CSDS 391: Introduction to AI
Spring 2023
PA 2
Due: April 16, 2023, before 11:59 pm,
Total Points: 75

In this assignment, you will implement two approaches to classification. The first is k-means clustering, the second is training a linear classifier (a one-layer neural network). In both cases you will use the iris dataset. It will be convenient to do this assignment in a language with good plotting libraries such as Matlab, python, or julia. Submit your write-up as a pdf file. Put your code and any related files in a zip folder (please use zip and not some other compression format). Submit those two files to Canvas so that your write up is visible within canvas (and not contained within the zipped folder).

Write-up: Your write-up should answer the questions like in the written assignments. When relevant, include the source code in the write-up if it is sufficiently short (say half a page or less), otherwise refer to the file.

Clustering

a) Write a program that implements the k-means clustering algorithm on the iris data set. You should use the objective function and learning rule that you will derive in W4 Q1 (and your implementation will help you uncover any errors, so it is helpful to do this step before you turn in W4.) (15 marks)

b) Plot the value of the objective function as a function of the iteration

$$D = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k} \|x_n - \mu_k\|^2$$

to show that your learning rule and implementation minimizes this expression. (5 marks)

c) Plot the results of the learning process by showing the initial, intermediate, and converged cluster centers overlaid on the data for $k = 2$ and $k = 3$. (5 marks)

d) Devise a method to plot the decision boundaries for this dataset using the optimized parameters. Explain your approach and plot your results. (10 marks)

Neural networks

a) Write a program that calculates the mean-squared error of the iris data for simple one-layer neural network using a sigmoid non-linearity (linear classification with sigmoid non-linearity). The function should take three arguments: the data vectors, the parameters defining the neural network, and the pattern classes. (10 marks)

b) Compute the mean squared error for two different settings of the weights (i.e. two different decision boundaries). Select these by hand and choose settings that give large and small errors respectively. Plot both boundaries on the dataset. You will only use the 2nd and 3rd iris classes in this problem. (5 marks)

c) Give a mathematical derivation of the gradient of the objective function above with respect to the neural network weights. You will have to use chain rule and the derivative of the sigmoid function as discussed in class. Use w_0 to represent the bias term. You should show and explain each step. (10 marks)

d) Show how the gradient can be written in both scalar and vector form. (5 marks)

e) Write code that computes the summed gradient for an ensemble of patterns. Illustrate the gradient by showing (i.e. plotting) how the decision boundary changes for a small step. (10 marks)