Optimization and Machine Learning, Fall 2023

Homework 5

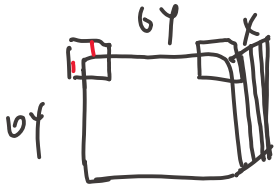
(Due Thursday, Jan 11 at 11:59pm (CST))

1. [10 points] [Deep Learning Model]

   (a) Consider a sequential 2D convolution block consist of 10 layers. Suppose the input size is 4 × 64 × 64 × (channel, width, height) and we use 3 × 3 (width, height) Conv2D 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 the 2 [5 points]

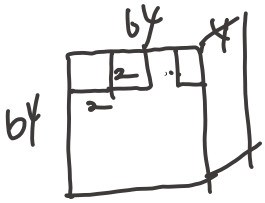
   (b) The convolution layer is followed by a max pooling layer with 2 × 2 (width, height) filter and stride = 2. What is the output size of the pooling layer? How many parameters do we have in the pooling layer? [5 points]

a).



1° width and height: $64-3+1+2=64$.
channel: ∵ ① ∴ 4 channels.
The Output size is, $64\times64\times4$.

2° parameters: $(3\times3\times4+1)\times10=370$



1° 1 2 3 4 5 6 .... 61 64.
width and height: $64\div2=32$.
The Output size is: $32\times32\times4$

2° no parameters.

$+ \quad (9-7)^2 + (3-\frac{14}{3})^2 + (6-7)^2 + (4-\frac{14}{3})^2 + (8-7)^2 + (4-\frac{14}{3})^2 + (1-1)^2 + (2-2)^2$

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 $x^{(1)}$.
To ensure consistent results, please use random numbers in the order shown in the table below. When selecting a center, arrange it in ascending order of sequence number. For example, when the normalized weights of 5 nodes are 0.2, 0.1, 0.3, 0.3, and 0.1, if the random number is 0.3, the selected node is the third one. Note that you don't necessarily need to use all of them.

0.6	0.2	0.5	0.9	0.3

(c) .

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

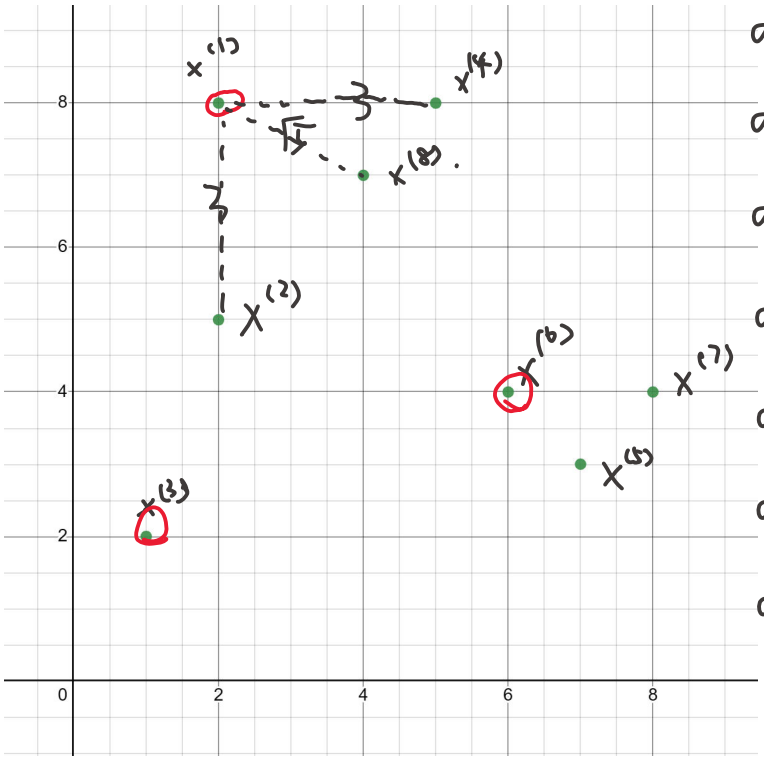
(b) Calculate the loss function

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

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).



$d(x^{(1)}, x^{(1)}) = 0$

$d(x^{(1)}, x^{(2)}) = 3^2 = 9$

$d(x^{(1)}, x^{(3)}) = 1^2 + 6^2 = 37$

$d(x^{(1)}, x^{(4)}) = 3^2 = 9$

$d(x^{(1)}, x^{(5)}) = 5^2 + 5^2 = 50$

$d(x^{(1)}, x^{(6)}) = 4^2 + 4^2 = 32$

$d(x^{(1)}, x^{(7)}) = 6^2 + 4^2 = 52$

$d(x^{(1)}, x^{(8)}) = 2^2 + 1^2 = 5$

$\sum_{i=2}^{8} d(x^{(1)}, x^{(i)}) = 9 + 37 + 9 + 50 + 32 + 52 + 5 = 194$

Normalized: $\frac{0+9+37+9+50}{\sum_{i=2}^{8} d(x^{(1)}, x^{(i)})} \approx 0.5412$; $\frac{0+9+37+9+50+32}{\sum_{i=2}^{8} d(x^{(1)}, x^{(i)})} \approx 0.7062$

$0.6 \in (0.5412, 0.7062)$.

$\therefore$ second center is $x^{(6)}$

$d(x^{(1)}, x^{(1)}) = 0 = 0 < d(x^{(6)}, x^{(1)}) = 4^2+4^2 = 32$

$d(x^{(1)}, x^{(2)}) = 3^2 = 9 < d(x^{(6)}, x^{(2)}) = 4^2+1 = 17$

$d(x^{(1)}, x^{(3)}) = 1^2+6^2 = 37 > d(x^{(6)}, x^{(3)}) = 5^2+2^2 = 29$

$d(x^{(1)}, x^{(4)}) = 3^2 = 9 < d(x^{(6)}, x^{(4)}) = 4^2+1^2 = 17$

$d(x^{(1)}, x^{(5)}) = 5^2+5^2 = 50 > d(x^{(6)}, x^{(5)}) = 1^2+1^2 = 2$

$d(x^{(1)}, x^{(6)}) = 4^2+4^2 = 32 > d(x^{(6)}, x^{(6)}) = 0 = 0$

$d(x^{(1)}, x^{(7)}) = 6^2+4^2 = 52 > d(x^{(6)}, x^{(7)}) = 2^2 = 4$

$d(x^{(1)}, x^{(8)}) = 2^2+1^2 = 5 < d(x^{(6)}, x^{(8)}) = 2^2+3^2 = 13.$

$\sum_{i=1}^{8} \min_{j<i} \|x^i - c_j\|^2 = 0+9+29+9+2+0+4+5 = 58$

Normalize: $\frac{0+9}{58} \approx 0.1552$; $\frac{0+9+29}{58} \approx 0.6552$

$0.2 \in (0.1552, 0.6552)$

$\therefore$ third center is $x^{(3)}$

Three groups and centroids:

Group 1: $x^{(1)}, x^{(4)}, x^{(8)}, x^{(2)} : \left(\frac{2+5+2+4}{4}, \frac{8+5+8+7}{4}\right) = \left(\frac{13}{4}, 7\right)$.

Group 2: $x^{(6)}, x^{(5)}, x^{(7)} : \left(\frac{7+6+8}{3}, \frac{3+4+4}{3}\right) = \left(7, \frac{11}{3}\right)$.

Group 3: $x^{(3)} : (1,2)$.

b) $Q(r,c) = \frac{1}{8}\left[\|x^{(1)} - x^{(1)}\|^2 + \|x^{(2)} - x^{(1)}\|^2 + \|x^{(4)} - x^{(1)}\|^2 + \right.$

$\|x^{(8)} - x^{(1)}\|^2 + \|x^{(5)} - x^{(6)}\|^2 + \|x^{(6)} - x^{(6)}\|^2 +$

$\left. \|x^{(7)} - x^{(6)}\|^2 + \|x^{(3)} - x^{(3)}\|^2 \right]$

$= \frac{1}{8}(0+9+9+5+2+0+4+0) = \frac{29}{8}$.

c) With one more iteration,

it will converge with $c_1 = (\frac{13}{4}, 7)$. $c_2 = (1, \frac{11}{3})$. $c_3 = (1, 2)$.

$Q(r,c) = \frac{1}{8}\left[(2-\frac{13}{4})^2 + (8-7)^2 + (2-\frac{13}{4})^2 + (5-7)^2 + (5-\frac{13}{4})^2 + (8-7)^2 + (4-\frac{13}{4})^2 + (7-7)^2 \right.$

$\left. + (7-7)^2 + (3-\frac{11}{3})^2 + (6-7)^2 + (4-\frac{11}{3})^2 + (8-7)^2 + (4-\frac{11}{3})^2 + (1-1)^2 + (2-2)^2 \right]$

$= \frac{1}{8}\left(\frac{25}{16} + 1 + \frac{25}{16} + 4 + \frac{49}{16} + 1 + \frac{9}{16} + \frac{4}{9} + 1 + \frac{1}{9} + 1 + \frac{1}{9} \right)$

$= \frac{1}{8}\left[\frac{108}{16} + \frac{6}{9} + 8 \right) = \frac{1}{8}\left(\frac{27}{4} + \frac{2}{3} + 8 \right] = \frac{185}{96} \approx 1.9271$

3. [10 points] Name 2 deep generation networks. [2 points] Briefly describe the training procedure of a GAN model.(What's the objective function?①How to update the parameters in each stage?)②[8 points]

1）生成对抗网络（GAN，Generative Adversarial Networks）；变分自编码器（VAE，Variational Autoencoders）。

2）对于 GAN 模型的训练过程：GAN 模型包括两个部分，生成器（Generator）和判别器（Discriminator）。

① objectle F: $\min_{\theta_g} \max_{\theta_d} \left(E_{x \sim p_{data}(x)} \log D_{\theta_d}(x) + E_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \right)$

② Update: Gradient descent on discriminator.

$$\max_{\theta_d} \left[E_{x \sim p_{data}} \log D_{\theta_d}(x) + E_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

then gradient descent on generator

$$\min_{\theta_g} E_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))).$$

Then update the discriminator and the generator.

The process will stop until they converge or up to iterations.