

Assignment 1

Name - Pranav Mishra

Roll no. - 2018IMT-068

Course Name: Machine Learning Lab

Course code: ITIT-4107-2021

1a.

Aim: To create a random dataset of 2 input variables named X1,X2 with 20 observations with mean 8 and standard deviation of 1.5, another 20 observations with mean 15 and standard deviation 1.5 an output variable y 0 for first 20 observations and 1 for the rest 20.

Procedure: We will use the **Scipy** module for creation of sample data for both input features that will follow the **Normal Distribution** and the output feature will also be created using the same module that will follow the **Binomial Distribution**. We will use the **Pandas** library to create a Dataframe between both input features and the output label and all the plotting will be done using the **Matplotlib** Module.

code:

```
import numpy as np
import matplotlib.pyplot as plt

rand = np.random

fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(12, 7))

i0input1=rand.normal(8,1.5,20)
i0input2 =rand.normal(8,1.5,20)
i1input1 =rand.normal(15,1.5,20)
i1input2 =rand.normal(15,1.5,20)

plt.subplot(2, 2,1)
plt.title('instance 0 input 1')
plt.hist(i0input1, bins=50,color='yellow')

plt.subplot(2, 2,2)
plt.title('instance 0 input 2')
```

```
plt.hist(i0input2, bins=50,color='blue')
```

```
plt.subplot(2, 2,3)
```

```
plt.title('instance 1 input 1')
```

```
plt.hist(i1input1, bins=50,color='red')
```

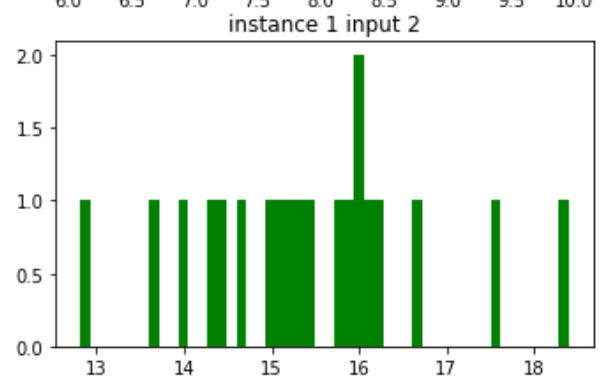
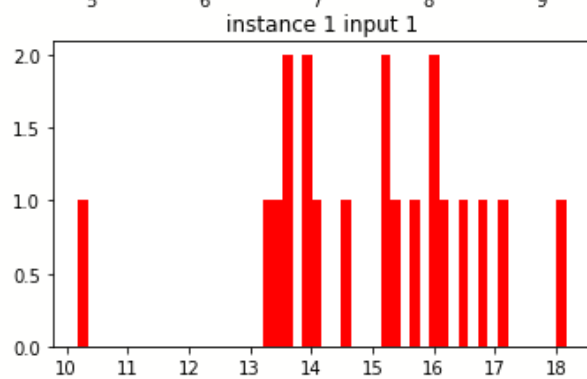
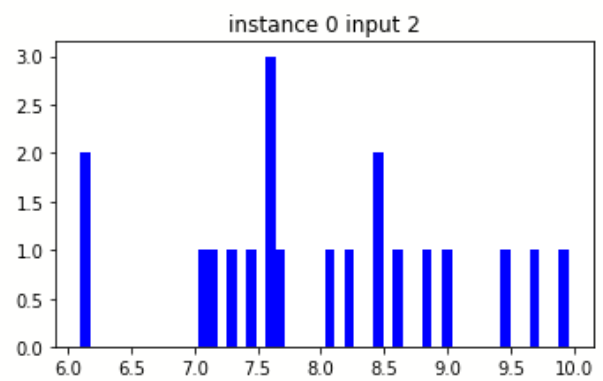
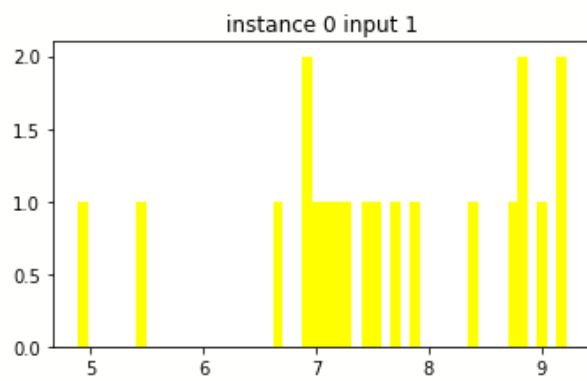
```
plt.subplot(2, 2,4)
```

```
plt.title('instance 1 input 2')
```

```
plt.hist(i1input2, bins=50,color='green')
```

```
plt.show()
```

output:



1b.

Aim: To create a random dataset of 2 input variables named X1,X2 with mean 8 and standard deviation of 1.5 and an output variable y sampled from a binomial distribution with probability of 1 as 0.6.

Procedure: We will use the **Scipy** module for creation of sample data for both input features that will follow the **Normal Distribution** and the output feature will also be created using the same module that will follow the **Binomial Distribution**. We will use the **Pandas** library to create a Dataframe between both input features and the output label and all the lotting will be done using the **Matplotlib** Module.

Code:

```
import numpy as np
import matplotlib.pyplot as plt

rand = np.random

fig, ax = plt.subplots(nrows=1, ncols=3, figsize=(12, 7))

input1=rand.normal(8,1.5,20)
input2=rand.normal(8,1.5,20)
output=rand.normal(100,0.6,20)

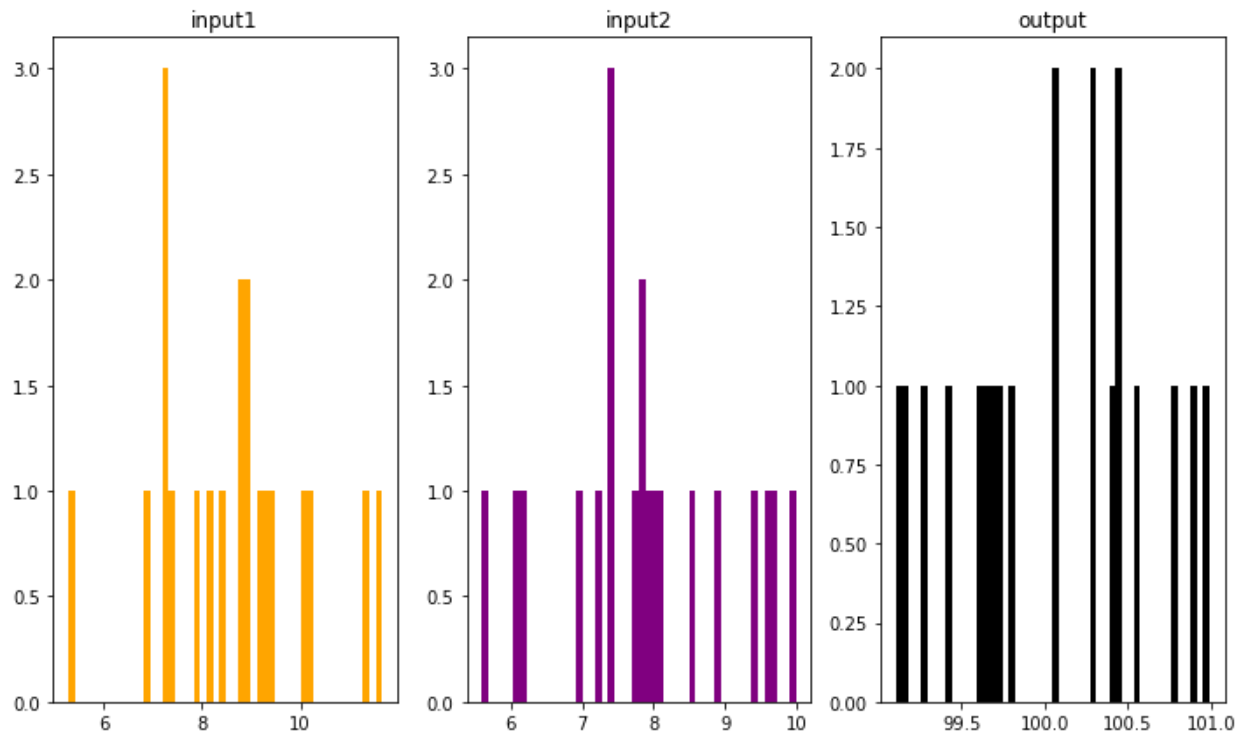
plt.subplot(1, 3,1)
plt.title('input1')
plt.hist(input1, bins=50,color='orange')

plt.subplot(1, 3,2)
plt.title('input2')
plt.hist(input2, bins=50,color='purple')

plt.subplot(1, 3,3)
plt.title('output')
plt.hist(output, bins=50,color='black')

plt.show()
```

Output:



1c

Aim: To create synthetic dataset using `make_moon`, `make_circles` and `make_blobs` from sklearn library.

Procedure: We will use the python **Sklearn** module that is used in creating synthetic dataset for the classification and clustering problem Plotting is done using the **Matplotlib** Module.

Code:

```
from sklearn.datasets import make_moons
from sklearn.datasets import make_blobs
from sklearn.datasets import make_circles
import matplotlib.pyplot as plt
fig, ax = plt.subplots(nrows=3, ncols=3, figsize=(7, 12))

# Make Moons
X1, y1 = make_moons(n_samples=100, noise=0.05)

plt.subplot(3, 1, 1)
plt.title('Make_Moons')
```

```
plt.scatter(X1[:, 0], X1[:, 1], marker='o', c=y1, s=25, edgecolor='k')

# Make Circles
X2, y2 = make_circles(n_samples=100, noise=0.1)

plt.subplot(3,1,2)
plt.title('Make_Circle')
plt.scatter(X2[:, 0], X2[:, 1], marker='o', c=y2, s=25, edgecolor='k')

# Make Blobs
X3, y3 = make_blobs(n_samples=100, centers=3, n_features=2)

plt.subplot(3,1,3)
plt.title('Make_Blobs')
plt.scatter(X3[:, 0], X3[:, 1], marker='o', c=y3, s=25, edgecolor='k')

plt.show()
```

Output:

