

COEN 240 Machine Learning

Homework #2

Guideline: Please complete the following problems and generate a PDF file. Please submit the PDF file and a separate zip file that contains all source code to Camino. Please refer to HomeworkFormat.pdf for the format of the submitted PDF file.

Problem 1 For the K-means clustering problem, when the binary indicators (responsibilities) r_{kn} 's are fixed for $k=1, 2, \dots, K$ and $n=1, 2, \dots, N$, derive for the cluster centers \mathbf{m}_k , $k=1, 2, \dots, K$, such that the following objective function J is minimized:

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{kn} \|\mathbf{m}_k - \mathbf{x}_n\|_2^2$$

Problem 2 Iris.xls contains 150 data samples of three Iris categories, labeled by outcome values 0, 1, and 2. Each data sample has four attributes: sepal length, sepal width, petal length, and petal width.

Implement the K-means clustering algorithm to group the samples into $K=3$ clusters. Randomly choose three samples as the initial cluster centers. Calculate the objective function value J as defined in **Problem 1** after the assignment step in each iteration. Exit the iterations if the following criterion is met: $J(\text{Iter} - 1) - J(\text{Iter}) < \epsilon$, where $\epsilon = 10^{-5}$, and Iter is the iteration number. Plot the objective function value J versus the iteration number Iter. Comment on the result. Attach the code at the end of the homework.

Problem 3 Assume a data sample $\mathbf{x} \in \mathbb{R}^D$ comes from one of two classes, C_1 and C_2 . Use logistic regression to do classification.

- Write the math expression of the logistic regression output, and the criterion used for the final classification.
- How many parameters (weights) need to be calculated/trained in this method?

Problem 4 Assume a data sample $\mathbf{x} \in \mathbb{R}^D$ comes from one of K classes, C_1, C_2, \dots, C_K . Use logistic regression to do classification.

- Write the math expression of the logistic regression output, and the criterion used for the final classification.
- How many parameters (weights) need to be calculated/trained in this method?