

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 $4 \times 64 \times 64 \times (\text{channel, width, height})$ 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]

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:

$$\begin{aligned}x^{(1)} &= (2, 8), \quad x^{(2)} = (2, 5), \quad x^{(3)} = (1, 2), \quad x^{(4)} = (5, 8), \\x^{(5)} &= (7, 3), \quad x^{(6)} = (6, 4), \quad x^{(7)} = (8, 4), \quad x^{(8)} = (4, 7).\end{aligned}$$

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

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

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]

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]