COMP 4331: Assignment\_2 - Classification

# Assignment Description:

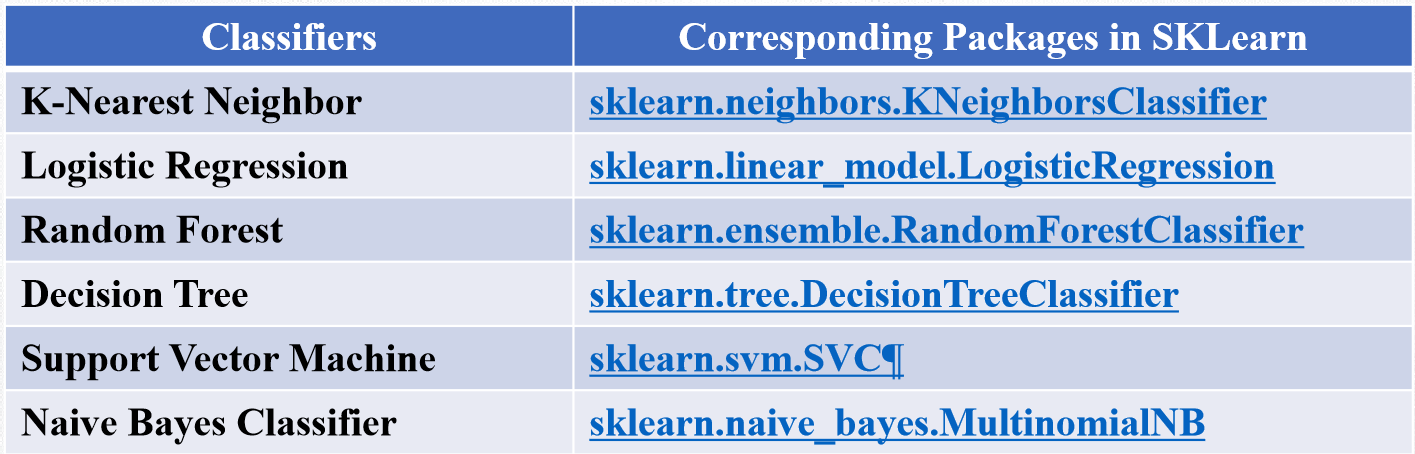
## Dataset Description

The dataset consists of a training set of 10,000 examples and a test set of 1,000

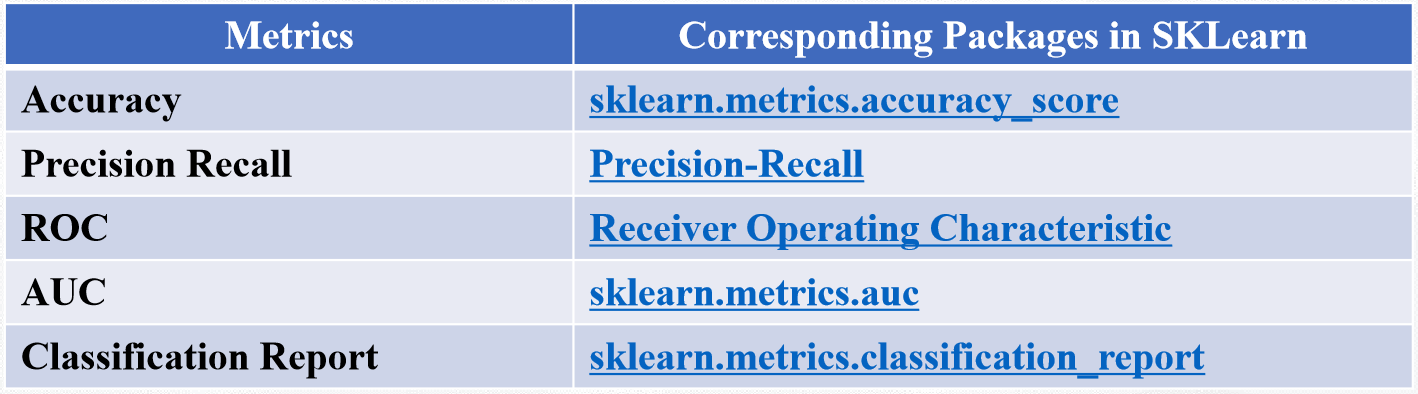
examples. Each example is a 28x28 grayscale image, associated with a label

from 10 classes.

## Classifiers referenced by tutorial notes



## Metrics reference by the tutorial notes



# Comparison of Classifiers:

## Decision Tree

### Basic information:

08ClassBasic.ppt slide 20

### Requirements:

Use the ***decision tree classifier*** under different criterions “***gini***” and “***entropy***”.

For each criterion, set the ***maximum depth*** of the decision tree as

***5*** and ***10*** separately.

You are required to compare the performance (precision, accuracy, f1 score

and recall) under different settings and give a brief discussion. (Totally 4

classifiers) (15 marks)

## KNN, SVM, Random Forest

### Requirements:

Apply three different classifiers KNN, SVM and Random Forest onto the

dataset. For each classifier, evaluate the performance (precision, accuracy,

f1 score and recall) and calculate the training time.

You are required to compare the performance of different classifiers and

give a brief discussion. (Totally 3 classifiers) (10 marks)

## Problems to ponder

1. How to choose the optimal K for KNN model? (Elbow Method)

<https://www.quora.com/How-can-I-choose-the-best-K-in-KNN-K-nearest-neighbour-classification>

<https://stats.stackexchange.com/questions/126051/choosing-optimal-k-for-knn>

1. How to tune the SVM parameter?

<https://medium.com/@aneesha/svm-parameter-tuning-in-scikit-learn-using-gridsearchcv-2413c02125a0>

## Multilayer perception

### Requirements:

Apply the multilayer perception network with different structures onto the

dataset. First, use just one single hidden layer with different numbers of

neurons as 50 and 100. Second, adopt two hidden layers with different

numbers of neurons as [100, 10] and [50, 20]. For each multilayer perception

network, evaluate the performance (precision, accuracy, f1 score and

recall) and calculate the training time.

You are required to compare the performance of the four multilayer perception

networks and give a brief discussion. (Totally 4 classifiers) (15

marks)

## Sklearn function name

MLPClassifier (hidden\_layer\_sizes=(100, ), activation=’relu’, solver=’adam’, alpha=0.0001, batch\_size=’auto’, learning\_rate=’constant’, learning\_rate\_init=0.001, power\_t=0.5, max\_iter=200, shuffle=True, random\_state=None, tol=0.0001, verbose=False, warm\_start=False, momentum=0.9, nesterovs\_momentum=True, early\_stopping=False, validation\_fraction=0.1, beta\_1=0.9, beta\_2=0.999, epsilon=1e-08)

## Data Split

### Requirements:

First, merge the training and the test sets in the original dataset, and

randomly split the whole dataset into new training (10,000 samples) and

test sets (1,000 samples). Apply KNN and SVM classifiers to the new

training and test sets.

You are required to compare the performance of each classifier on the original

training and test sets with that on the new training and test sets, and

give a brief discussion. (10 marks)