**Homework 3**

**Please submit this document to iCollege by 2021/4/18 11:00 pm. Late submission will be penalized.**

1. (10 points) What is the difference between supervised learning and unsupervised learning?

Supervised learning is when there is a specific mapping that takes place between an input and output based on some input-output pairs (extract y).

Unlike supervised machine learning, unsupervised machine learning deals with un-labelled data and allows the computer to make its own models (extract x).

1. (10 points) What are the definitions of regression and classification problems? Give an example (i.e. application) for each of them.

Classification helps to predict the class/group of certain things for some given observations. Its main goal is to predict how some input variables (X) are related to some discrete output variables (Y). This is used in filtering spam and non-spam emails.

Regression problems help with predicting a specific quantity. This maps how some input variables (X) are related to some continuous output variables (Y). This is used in predicted values in the stock market.

1. (10 points) What are training dataset, validation dataset and test dataset used for? What is the risk of training and evaluating a model on the same dataset?

Training dataset is a set of examples used to train the model and fit all the parameters in a model. Validation dataset is a set of examples used to “fine-tune” the architecture of the classifier. It is also called the development set. Test dataset is an independent dataset used to assess the performance of the training model.

Using the same dataset to train and evaluate a model would cause it to memorize and will not give you a clear picture of complete generalization in the data.

1. (10 points) How can you deal with the missing data? Explain it with pandas functions (e.g. dropna, fillna, etc.).

You can always fill missing values or drop them in dataframes in pandas. DataFrame.fillna(val) would replace every “null” or “missing” value in the dataframe with the value in the “val” variable. The DataFrame.dropna() would delete every set of values if even one entry in them is “null” or “missing.” This helps with cleaning up the dataframe, unlike fillna(), which populates it and keeps it intact.

1. (10 points) Calculate the mean, median, mode, range and standard deviation of following numbers.
2. 1, 2, 0, -2, 5, 8, -9

Mean: 5/7 = 0.714

Median: 1

Mode: NA (all appear only once)

Range: 8 - (-9) = 17

Standard Deviation: 5.006 ~ sqrt(175.43/7)

1. -3, 2, -5, 2, 5, -10, 9

Mean: 0/7 = 0

Median: 2

Mode: 2

Range: 9 - (-10) = 19

Standard Deviation: 5.95 ~ sqrt(248/7)

1. (10 points) Answer the following questions with codes. Assume the dictionary is named as *d*.

1). Get all the keys of a dictionary.

for i in d:

print(i)

2). Get all the items of a dictionary.

for i in d:

print(d[i])

3). Sort a dictionary by keys.

for i in sorted(d.keys()):

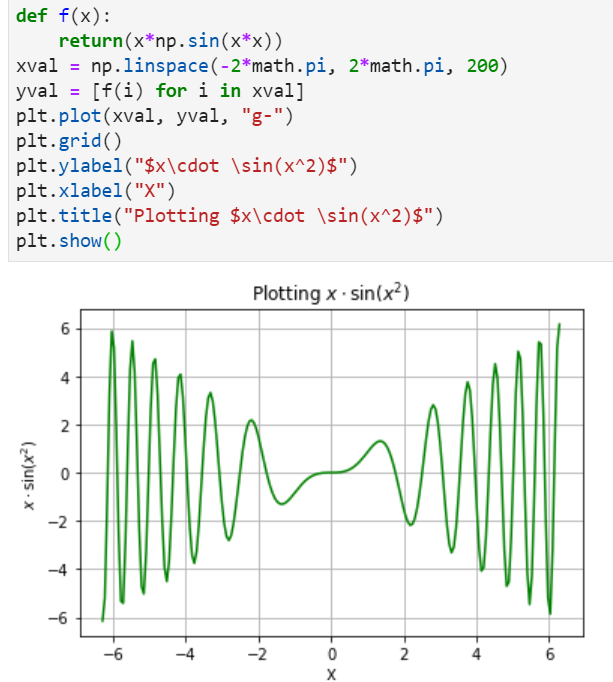
print(i, d[i])

4). Sort a dictionary by values.

for key, val in sorted(d.items(), key = lambda x: x[1]):

print(key, value)

1. (20 points) Plot y=x\*sin(x2) from -2π to 2π. Show your code and figure.



1. (20 points) Suppose *df* has following value.

Table

Description automatically generated

Are the following codes valid? Explain the meaning of each line if it is correct.

1. df[0]: This is invalid
2. df[‘0’]: This is invalid
3. df[‘sepal length (cm)’]: This fetches the values of the column “sepal length (cm)” along with the 0-5 column to its left.
4. df.loc[0]: This prints the first row of the dataframe with column names. In this case, it would print: sepal length (cm) 5.1

sepal width (cm) 3.5

petal length (cm) 1.4

petal width (cm) 0.2

1. df.loc[‘sepal length (cm)’]: This is invalid
2. df.iloc[0]: This prints the first row of the dataframe with column names. In this case, it would print: sepal length (cm) 5.1

sepal width (cm) 3.5

petal length (cm) 1.4

petal width (cm) 0.2

1. df.iloc[‘sepal length (cm)’]: This is invalid

Do you have any suggestions for instructor or TA to improve this class? Any comment is welcomed!

Please feel free to leave it below ☺