Angewandte KI in der Medizin
Exercise No. 1
due date: 1402/10/01
By:

Amirhossein Tabatabaei

December 12, 2023

Theoretical part

1. Please mention all different learning types with two examples for each. Is the number of clusters pre-determined in advance prior to clustering (unsupervised learning)? Support your reply with an example.

2. Please describe the Bayes decision rule for a sample binary classification problem. How this results can be extended to a 3-class classification problem?

3. Can we extend the Bayes concept to the regression problem wherein the feature data are continuous?

4. Please explain the concept of $K - NN$ classification method. Is is considered to be a linear or non-linear classifier? How can we determine an optimal value for K? Does it make sense to use it for a regression problem as well?

Practical part

There are some benchmark data normally available in Machine Learning in both classification and regression part. Here, we are going to use some of them in a real practice.

Data visualization

1. MNIST data set:
   (a) Download a subset of MNIST dataset using one of the repositories/resources in the Internet.
   (b) Use scatter graph to show the distribution of the classes.
   (c) The pixel values are playing the role of features, so for each image of size 28×28 there is a vector of size 784 elements. Please calculate the a-priori probability for each feature value per class.

Linear Classification and regression

2. Multiclass classification using linear classification
   (a) Consider the above multi-class classification which must be solved using a binary classification technique. At first split the data set into 70% train and 30% test. The use one-versus-rest to classify all objects in the train set. For example to classify 0 from the others there will be one label (e.g., 0) for 0s and another label (e.g., 1) for the rest and so on for other digits. Use linear classifier with Stochastic Gradient Method to come up with the 10 classifiers.
   (b) Using the extracted 10 classifiers in the last part, predict the image objects in the test set.
   (c) Calculate the accuracy of your classification scheme just for each classifier.
   (d) Repeat the above steps for different ratio of train/set size wherein 80% is dedicated to train and 20% to test set respectively.
   (e) Boston house price dataset:
      i. Download Boston house price dataset from one of the repositories.
      ii. visualize the data and describe the feature values.
      iii. Consider 80% of the entries as a train subset and the rest as a test subset and calculate the regression predictor based on a least square method.

3. Boston housing price dataset. This data set is a dataset collected by the U.S. Census Service concerning housing in the area of Boston MA which includes about 14 features. Repeat the above exercise for this dataset as well (Hint: due to the continuity of the features, regression task is performed here.)