

Artificial Intelligence

Home Work – 4

Akshara Boppidi

Solution to 1(A):

Filename in git: hw4a.py

To generate 1000 data points, by using given python code:

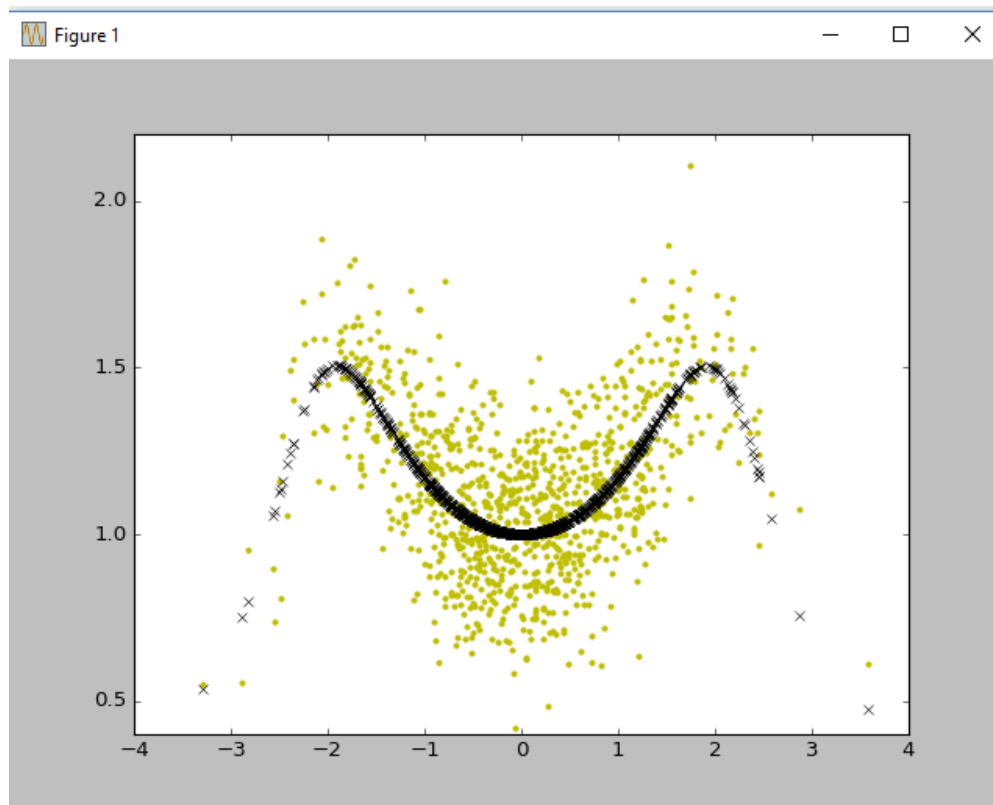
```
x=np.random.normal(0,1,N)
```

// here x is a input vector and np.random.normal(0,1,N) draws the random samples from a normal (Gaussian) distribution. Where, 0 is the mean of the distribution, 1 is the standard deviation and N is the size.

```
y=ytrue+noise
```

//here a random noise is added to true y and ‘y’ is the noisy output, ‘ytrue’ is the output without any noise.

Yellow (.) indicates the noisy output and black (+) indicates the true value.



Solution to 1(B):

File name in git: hw4.py

Using 10-fold CV, here we are splitting data into ten groups (training/test) by the dataset of 1(a). And we are varying `n_neighbors` in range from 1, 3, 5... $2\lfloor \frac{N+1}{2} \rfloor - 1$. As we are taking Dataset $N=1000$, $2\lfloor \frac{1000+1}{2} \rfloor - 1 \Rightarrow 2\lfloor 500.5 \rfloor - 1 \Rightarrow 1000-1 \Rightarrow 999$. So, we are varying `n_neighbors` in range from 1, 3, 5...999. For example, here we are taking 900 training dataset and 100 test dataset for a dataset of 1000. We are generating a dataset by using the training set, which is tested on the validation/test dataset and the best `k` is picked. We are obtaining three best values of `k-neighbors` that yield the best CV Eout. I.e. by knowing its corresponding index and its corresponding values of the `k-neighbors` which can be observed in below output.

```
0.274946000102
323
0.274927695045
297
0.274927667968
293
C:\Users\akshara\Documents\AI>
```

Solution to 1(C):

To report the best CV Eout:

```
C:\Users\akshara\Documents\AI>hw4.py
Eout:
1.0
Eouttrue
-0.79467483009
```

Here, by taking the three best values of `k-neighbors` yielded above 323, 297, 293 and by taking average of these three values, we obtain a value: 304 and we replace the `n_neighbors` in the instance of neighbor's regressor by this value.

The resultant produced, Eout is the perfect positive correlation coefficient. The perfect positive correlation means correlation coefficient is exactly equal to 1. That means the two datasets are plotted in the same direction. Whereas, a negative correlation means that two datasets move in opposite directions. Correlation coefficient values fall between -1 and 1.

Solution to 2:

By using the same dataset, we are repeating the experiment 100 times for storing the best three `k` number of neighbors in every single trail. For which, we are firstly creating an empty array named

'result' to store best three k values and creating a for loop for repeating the program 100 times. After getting the sorted array of best k values, we are storing the three best values and appending these values into our result array. Now, after the end of all trials we are plotting a histogram for all the values of k we have saved, which is seen in the figure below.

Histogram:

