Solution approach

(2)

$$\frac{\partial \hat{x}_i}{\partial \mu} = \frac{\partial}{\partial \mu} \frac{x_i - \mu}{\sqrt{v + \epsilon}} = \boxed{-\frac{1}{\sqrt{v+\epsilon}}} \longrightarrow$$

$$(x_i - \mu) \cdot (v + \epsilon)^{-\frac{1}{2}}$$

$D, 1$

$$\begin{bmatrix} v_1 + \epsilon^{-\frac{1}{2}} \\ v_2 + \epsilon^{-0.5} \\ v_3 + \epsilon^{-0.5} \end{bmatrix}$$

(4)

$$\frac{\partial v}{\partial \mu} = \frac{\partial}{\partial \mu} \frac{1}{M} \sum_{i=1}^{M} (x_i - \mu)^2 = \boxed{\frac{1}{M} \sum_{i=1}^{M} 2(x_i - \mu)(-1)}$$

We are missing $\frac{\partial J}{\partial v}$: $\boxed{\frac{\partial J}{\partial v}} = \frac{\partial J}{\partial \hat{x}} \cdot \frac{\partial \hat{x}}{\partial v}$ (3B) (3A)

(D, D) (3A)

$$\frac{\partial \hat{x}_i}{\partial v} = \frac{\partial}{\partial v} \sum_{i=1}^{M} \overbrace{\frac{x_i - \mu}{\sigma}}^{\hat{x}} = \frac{\partial}{\partial v} \sum_{i=1}^{M} (x_i - \mu) \cdot (v + \epsilon)^{-\frac{1}{2}}$$

$$= \sum_{i=1}^{M} (x_i - \mu) \cdot \left(-\frac{1}{2}\right) (v + \epsilon)^{-\frac{3}{2}} = \boxed{-\frac{1}{2} \sum_{i=1}^{M} (x_i - \mu)(v + \epsilon)^{-\frac{3}{2}}}$$ (3A)

$$\frac{\partial J}{\partial \mu} = \frac{\partial J}{\partial \hat{x}} \cdot \frac{\partial \hat{x}}{\partial \mu} + \frac{\partial J}{\partial v} \cdot \frac{\partial v}{\partial \mu}$$

(1) (2) (3) (4)

$$\cancel{\sum_{i=1}^{M} \frac{\partial J}{\partial \hat{x}_i}} = \left(\sum_{i=1}^{M} \frac{\partial J}{\partial \hat{x}_i} \circ \frac{-1}{\sqrt{v+\epsilon}} + \frac{\partial J}{\partial v} \sum_{i=1}^{M} \frac{\partial J}{\partial \hat{x}_i} \cdot \left(-\frac{1}{2}\right)(x_i - \mu)(v + \epsilon)^{-\frac{3}{2}} \right)$$

(D, 1) (1) (2) (D, 1) (D, 1) (3B)

$$= \sum_{i=1}^{M} \frac{\partial J}{\partial \hat{x}_i} \circ \frac{-1}{\sqrt{v+\epsilon}} + \frac{\partial J}{\partial v} \circ \frac{1}{M} \sum_{i=1}^{M} -2(x_i - \mu)$$

$$= \sum_{i=1}^{M} \frac{\partial J}{\partial \hat{x}_i} \circ \frac{-1}{\sqrt{v+\epsilon}} + \frac{\partial J}{\partial v} \circ (-2) \cdot \frac{1}{M} \cdot \sum_{i=1}^{M} x_i - \frac{1}{M} \sum_{i=1}^{M} \mu$$

$$= \sum_{i=1}^{M} \frac{\partial J}{\partial \hat{x}_i} \circ \frac{-1}{\sqrt{v+\epsilon}} + \frac{\partial J}{\partial v} \circ (-2) \cdot \overbrace{\mu - \frac{M \cdot \mu}{M}}^{0}$$

$$= \boxed{\sum_{i=1}^{M} \frac{\partial J}{\partial \hat{x}_i} \circ \frac{-1}{\sqrt{v+\epsilon}}}$$

21/10/2018

$f(y) \qquad y(\hat{x}, \gamma, \beta) \qquad \hat{x}(\mu, \sigma^2, x)$

$(1, D)$

$$\boxed{\frac{\partial f}{\partial y} = dout}$$ // upstream gradient

$$\boxed{\frac{\partial f}{\partial \gamma} = \sum_{i=1}^{M} \overset{1,D}{\frac{\partial f}{\partial y_i}} \circ \overset{1,D}{\hat{x}_i}}$$

1

$$\boxed{\frac{\partial f}{\hat{x}_i} = \overset{1,D}{\frac{\partial f}{\partial y_i}} \circ \overset{1,D}{\gamma}}$$

$$\boxed{\frac{\partial f}{\partial \beta} = \sum_{i=1}^{M} \overset{1,D}{\frac{\partial f}{\partial y_i}}}$$

$R^D \to R^D$

$$\mu = \frac{1}{M} \sum_{i=1}^{M} x_i \qquad v = \frac{1}{M} \sum_{i=1}^{M} (x_i - \mu)^2 \qquad \hat{x}_i = \frac{x_i - \mu}{\sigma}$$

Jacobian (D,D) Vector

$$\frac{\partial f}{\mu} = \overset{(1,D)^T}{\frac{\partial f}{\partial \hat{x}_i}} \cdot \frac{\partial \hat{x}_i}{\partial \mu} + \overset{(1,D)}{\frac{\partial f}{\partial v}} \cdot \overset{(D,D)}{\frac{\partial v}{\partial \mu}}$$

3

→ This is how it is done in the tutorial. Let's try it with the quotient rule first.

Lokal

— ✱ —

Quotient rule attempt

$$\hat{x}_i = \frac{x_i - \cancel{\frac{\sum_{i=1}^{M} x_i}{M}}}{\sqrt{\frac{\sum_{i=1}^{M}\left(x_i - \cancel{\frac{\sum_{i=1}^{M} x_i}{M}}\right)^2}{M}}} \begin{array}{l} \to f(x) \\ \\ \to g(x) \end{array}$$

Quotient rule: $\frac{\partial}{\partial x} \frac{f(x)}{g(x)} = \frac{f'(x) \cdot g(x) - f(x) g'(x)}{g(x)^2}$

This doesn't work?

$f'(\mu) = -1$

$$g'(\mu) = \frac{\partial}{\partial \mu} \sqrt{\frac{1}{M} \sum_{i=1}^{M} (x_i - \mu)^2} = \frac{1}{2} \cdot \frac{1}{\sigma} \cdot \frac{1}{M} \cdot \sum 2 (x_i - \mu) \cdot (-1)$$

$$= \frac{\mu - x_i}{\sigma \cdot M} \Rightarrow \frac{\partial}{\partial \cancel{x}} \frac{x - \mu}{\sigma} = \frac{(-1) \cdot \sigma - (x_i - \mu) \cdot \frac{\mu - x_i}{\sigma \cdot M}}{\sigma^2}$$

$$= \frac{-\sigma - \overbrace{\cancel{}\, x_i \mu - x_i^2 - \mu^2 + \mu x_i}^{-(-x_i^2 + 2\mu x_i - \mu^2)}}{\sigma^2}$$

— ✱ —

$$\frac{\partial J}{\partial x_i} = \frac{\partial J}{\partial \hat{x}_i} \boxed{\frac{\partial \hat{x}_i}{\partial x_i}} + \frac{\partial J}{\partial \mu} \boxed{\frac{\partial \mu}{x_i}} + \frac{\partial J}{\partial v} \cdot \boxed{\frac{\partial v}{\partial x_i}}$$

(1) (D, D)

$$\boxed{\frac{\partial \hat{x}_{ij}}{\partial x_{ik}}} = \begin{cases} \frac{1}{\sigma} & i = j \\ 0 & i \neq j \end{cases}$$

(D, D) (2)

$$\boxed{\frac{\partial \mu_k}{x_{il}}} = \begin{cases} \frac{1}{M} & l = k \\ 0 & l \neq k \end{cases}$$

$$\cancel{\frac{\partial v_k}{\partial x_{il}} = \{}$$

(3)

$$\boxed{\frac{\partial v_k}{\partial x_{il}}} = \begin{cases} \frac{1}{M} \cdot 2(x_i - \mu) \cdot \underbrace{(-1) \cdot \frac{1}{M}}_{\text{Apparently this is wrong (?)}} & k = l \\ 0 & k \neq l \end{cases}$$

$$\frac{\partial J}{\partial x_i} = \left(\frac{\partial J}{\partial \hat{x}_i} \cdot \frac{1}{\sigma}\right) + \left(\frac{\partial J}{\partial \mu} \cdot \frac{1}{M}\right) + \left(\frac{\partial J}{\partial v} \cdot \frac{2(x_i - \mu)}{M}\right)$$

$$= \left(\frac{\partial J}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{v+\epsilon}}\right) + \left(\frac{1}{M} \cdot \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} \cdot \frac{-1}{\sqrt{v+\epsilon}}\right) - \left(\frac{1}{2} \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} \cdot (x_j - \mu)(v+\epsilon)^{-\frac{3}{2}} \cdot \frac{2(x_i - \mu)}{m}\right)$$

$$= \left(\frac{\partial J}{\partial \hat{x}_i} \cdot (v+\epsilon)^{-\frac{1}{2}}\right) + \left(-\frac{(v+\epsilon)^{-\frac{1}{2}}}{M} \cdot \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j}\right) - \left(\frac{(v+\epsilon)^{-\frac{1}{2}}}{M} \cdot \frac{x_i - \mu}{\sqrt{v+e}} \cdot \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} \cdot \frac{x_j - \mu}{\sqrt{v+\epsilon}}\right)$$

$\sigma^2 = v$

Fancy trick: $(\sigma^2 + \epsilon)^{-\frac{3}{2}} = (\sigma^2 + \epsilon)^{-\frac{1}{2}} \cdot (\sigma^2 + \epsilon)^{-1} = (\sigma^2 + \epsilon)^{-\frac{1}{2}} \cdot \frac{1}{\sqrt{v+\epsilon}} \cdot \frac{1}{\sqrt{v+\epsilon}}$

$$= \frac{\partial J}{\partial \hat{x}_i} (v+\epsilon)^{-\frac{1}{2}} - \frac{(v+\epsilon)^{-\frac{1}{2}}}{M} \cdot \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} - \frac{(v+\epsilon)^{-\frac{1}{2}}}{M} \cdot \frac{x_i - \mu}{\sqrt{v+\epsilon}} \cdot \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} \cdot \frac{x_j - \mu}{\sqrt{v+\epsilon}}$$

$$= \left(\frac{\partial J}{\partial \hat{x}_i} \cdot (v+\epsilon)^{-\frac{1}{2}}\right) - \left(\frac{(v+\epsilon)^{-\frac{1}{2}}}{M} \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j}\right) - \left(\frac{(v+\epsilon)^{-\frac{1}{2}}}{M} \cdot \hat{x}_i \cdot \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} \cdot \frac{x_j - \mu}{\sqrt{v+\epsilon}}\right)$$

$$= \frac{(v+\epsilon)^{-\frac{1}{2}}}{M} \left(M \cdot \frac{\partial J}{\partial \hat{x}_i} - \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} - \hat{x}_i \cdot \sum_{j=1}^{M} \frac{\partial J}{\partial \hat{x}_j} \cdot \underbrace{\frac{x_j - \mu}{\sqrt{v+\epsilon}}}_{\hat{x}_j}\right)$$

21/10/18

Recap :

$$\boxed{\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^{M} \frac{\partial \ell}{\partial y_i} \\
\frac{\partial \ell}{\partial \gamma} &= \sum_{i=1}^{M} \frac{\partial \ell}{\partial y_i} \circ \hat{x}_i \\
\frac{\partial \ell}{\partial x_i} &= \frac{1}{M \cdot \sqrt{v+\epsilon}} \cdot \left(M \frac{\partial \ell}{\partial \hat{x}_i} - \sum_{j=1}^{M} \frac{\partial \ell}{\partial \hat{x}_j} - \hat{x}_i \sum_{j=1}^{M} \frac{\partial \ell}{\partial \hat{x}_j} \cdot \hat{x}_j \right)
\end{aligned}}$$