Homework 6 Solution

April 14, 2018

Problem 1 (20 points)

Forward and backward propagation for a simple GAN.

Forward Propagation (10 points)

4 points for providing the correct formula and **6 points** for numerical calculations, **2 points** for each of h_0, h_1, and h_2.

$$z = (0.4, 0.8)^T$$

$$h_0 = \mathrm{ReLu}(w_0^t z + b_0) = \begin{bmatrix} 0.38 \\ 1.40 \\ 0.96 \\ 1.06 \\ 0.70 \\ 1.54 \\ 1.62 \\ 1.06 \end{bmatrix}$$

The values on h_1 layer are

$$h_1 = \mathrm{reshape}(h_0) = [h_1^1, h_1^2]$$

$$h_1^1 = \begin{bmatrix} 0.38 & 0.96 \\ 1.40 & 1.06 \end{bmatrix}$$

$$h_1^2 = \begin{bmatrix} 0.70 & 1.62 \\ 1.54 & 1.06 \end{bmatrix}$$

The values on convolution layer h_2 are

$$h_2 = \mathrm{ReLu}(\mathrm{Deconv}(h_1, w_1, b_1)) = \begin{bmatrix} 0.324 & 1.066 & 1.134 & 1.002 \\ 1.060 & 2.820 & 4.064 & 1.324 \\ 1.058 & 4.358 & 3.022 & 0.844 \\ 1.624 & 1.880 & 0.978 & 0.318 \end{bmatrix}$$

$$+\begin{bmatrix} 0.1 & 0.5 & 0.6 & 0.3 \\ 0.2 & 0.4 & 0.8 & 0.1 \\ 0.1 & 0.3 & 0.7 & 0.2 \\ 0.3 & 0.1 & 0.2 & 0.1 \end{bmatrix} = \begin{bmatrix} 0.424 & 1.566 & 1.734 & 1.302 \\ 1.260 & 3.220 & 4.864 & 1.424 \\ 1.158 & 4.658 & 3.722 & 1.044 \\ 1.924 & 1.980 & 1.178 & 0.418 \end{bmatrix}$$

After reshape, we have

$H_2 = [0.424, 1.260, 1.158, 1.924, 1.566, 3.220, 4.658, 1.980, 1.734, 4.864, 3.722, 1.078, 1.302, 1.424, 1.044, 0.418]^T$

The final value at the binary discriminator is

$$p(y=0) = \sigma(w_2^t H_2 + b_2) = 0.999994394742746$$

Backward Propagation (10 points)

5 points for providing the correct formula and **5 points** for numerical calculations. In each case **1 point** for each of ∇y, ∇h_2, ∇w_1, $\nabla h1$ and ∇w_0.

Set $L = \log p(y=0)$, the gradient of y is

$\nabla H_2 = \dfrac{p(y=0)\,[1-p(y=0)]\,w_2}{p(y=0)} = [0.0544, 0.1088, 0.1632, 0.1088, 0.2176, 0.3264,$

$0.3808, 0.0544, 0.0544, 0.1088, 0.2720, 0.1632, 0.0544, 0.1088, 0.3264, 0.4896]^t \times 10^{-5}$

$$\nabla h_2 = \text{reshape}(\nabla H_2) = \begin{bmatrix} 0.0544 & 0.2176 & 0.0544 & 0.0544 \\ 0.1088 & 0.3264 & 0.1088 & 0.1088 \\ 0.1632 & 0.3808 & 0.2720 & 0.3264 \\ 0.1088 & 0.0544 & 0.1632 & 0.4896 \end{bmatrix} \times 10^{-5}$$

$$\nabla w_1 = \frac{\partial h_2}{\partial w_1} \nabla h_2$$

As all elements in h_2 are larger than 0, we have

$$\nabla w_1[i][j][k] = \sum_{r=1}^{4} \sum_{c=1}^{4} h_1[r+i-1][c+j-1][k] \nabla h_2[r][c]$$

The gradient of w_1 is list as

$$\nabla w_1 = [\nabla w_1^1, \nabla w_1^2]$$

$$\nabla w_1^1 = \begin{bmatrix} 0.1164 & 0.0655 & 0.0638 \\ 0.0873 & 0.1050 & 0.0987 \\ 0.0341 & 0.0707 & 0.0728 \end{bmatrix} \times 10^{-4}$$

$$\nabla w_1^2 = \begin{bmatrix} 0.1489 & 0.0964 & 0.0956 \\ 0.1017 & 0.1279 & 0.1260 \\ 0.0409 & 0.0858 & 0.0904 \end{bmatrix} \times 10^{-4}$$

the gradients of h_1 and h_0 are respectively

$$\nabla h_1[r+i-1][c+j-1][l] = \sum w_1[i][j][l] \nabla h_2[r][c]$$

$$\nabla h_1^1 = \begin{bmatrix} 0.6256 & 0.5494 \\ 0.5875 & 0.5657 \end{bmatrix} \times 10^{-5}$$

$$\nabla h_1^2 = \begin{bmatrix} 0.6256 & 0.6201 \\ 0.7126 & 0.7398 \end{bmatrix} \times 10^{-5}$$

$$\nabla h_0 = [0.6256, 0.5875, 0.5494, 0.5657, 0.6256, 0.7126, 0.6201, 0.7398]^t \times 10^{-5}$$

then the final gradient of w_0 is

$$\nabla w_0 = \nabla h_0 z_1^t = \begin{bmatrix} 0.2502 & 0.5004 \\ 0.2350 & 0.4700 \\ 0.2198 & 0.4395 \\ 0.2263 & 0.4526 \\ 0.2502 & 0.5004 \\ 0.2850 & 0.5701 \\ 0.2480 & 0.4961 \\ 0.2959 & 0.5918 \end{bmatrix} \times 10^{-5}$$

Problem 2 (20 points)

5 points for each of ∇x, ∇h, ∇w_1, and ∇w_0. If you make minor mistakes in any of them you will earn **3 points**.

For encoding and decoding processes of the auto-encoder are listed as below:

$$h = \text{ReLu}(w_1^t x + b_1)$$

$$x' = \sigma(w_2^t h + b_2)$$

We want to final the weight w_1 and w_2 such that

$$w_1, w_2 = \arg\min_{w_1, w_2} \sum_{i=1}^{N} (x_i - x_i')^2$$

by backward propagation

$$\nabla x_j' = \frac{\partial \sum_{i=1}^{N} (x_i - x_i')^2}{\partial x_j'} = 2(x_j' - x_j)$$

$$\nabla h_j = \frac{\partial x_j'}{\partial h_j} \nabla x_j' = 2(x_j' - x_j) x_j' (1 - x_j') w_2$$

$$\nabla w_2 = \sum_{i=1}^{N} \frac{\partial x_i'}{\partial h_i} \nabla x_i' = \sum_{i=1}^{N} 2(x_i' - x_i) x_i' (1 - x_i') h_i$$

$$\nabla w_1 = 2 \sum_{i=1}^{N} \frac{\partial h_i}{\partial w_1} (x_i' - x_i) x_i' (1 - x_i') w_2$$

where

$$\frac{\partial h_i}{\partial w_1} = \begin{cases} x_i & \text{if } w_1^t x_i + b_1 > 0 \\ 0 & \text{else} \end{cases}$$

Problem 3 (20 points)

For maximizing $p(x|\phi)$ we can maximize its lower bound, which can be obtained as follows:

$$
\begin{aligned}
\log p(x|\phi) &= \log \mathbb{E}_{p(z|\phi)}\{p(x|z,\phi)\} &(1)\\
&= \log \mathbb{E}_{p(z|\phi)}\left\{\frac{q(z|x,\theta)}{q(z|x,\theta)}p(x|z,\phi)\right\} &(2)\\
&= \log \mathbb{E}_{q(z|x,\theta)}\left\{\frac{p(z|\phi)}{q(z|x,\theta)}p(x|z,\phi)\right\} &(3)\\
&\geq \mathbb{E}_{q(z|x,\theta)}\left\{\log\left(\frac{p(z|\phi)}{q(z|x,\theta)}p(x|z,\phi)\right)\right\} &(4)\\
&= \mathbb{E}_{q(z|x,\theta)}\left\{\log\left(\frac{p(z|\phi)}{q(z|x,\theta)}\right)\right\} + \mathbb{E}_{q(z|x,\theta)}\{\log(p(x|z,\phi))\} &(5)\\
&= \mathbb{E}_{q(z|x,\theta)}\{\log p(x|z,\phi)\} - KL(q(z|x,\theta)\|p(z|\phi)) &(6)\\
&= \mathbb{E}_{q(z|x,\theta)}\{\log p(x|z,\phi)\} - KL(q(z|x,\theta)\|p(z)) &(7)
\end{aligned}
$$

where the inequality term in (4) is due to the Jensen's inequality, and the last equality is due to the fact that random variable z is independent of the decoder parameters ϕ.

1. Expansion of $p(x|\phi)$ as an expectation: **5 points**
2. Using the Jensen's inequality on the results: **5 points**
3. Changing the expectation measure: **5 points**
4. Using the fact that z is independent of ϕ: **5 points**