^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j)$$

Note that $\circ$ is the element wise multiplication operator

Lets expand this expression.

$$\begin{aligned} E &= (Q_j \circ (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j))^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j) && (30) \\ &= (Q_j \circ \boldsymbol{y} - Q \circ \boldsymbol{X}\boldsymbol{w}_j)^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j) && (31) \\ &= (Q_j^T \circ \boldsymbol{y^T} - Q^T \circ \boldsymbol{w}_j^T \boldsymbol{X}^T)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_j) && (32) \\ &= Q_j^T \circ \boldsymbol{y^T}\boldsymbol{y} - Q^T \circ \boldsymbol{y^T}\boldsymbol{X}\boldsymbol{w}_j - Q^T \circ \boldsymbol{w}_j^T \boldsymbol{X}^T \boldsymbol{y} + Q^T \circ \boldsymbol{w}_j^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w}_j && (33) \end{aligned}$$

We need to convert this form into a friendlier one where we know the derivative,

$$\text{let } Q'_j = \begin{bmatrix} Q_j^T \\ Q_j^T \\ \vdots \\ Q_j^T \end{bmatrix}$$

$$\begin{aligned} E &= Q^T \circ \boldsymbol{y^T}\boldsymbol{y} - Q_j^T \circ \boldsymbol{y^T}\boldsymbol{X}\boldsymbol{w}_j - Q_j^T \circ \boldsymbol{w}_j^T \boldsymbol{X}^T \boldsymbol{y} + Q_j^T \circ \boldsymbol{w}_j^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w}_j && (34) \\ &= Q_j^T \circ \boldsymbol{y^T}\boldsymbol{y} - Q_j^T \circ \boldsymbol{y^T}\boldsymbol{X}\boldsymbol{w}_j - \boldsymbol{w}_j^T \boldsymbol{X}^T (Q_j \circ \boldsymbol{y}) + Q_j^T \circ \boldsymbol{w}_j^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w}_j && (35) \\ &= Q_j^T \circ \boldsymbol{y^T}\boldsymbol{y} - Q_j^T \circ \boldsymbol{y^T}\boldsymbol{X}\boldsymbol{w}_j - \boldsymbol{w}_j^T \boldsymbol{X}^T (Q_j \circ \boldsymbol{y}) + \boldsymbol{w}_j^T (Q'^T_j \circ \boldsymbol{X}^T)\boldsymbol{X}\boldsymbol{w}_j && (36) \end{aligned}$$

Now lets differentiate and set to 0

$$\begin{aligned} \frac{\partial E}{\partial \boldsymbol{w}_j} &= 0 - \boldsymbol{X^T}(Q_j \circ \boldsymbol{y}) - \boldsymbol{X^T}(Q_j \circ \boldsymbol{y}) + ((Q'^T_j \circ \boldsymbol{X^T})\boldsymbol{X} + \boldsymbol{X}^T(Q'_j \circ \boldsymbol{X}))\boldsymbol{w_j} && (37) \\ &= -2\boldsymbol{X^T}(Q_j \circ \boldsymbol{y}) + 2(Q'^T_j \circ \boldsymbol{X}^T)\boldsymbol{X}\boldsymbol{w}_j && (38) \end{aligned}$$

Setting this to 0, we get

$$\boldsymbol{w_j}^* = (Q'^T_j \circ \boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X^T}(Q_j \circ \boldsymbol{y})$$