|  |  |
| --- | --- |
| [Machine Learning]  [2021-1] |  |
| Homework 2  Lec 5, 6, 7 |  |
| [Due Date] 2021.04.23  Student ID :  Name :  Professor : Juntae Kim | logo-placeholder |

1. Answer following questions (30 pts)

* 1. Describe the odds ratio, logit, logistic function(including math formula) and their meaning.
  2. Describe the two Impurity measures for Decision Tree Learning(including math formula) and their meaning.
  3. Describe the Naïve Bayesian classifier(including math formula) and its meaning. Also explain how you can deal with the continuous feature values.

|  |
| --- |
| Your Answer |
|  |

2. Apply Logistic Regression, Decision Tree, Naïve Bayesian Classifier, k-Nearest Neighbor on Wine Dataset to predict the origin of wines. Describe the learned model, and compare the accuracies. (30 pts)

* Dataset

<https://archive.ics.uci.edu/ml/datasets/Wine>

The dataset is the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the 3 types of wines.

* Use downloaded raw data or scikit-learn library

<https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html>

from sklearn.datasets import load\_wine

wine = load\_wine()

X = wine.data

y = wine.target

* Check test accuracy (use 30% for test)

lr.score(X\_test\_std, y\_test)



|  |
| --- |
| Code |
|  |
| Result(Captured images) |
|  |
| Description |
|  |

3. Describe why the K-Nearest Neighbors method is not appropriate for dataset with large number of features. (10 pts)

|  |
| --- |
| Your Answer |
|  |

4. Apply Multi-layer Perceptron on Olivetti Faces Dataset to identify persons from images. Describe the learned model. (30 pts)

* Dataset

<https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_olivetti_faces.html>

The Olivetti Faces dataset contains a set of face images taken at AT&T Laboratories Cambridge. The sklearn.datasets.fetch\_olivetti\_faces function is the data fetching function that downloads the data.

There are 10 different images of each of 40 distinct persons. For some persons, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses).

The 64x64 pixels image is quantized to 256 grey levels and stored as unsigned 8-bit integers; the loader will convert these to floating point values on the interval [0, 1]. The target for this database is an integer from 0 to 39 indicating the identity of the person pictured.

* Check test accuracy (use 20% for test)

print("Training set score: %f" % mlp.score(X\_train, y\_train))

print("Test set score: %f" % mlp.score(X\_test, y\_test))



* Plotting several images (person 0, 1, 2)

from sklearn import datasets

from matplotlib import pyplot as plt

face = datasets.fetch\_olivetti\_faces()

X = face.data

y = face.target

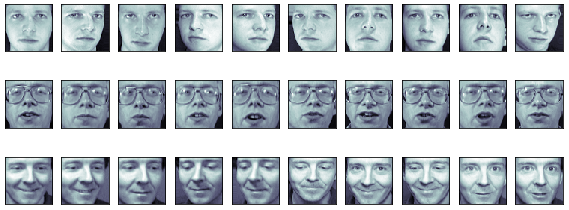
%matplotlib inline

fig = plt.figure(figsize=(10, 4))

for i in range(30):

ax = fig.add\_subplot(3, 10, i + 1, xticks=[], yticks=[])

ax.imshow(face.images[i], cmap=plt.cm.bone)



|  |
| --- |
| Code |
|  |
| Result(Captured images) |
|  |
| Description |
|  |

**Note**

1. Summit the file to e-class as pdf.

2. Specify your file name as “hw2\_<StudentID>\_<Name>.pdf”