Self-Check Practice 4/10/2020   Data Science and Machine Learning
Part I

1. Time series _____ looks at existing time series data for pattern, help us to understand data. Time series _____ uses data from the past to predict the future.
   a) fetures, predict
   b) analysis, forecasting
   c) data, pattern
   d) mapping, analysis

2. The process of choosing the best value of k for the k nearest neighbors algorithm is called _____
   a) hyperparameter tuning
   b) machine tuning
   c) parameter analysis
   d) precision tuning

3. A Bunch object's _____ and _____ attributes are Numpy arrays containg the dataset's samples and labels repectively.
   a) information, set
   b) sample, category
   c) sample, label
   d) data, target

4. The process of famililiarizing yourself with data us called _____.
   a) data exploration
   b) data analysis
   c) data manipulation
   d) data mining

5. By default the train_test_split reserve _____ for data training and _____ for testing.
   a) 50%, 50%
   b) 75%, 25%
   c) 70%, 30%
   d) 90%, 10%

6. To change the default test size of train_test_split to 40%, one can provide the following as the parameter.
   a) test_size = 0.40 (or train_size = 0.60)
   b) test = 0.40 (or train = 0.60)
   c) frac = 0.40
   d) test_percent = 40

7. The KneighborClassifier is said to be _____ because its work is performed only when you use it to make prediction.
   a) slow
   b) lazy
   c) hidden
   d) smart

8. A Seaborn_____ displays values as colors, often with higher magnitude displayed as more intense colors..
   a) Plot
   b) heat map
   c) color map
   d) subplot

9. By default, LinearRegression estimator perform
   a) Multiple linear regression
   b) Simple linear regression
   c) Lasso regression
   d) Ridge regression

10 DataFrame method_____ return a randomly selected subset of the DataFrame's rows.
   a) random
   b) sample
   c) choose
   d) fold

Part II

1. Consider the NYC temperature application in 10_16.ipynb. Assuming linear trend continues, use the slope and intercept values, in what year the NYC average January temperature will reach 40 degree Fahrenheit?

2. For Digits dataset, how do we display the 8 x 8 image data and its numeric value of the sample number 45?

3. Continue from problem above, display the image of sample number 45.

```
year = 2019
slope = linear_regression.slope
intercept = linear_regression.intercept
temperature = slope * year + intercept
while temperature <40.0:

year +=1

temperature = slope*year+ intercept

year
```

```
digits.images[45]
digits.target[45]
```

```
axes = plt.subplot()
image = plt.imshow(digits.images[45], cmap = plt.cm.gray_r)
xticks = axes.set_xticks([])
yticks = axes.set_yticks([])
plt.show()
```