Clustering Assignment in KNIME

3 questions

1. Create a KNIME workflow that utilizes the K-means clustering learning method to train a model on the Iris training data set. Set the initial value of K to default k=3. This corresponds well to the 3 existing classes of Iris categories as provided in the training data set.

   Was every instance of the Iris flower assigned to the "correct" corresponding type of Iris cluster?

   - ( ) Yes - every instance was assigned to a "correct" cluster
   - (•) No - some Iris-versicolor and Iris-virginica instances were assigned to a mix of clusters
   - ( ) No - some Iris-setosa instances were assigned to a mix of clusters

2. Download the bank training data set attached here and create your own K-means clustering model workflow.

   bank_data_Example.csv

   The marketing department of a financial firm keeps records on customers, including demographic information and the type of accounts. When launching a new product, such as a "Personal Equity Plan" (PEP), a direct mail piece, advertising the product, is sent to existing customers, and a record kept as to whether that customer responded and bought the product. The data contains the following fields

   id a unique identification number

   age age of customer in years (numeric)

   sex MALE / FEMALE

   region inner_city/rural/suburban/town

   income income of customer (numeric)

   married is the customer married (YES/NO)

   children number of children (numeric)

   car does the customer own a car (YES/NO)

   save_acct does the customer have a saving account (YES/NO)

   current_acct does the customer have a current account (YES/NO)

   mortgage does the customer have a mortgage (YES/NO)

   pep did the customer buy a PEP (Personal Equity Plan) after the last mailing (YES/NO)

   Experiment with several different values for K. Analyze the cluster assignment for the ID12101. For k=3 and K=5 was the ID12101 assigned to the same of different cluster?

   - ( ) ID12101 was assigned to the same cluster for both k=3 and k=5
   - (•) ID12101 was not assigned to the same cluster for both k=3 and k=5

3. For K-means clustering on the Bank data with K=6 what are the average characteristics for age and income of the largest cluster in the Bank data set? (without normalization)

   - ( ) Average Age=33; Average Income=$17200
   - ( ) Average Age=25; Average Income=$12548
   - ( ) Average Age=62; Average Income=$53786
   - (•) Average Age=42; Average Income=$23753

Submit Quiz