Ex. 1 - DL basics

1. a. The shape of X: $m\times 10$ where m is the batch size.
   b. The shape of W_h: 10×50, b_h is 1×50.
   c. The shape of W_o: 50×3, b_o is 1×3.
   d. The shape of output Y: $m\times 3$, m is the batch size.
   e. $Y = ReLU(\underbrace{ReLU(X\cdot W_h + b_h)}_{\text{output of hidden layer}}\cdot W_o + b_o)$

2. Given 3 convolutional layers with 3×3 kernels we will define the following for each step:

   n_i^O - number of outputs of step i.

   n_i^I - number of input channels to step i.

   The total number of parameters, N_T, will be then

$$N_T = \sum_{i=1}^{3} n_i^O \cdot (\underbrace{3\times 3}_{\text{kernel size}}\times n_i^I + 1) =$$

$$= n_1^O\cdot(3\times 3\times n_1^I + 1) + n_2^O\cdot(3\times 3\times n_2^I + 1) + n_3^O\cdot(3\times 3\times n_3^I + 1)$$

$$= 100\cdot(3\cdot 3\cdot \underset{\text{RGB}}{3} + 1) + 200\cdot(3\times 3\cdot \underset{\text{output of previous layer}}{100} + 1) + 400\cdot(3\cdot 3\cdot 200 + 1) =$$

$$\boxed{N_T = 2{,}800 + 180{,}200 + 720{,}400 = 903{,}400}$$

3. a. $\frac{\partial f}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial f}{\partial y_i} \cdot \frac{\partial y_i}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial f}{\partial y_i} \cdot \hat{x}_i$

b. $\frac{\partial f}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial f}{\partial y_i} \cdot \underbrace{\frac{\partial y_i}{\partial \beta}}_{= 1} = \sum_{i=1}^{m} \frac{\partial f}{\partial y_i}$

c. $\frac{\partial f}{\partial \hat{x}_i} = \frac{\partial f}{\partial y_i} \cdot \frac{\partial y_i}{\partial x_i} = \frac{\partial f}{\partial y_i} \cdot \gamma$

d. $\frac{\partial f}{\partial \sigma^2} = \sum_{i=1}^{m} \frac{\partial f}{\partial \hat{x}_i} \cdot \frac{\partial \hat{x}_i}{\partial \sigma^2} = -\frac{1}{2} \sum_{i=1}^{m} \frac{\partial f}{\partial \hat{x}_i} \cdot \frac{x_i - \mu}{(\sigma^2 + \varepsilon)^{3/2}} =$

$$= -\frac{1}{2} \sum_{i=1}^{m} \underbrace{\frac{\partial f}{\partial y_i} \cdot \gamma}_{\frac{\partial f}{\partial \hat{x}_i}} \cdot \frac{x_i - \mu}{(\sigma^2 + \varepsilon)^{3/2}}$$

e. $\frac{\partial f}{\partial \mu} = \left(\sum_{i=1}^{m} \frac{\partial f}{\partial \hat{x}_i} \cdot \frac{\partial \hat{x}_i}{\partial \mu} \right) + \frac{\partial f}{\partial \sigma^2} \cdot \frac{\partial \sigma^2}{\partial \mu} = -\frac{\partial f}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma^2 + \varepsilon}} + \frac{\partial f}{\partial \sigma^2} \cdot \frac{\partial \sigma^2}{\partial \mu}$

חושב בסעיף c

חושב בסעיף d

<u>חישוב עזר:</u>

$$\frac{\partial \sigma^2}{\partial \mu} = \frac{2}{m} \sum_{i=1}^{m} (x_i - \mu) = \frac{2}{m} \cdot \left(\left(\sum_{i=1}^{m} x_i \right) - m \cdot \mu \right) = 2 \cdot (\mu - \mu) = \underline{\underline{0}}$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial \mu} = \sum_{i=1}^{m} \frac{\partial f}{\partial \hat{x}_i} \cdot \frac{\partial \hat{x}_i}{\partial \mu} = \sum_{i=1}^{m} \frac{\partial f}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma^2 + \varepsilon}}}$$

סעיף c　　סעיף e　　סעיף d

f. $\frac{\partial f}{\partial x_i} = \left(\frac{\partial f}{\partial \hat{x}_i} \cdot \frac{\partial \hat{x}_i}{\partial x_i} \right) + \left(\frac{\partial f}{\partial \mu} \cdot \frac{\partial \mu}{\partial x_i} \right) + \left(\frac{\partial f}{\partial \sigma^2} \cdot \underbrace{\frac{\partial \sigma^2}{\partial x_i}}_{= \frac{2(x_i - \mu)}{m}} \right) =$

סעיף e

$$= \frac{\partial f}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma^2 + \varepsilon}} + \left(\sum_{k=1}^{m} \frac{\partial f}{\partial \hat{x}_k} \cdot \frac{-1}{\sqrt{\sigma^2 + \varepsilon}} \right) \cdot \frac{1}{m} + \left(-\frac{1}{2} \sum_{n=1}^{m} \frac{\partial f}{\partial \hat{x}_n} \cdot \frac{(x_n - \mu)}{(\sigma^2 + \varepsilon)^{3/2}} \cdot \frac{2(x_i - \mu)}{m} \right)$$

$$= \frac{\partial f}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma^2 + \varepsilon}} - \frac{1}{m\sqrt{\sigma^2 + \varepsilon}} \sum_{k=1}^{m} \frac{\partial f}{\partial \hat{x}_k} - \frac{x_i - \mu}{m\sqrt{\sigma^2 + \varepsilon}} \sum_{n=1}^{m} \frac{\partial f}{\partial \hat{x}_n} \cdot \frac{(x_n - \mu)}{(\sigma^2 + \varepsilon)}$$

Scanned with CamScanner

$$\Rightarrow \frac{\partial f}{\partial x_i} = \frac{1}{m\sqrt{\sigma^2+\varepsilon}}\left(m\cdot\frac{\partial f}{\partial \hat{x}_i} - \sum_{k=1}^{m}\frac{\partial f}{\partial \hat{x}_k} - \underbrace{\frac{x_i-\mu}{\sqrt{\sigma^2+\varepsilon}}}_{=\hat{x}_i}\sum_{n=1}^{m}\frac{\partial f}{\partial \hat{x}_n}\cdot \underbrace{\frac{x_n-\mu}{\sqrt{\sigma^2+\varepsilon}}}_{=\hat{x}_n}\right)$$

$$\Rightarrow \frac{\partial f}{\partial x_i} = \frac{1}{m\sqrt{\sigma^2+\varepsilon}}\left(m\cdot\frac{\partial f}{\partial \hat{x}_i} - \sum_{k=1}^{m}\frac{\partial f}{\partial \hat{x}_k} - \hat{x}_i\sum_{n=1}^{m}\frac{\partial f}{\partial \hat{x}_n}\cdot \hat{x}_n\right)$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial x_i} = \frac{1}{m\sqrt{\sigma^2+\varepsilon}}\left(m\cdot\gamma\cdot\frac{\partial f}{\partial y_i} - \sum_{k=1}^{m}\gamma\cdot\frac{\partial f}{\partial y_k} - \frac{\partial f}{\partial y_i}\cdot\gamma\cdot\sum_{n=1}^{m}\gamma\cdot\frac{\partial f}{\partial y_n}\cdot \hat{x}_n\right)}$$

נציב:

$$\frac{\partial f}{\partial \hat{x}_i} = \frac{\partial f}{\partial y_i}\cdot\gamma$$