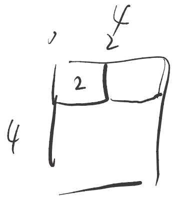
Optimization and Machine Learning, Fall 2023
Homework 5

(Due Thursday, Jan 11 at 11:59pm (CST))



1. [10 points] [Deep Learning Model]

   (a) Consider a 2D convolution layer. Suppose the input size is $4 \times 64 \times 64 \times$ (channel, width, height) and we use **ten** 3×3 (width, height) kernels with 4 channels input and 4 channels output to convolve with it. Set stride = 1 and pad = 1. What is the output size? Let the bias for each kernel be a scalar, how many parameters do we have in this layer? [5 points]

   (b) The convolution layer is followed by a max pooling layer with 2×2 (width, height) filter and stride = 2. What is the output size of the pooling layer? How many parameters do we have in the pooling layer? [5 points]

(a)

① first we know that we will have $4 \times 10 = 40$ output map

the shape should be $\frac{64-3+2}{1} + 1 = 64$

So we have size is $4 \times 64 \times 64$

parameter: $(3 \times 3 \times 4 + 1) \times 10 \times 4 = 1480$

(b)

$32 \times 32 \times 10$

0

2. [10 points] Use the k-means++ algorithm and Euclidean distance to cluster the 8 data points into $K = 3$ clusters. The coordinates of the data points are:

$$x^{(1)} = (2,8),\ x^{(2)} = (2,5),\ x^{(3)} = (1,2),\ x^{(4)} = (5,8),$$
$$x^{(5)} = (7,3),\ x^{(6)} = (6,4),\ x^{(7)} = (8,4),\ x^{(8)} = (4,7).$$

Suppose that initially the first cluster centers is $x^{(1)}$.

(a) Perform the k-means++ algorithm to initialize other centers and report the coordinates of the resulting centroids. [3 points]

(b) Calculate the loss function

$$Q(r,c) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K} r_{ij}||x^{(i)} - c_j||^2, \tag{1}$$

where $r_{ij} = 1$ if $x^{(i)}$ belongs to the j-th cluster and 0 otherwise. [2 points]

(c) How many more iterations are needed to converge? [3 points] Calculate the loss after it converged. [2 points]

(a) the first center is $X^{(1)}$. so we calculate the distance $D^2(x)$

0, 9, 37, 9, 50, 32, 52, 5

$\sum_i^8 D^2(X_i) = 194$

$P_r(8) = [0.973, 1)$ $P_r(7) = [0.705, 0.973)$ $P_r(6) = [0.540, 0.705)$

So we choose point 6.

we now calculate the $D^2(x)$ with $X_{(6)}$

32, 17, 29, 17, 2, 0, 4, 13

So now we have 0, 9, 29, 9, 2, 0, 4, 5

$\sum_i^8 D^2(X_i) = 58$

$P_r(2) = [0, 0.16)$ $P_r(3) = [0.16, 0.66)$

we choose point 3.

So we have $X^{(1)}(2,8)$ $X^{(3)}(1,2)$ $X^{(6)}(6,4)$

(b)

$x^{(1)} = (2,8)$, $x^{(2)} = (2,5)$, $x^{(3)} = (1,2)$, $x^{(4)} = (5,8)$,
$x^{(5)} = (7,3)$, $x^{(6)} = (6,4)$, $x^{(7)} = (8,4)$, $x^{(8)} = (4,7)$.

$x^{(1)}$, $x^{(2)}$, $x^{(4)}$, $x^{(8)}$

$x^{(3)}$.

$x^{(6)}$, $x^{(5)}$, $x^{(7)}$

$$Q(r,c) = \frac{1}{8} \cdot (9+9+5+2+4) = \frac{29}{8}$$

1

(c)

new center $\left(\frac{13}{4}, 7\right)$ $(1,2)$ $\left(7, \frac{11}{3}\right)$

$x^{(1)}$, $x^{(2)}$, $x^{(4)}$, $x^{(8)}$

$x^{(3)}$

$x^{(5)}$, $x^{(6)}$, $x^{(7)}$

no point will move to other class

$$Q(r,c) = \frac{1}{8}\left(\frac{25}{16} + 1 + \frac{25}{16} + 4 + \frac{49}{16} + 1 + \frac{9}{16}\right.$$

$$\left. + 0 + \frac{4}{9} + 1 + \frac{1}{9} + 1 + \frac{1}{9}\right)$$

$$= \frac{1}{8}\left(6 + \frac{108}{16} + 2 + \frac{6}{9}\right) = \frac{1}{8}\left(8 + \frac{27}{4} + \frac{2}{3}\right)$$

$$= 1 + \frac{89}{96}$$

the ieeration step

3. [10 points] Name 2 deep generation networks. [2 points] Briefly describe the training procedure of a GAN model.(What's the objective function? How to update the parameters in each stage?) [8 points]

(1) VAE, GAN

(2)

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))]$$

for number of training iterations do:

for k steps do:

sample minibatch of m noise samples $\{z^{(1)} \ldots z^{(m)}\}$ from noise prior $p_g(z)$

sample minibatch of m examples $\{x^{(1)} \ldots x^{(m)}\}$ from data generating distribution

update the discriminator by ascending stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} [\log D(x^{(i)}) + \log(1-D(G(z^{(i)})))]$$

end for

sample minibatch of m noise samples $\{z^{(1)} \ldots z^{(m)}\}$ from noise prior $p_g(z)$

update the generator by descending its stochastic gradient.

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1-D(G(z^{(i)})))$$

end for