Dataset1-Regression_output_1

October 7, 2021

1 Dataset 1 - Regression

1.1 Import Libraries

[1]:
```
import train_test
import ABC_train_test
import regressionDataset
import network
import statsModel
import performanceMetrics
import dataset
import sanityChecks
import torch
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import norm
from torch.utils.data import Dataset,DataLoader
from torch import nn
import warnings
warnings.filterwarnings('ignore')
```

1.2 Parameters

General Parameters

1. Number of Samples

Discriminator Parameters

1. Size : number of hidden nodes

ABC-Generator parameters are as mentioned below: 1. mean : 1 ($\beta \sim N(\beta^*, \sigma)$ where β^* are coefficients of statistical model) or 1 ($\beta \sim N(0, \sigma)$ 2. std : $\sigma = 1, 0.1, 0.01$ (standard deviation)

[2]:
```
n_features = 10
sample_size = 100
#Discriminator Parameters
hidden_nodes = 25
#ABC Generator Parameters
mean = 1
```

```
variance = 0.001
```

1.3 Dataset

Generate a random regression problem

$Y = 1 + \beta_1 x_1 + \beta_2 x_2 + \beta_2 x_3 + ... + \beta_n x_n + N(0, \sigma)$ where $\sigma = 0.1$

[3]:
```
X,Y = regressionDataset.regression_data(sample_size,n_features)
```

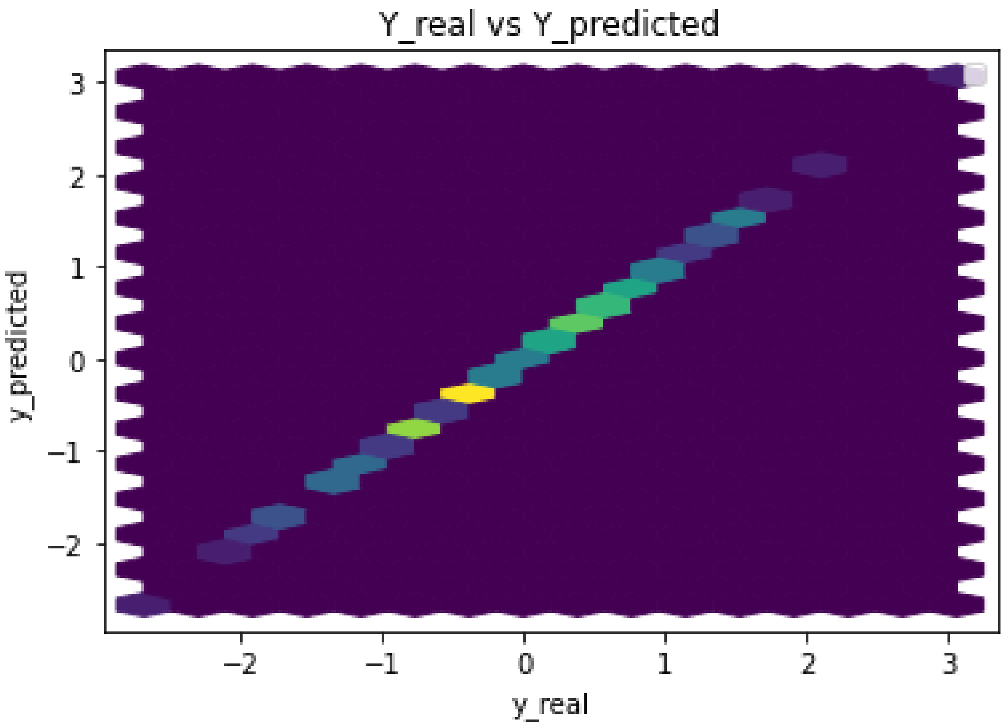
```
         X1        X2        X3        X4        X5        X6        X7  \
0 -0.467569  0.032126 -1.225677  0.841746  0.220220  1.431035 -0.260713
1 -0.329127  0.543150  1.036346 -0.794104 -0.234080  1.255848 -1.328471
2  1.203425 -1.393708 -1.208895 -1.615668  1.887532  0.106607  0.328814
3  0.007436 -0.583387 -0.687340 -1.406820  0.653945 -1.400086  0.047870
4  0.635016 -0.077755  0.251223 -1.051868 -1.349735 -0.565553  0.765214

         X8        X9       X10           Y
0  2.080232  2.399943  0.659622  166.198911
1 -0.140738  0.648646 -1.034550  -90.549344
2 -0.748738 -0.124472 -1.773646 -106.343704
3 -1.145719  0.073097 -1.004091 -197.592009
4  0.230325 -0.935316  1.314890  -27.918795
```

1.4 Stats Model

[4]:
```
[coeff,y_pred] = statsModel.statsModel(X,Y)
```

```
No handles with labels found to put in legend.

                            OLS Regression Results
==============================================================================
Dep. Variable:                      Y   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 1.539e+07
Date:                Thu, 07 Oct 2021   Prob (F-statistic):          5.23e-273
Time:                        07:15:05   Log-Likelihood:                 576.28
No. Observations:                 100   AIC:                            -1131.
Df Residuals:                      89   BIC:                            -1102.
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -3.816e-17   8.06e-05  -4.74e-13      1.000      -0.000       0.000
x1             0.3947   8.54e-05   4624.741      0.000       0.395       0.395
x2             0.4489   8.75e-05   5128.137      0.000       0.449       0.449
x3             0.1497   8.32e-05   1798.917      0.000       0.149       0.150
x4             0.4983   8.29e-05   6009.497      0.000       0.498       0.499
x5             0.3495   8.62e-05   4054.849      0.000       0.349       0.350
```

```
x6             0.1555   8.45e-05   1840.355      0.000       0.155       0.156
x7             0.3753   8.69e-05   4317.524      0.000       0.375       0.375
x8             0.3258   8.43e-05   3863.700      0.000       0.326       0.326
x9             0.1363   8.62e-05   1582.496      0.000       0.136       0.137
x10            0.2509   8.26e-05   3037.644      0.000       0.251       0.251
==============================================================================
Omnibus:                        7.753   Durbin-Watson:                   1.757
Prob(Omnibus):                  0.021   Jarque-Bera (JB):                7.723
Skew:                           0.523   Prob(JB):                       0.0210
Kurtosis:                       3.871   Cond. No.                         1.65
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
Parameters:  const   -3.816392e-17
x1       3.947487e-01
x2       4.489328e-01
x3       1.496617e-01
x4       4.983450e-01
x5       3.495472e-01
x6       1.554696e-01
x7       3.752957e-01
x8       3.258113e-01
x9       1.363344e-01
x10      2.509494e-01
dtype: float64
```




```
Performance Metrics
Mean Squared Error: 5.781476153247211e-07
Mean Absolute Error: 0.0005982846053422965
Manhattan distance: 0.059828460534229654
Euclidean distance: 0.007603601878877674
```

2 Generator and Discriminator Networks

GAN Generator

[5]:
```
class Generator(nn.Module):

  def __init__(self,n_input):
    super().__init__()
    self.output = nn.Linear(n_input,1)

  def forward(self, x):
    x = self.output(x)
    return x
```

GAN Discriminator

[6]:
```
class Discriminator(nn.Module):
```

```
    def __init__(self,n_input,n_hidden):

      super().__init__()
      self.hidden = nn.Linear(n_input,n_hidden)
      self.output = nn.Linear(n_hidden,1)
      self.relu = nn.ReLU()

    def forward(self, x):
      x = self.hidden(x)
      x = self.relu(x)
      x = self.output(x)
      return x
```

ABC Generator

The ABC generator is defined as follows:

$Y = 1 + \beta_1 x_1 + \beta_2 x_2 + \beta_2 x_3 + ... + \beta_n x_n + N(0, \sigma)$ where $\sigma = 0.1$

$\beta_i \sim N(0, \sigma^*)$ when $\mu = 0$ else

$\beta_i \sim N(\beta_i^*, \sigma^*)$ where $\beta_i^* s$ are coefficients obtained from stats model

Parameters : μ and σ^*

σ^* takes the values 0.01,0.1 and 1

[7]:
```
def ABC_pre_generator(x_batch,coeff,variance,mean,device):

    coeff_len = len(coeff)

    if mean == 0:
      weights = np.random.normal(0,variance,size=(coeff_len,1))
      weights = torch.from_numpy(weights).reshape(coeff_len,1)
    else:
      weights = []
      for i in range(coeff_len):
        weights.append(np.random.normal(coeff[i],variance))
      weights = torch.tensor(weights).reshape(coeff_len,1)

    y_abc =  torch.matmul(x_batch,weights.float())
    gen_input = torch.cat((x_batch,y_abc),dim = 1).to(device)
    return gen_input
```

3 GAN Model

[8]:
```
real_dataset = dataset.CustomDataset(X,Y)
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
```

[9]:
```
generator = Generator(n_features+2)
discriminator = Discriminator(n_features+2,hidden_nodes)

criterion = torch.nn.BCEWithLogitsLoss()
gen_opt = torch.optim.Adam(generator.parameters(), lr=0.01, betas=(0.5, 0.999))
disc_opt = torch.optim.Adam(discriminator.parameters(), lr=0.01, betas=(0.5, 0.
↪999))
```

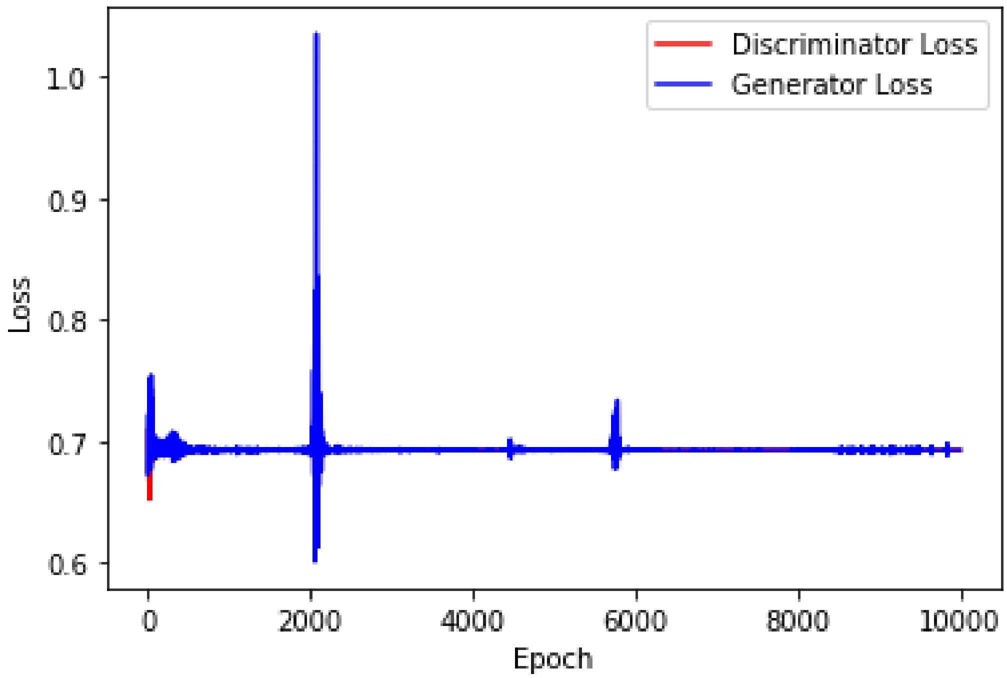
[10]:
```
print(generator)
print(discriminator)
```

```
Generator(
  (output): Linear(in_features=12, out_features=1, bias=True)
)
Discriminator(
  (hidden): Linear(in_features=12, out_features=25, bias=True)
  (output): Linear(in_features=25, out_features=1, bias=True)
  (relu): ReLU()
)
```

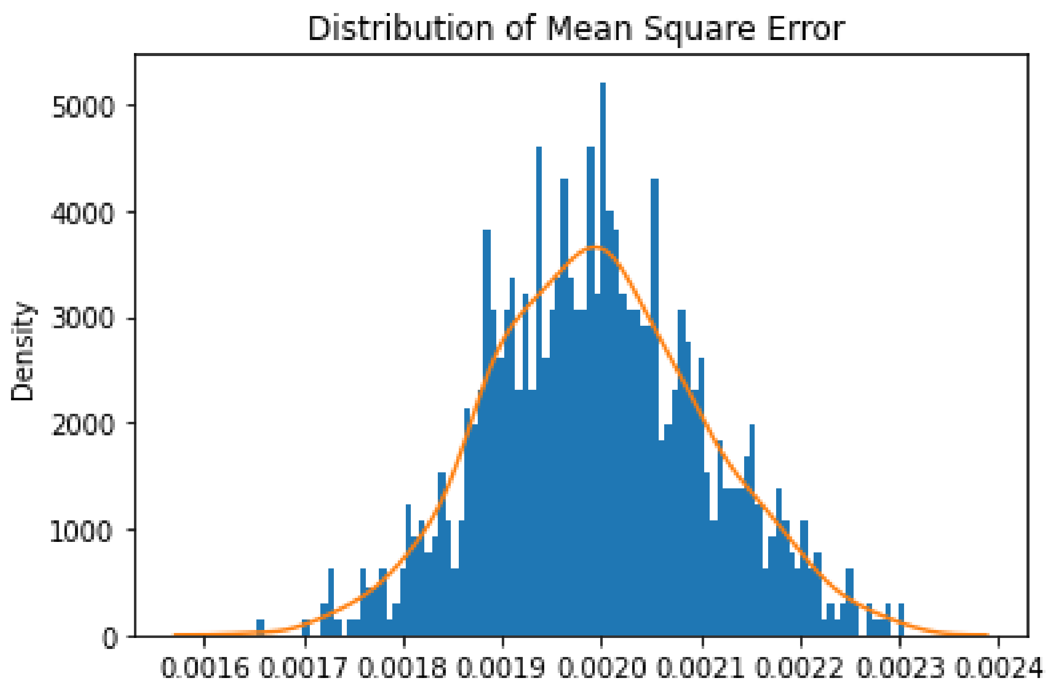
[11]:
```
n_epochs = 5000
batch_size = sample_size//2
```

[12]:
```
# Parameters
sample_size = 100
std = 1
mean = 1
```

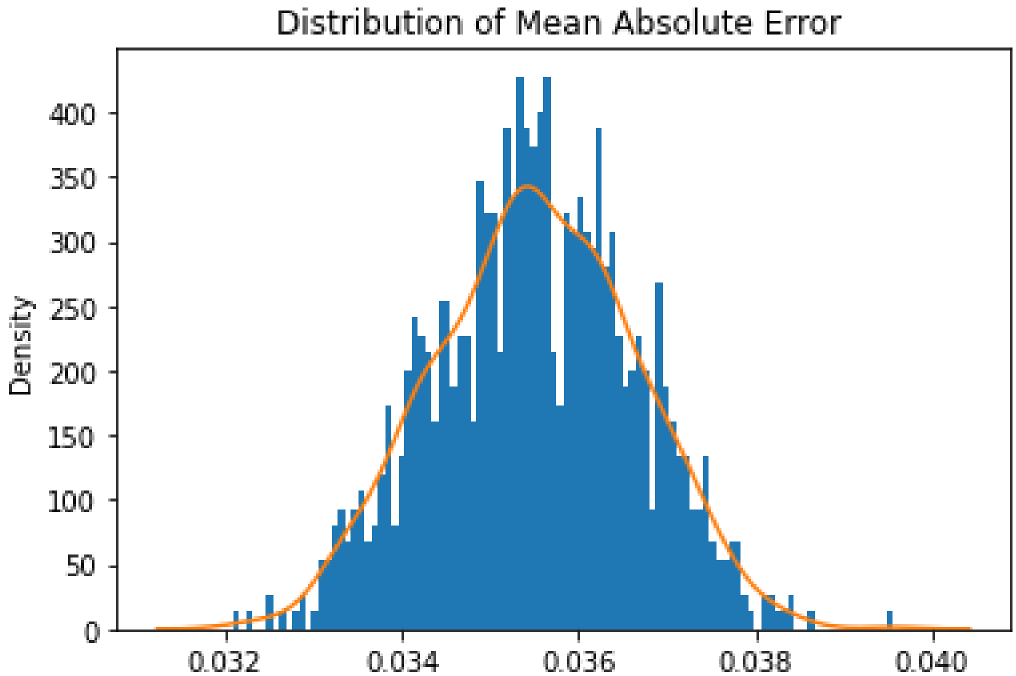
[13]:
```
train_test.
↪training_GAN(discriminator,generator,disc_opt,gen_opt,real_dataset,batch_size,␣
↪n_epochs,criterion,device)
```



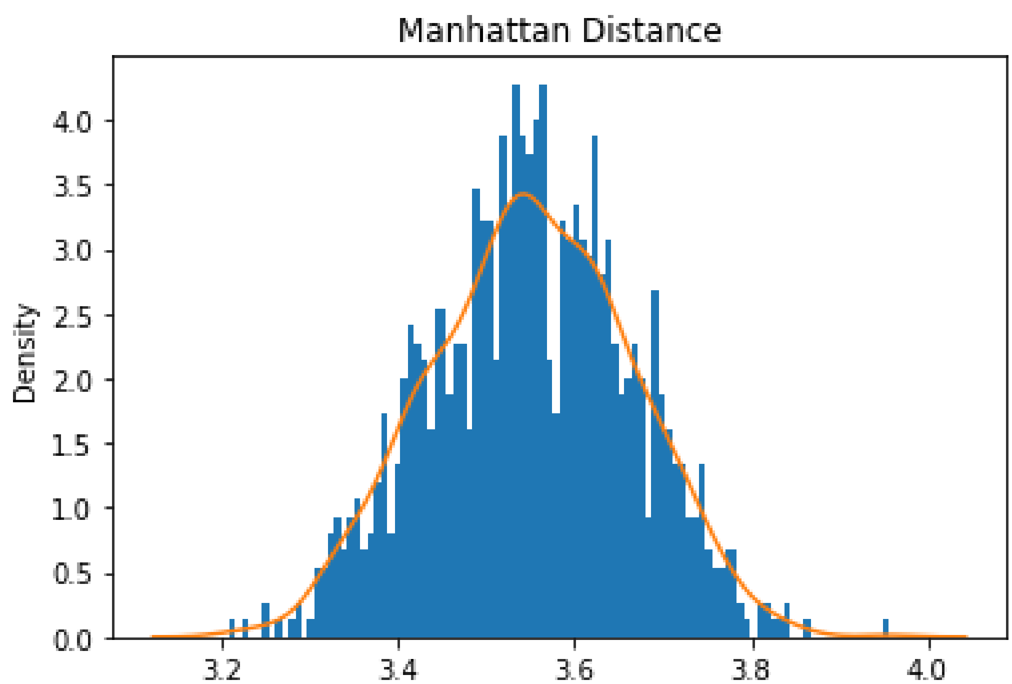
[14]:
```
train_test.test_generator(generator,real_dataset,device)
```



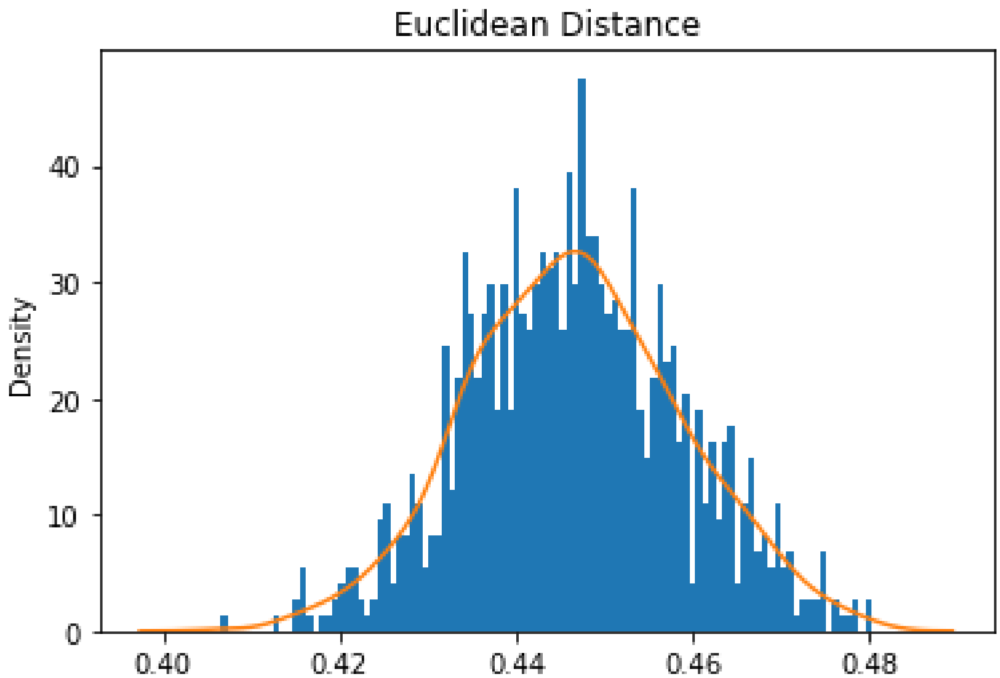
```
Mean Square Error: 0.0019959189323502853
```




```
Mean Absolute Error: 0.03550904961375519
```



Mean Manhattan Distance: 3.5509049613755197



Mean Euclidean Distance: 3.5509049613755197

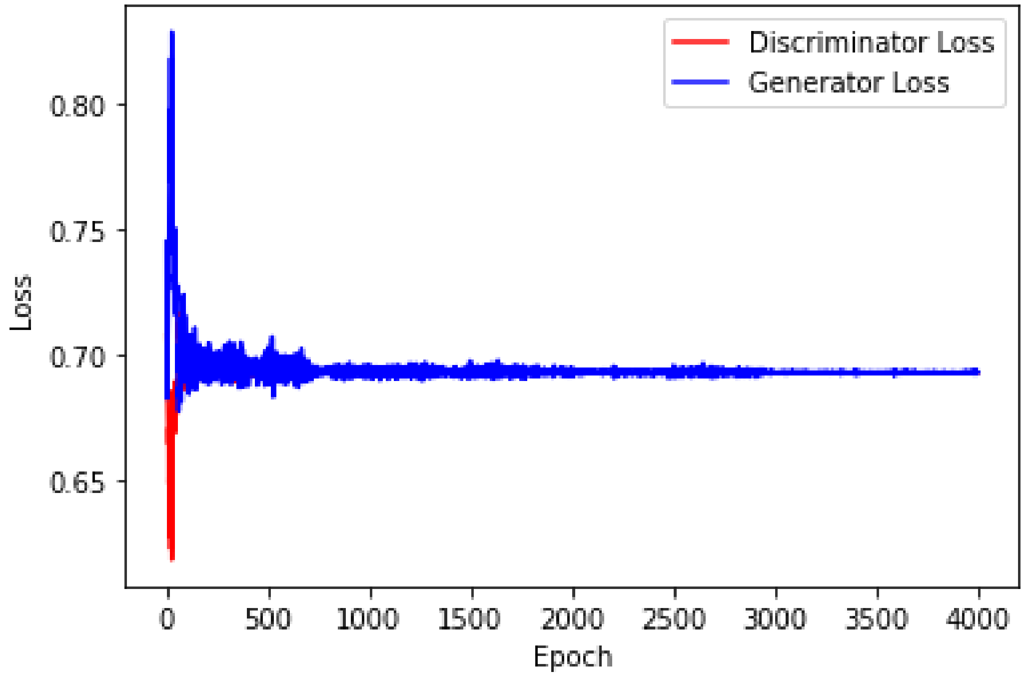
4 ABC GAN Model

Training the network

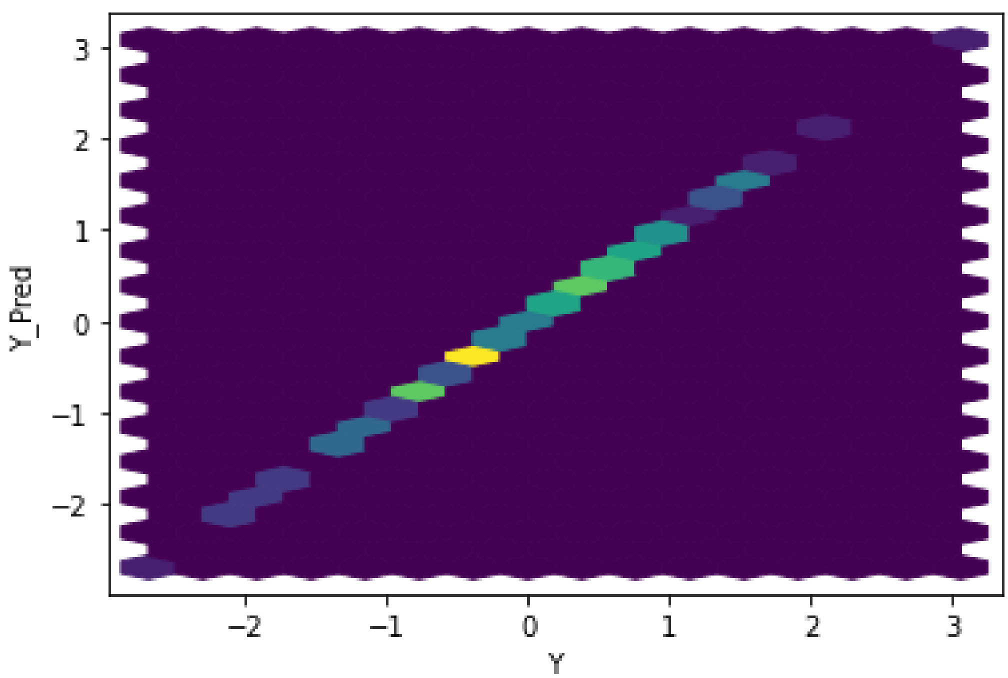
[15]:
```
gen = Generator(n_features+2)
disc = Discriminator(n_features+2,hidden_nodes)

criterion = torch.nn.BCEWithLogitsLoss()
gen_opt = torch.optim.Adam(gen.parameters(), lr=0.01, betas=(0.5, 0.999))
disc_opt = torch.optim.Adam(disc.parameters(), lr=0.01, betas=(0.5, 0.999))
```

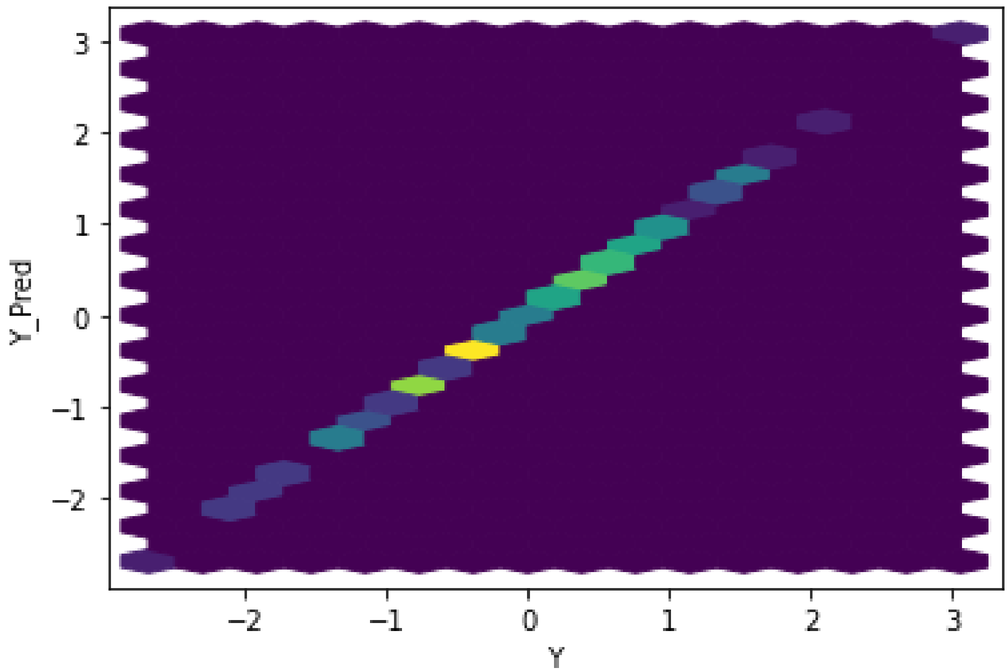
[16]:
```
n_epoch_abc = 2000
batch_size = sample_size//2
```

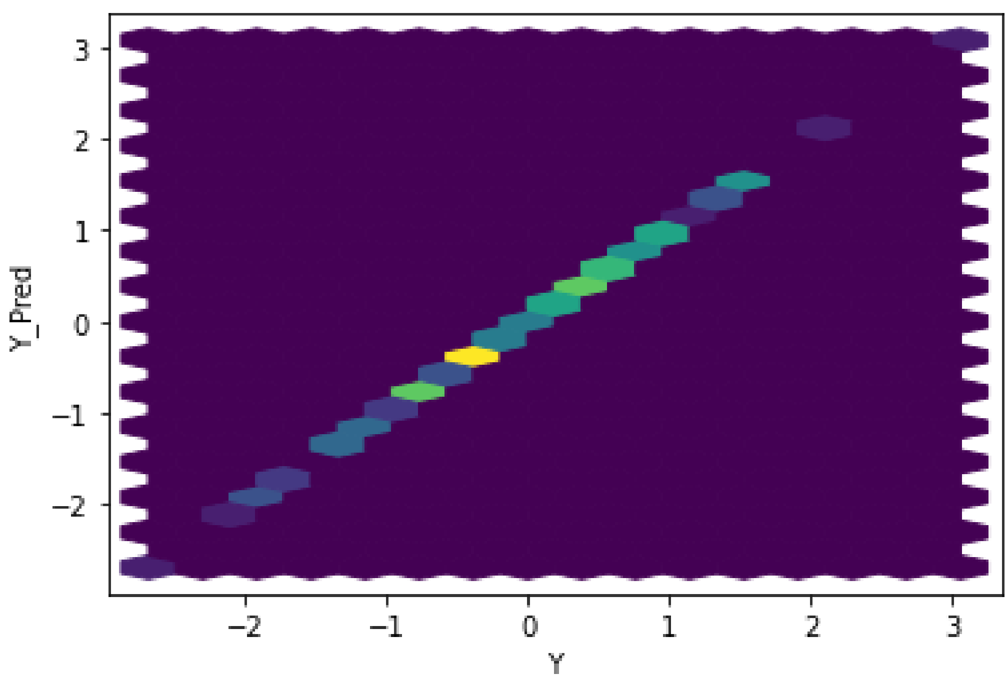
[17]:
```
ABC_train_test.training_GAN(disc, gen,disc_opt,gen_opt,real_dataset,␣
↪batch_size, n_epoch_abc,criterion,coeff,mean,variance,device)
```

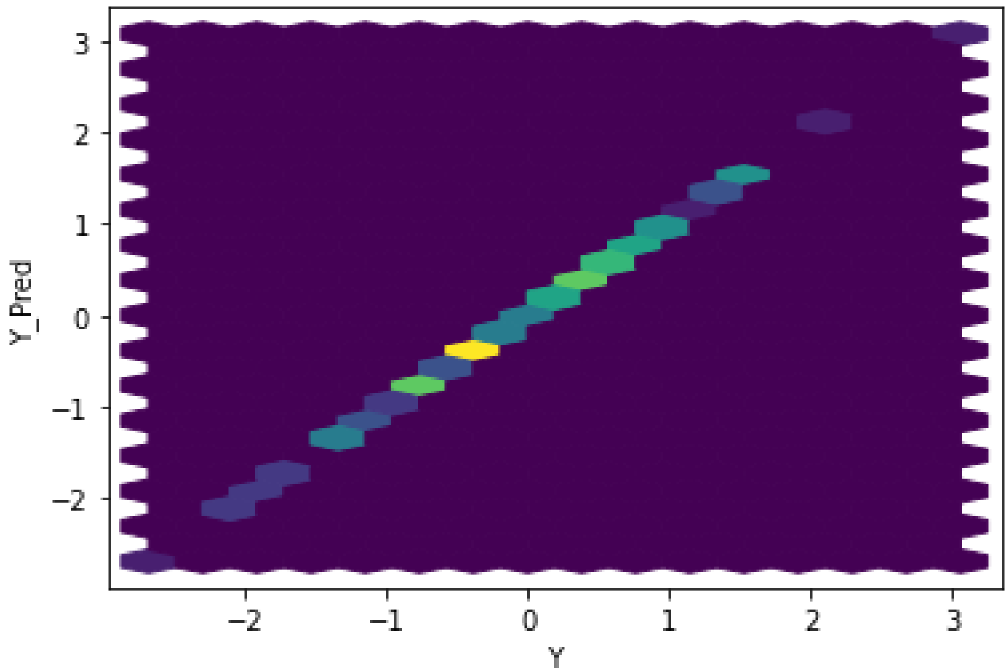


[18]:
```
ABC_train_test.test_generator(gen,real_dataset,coeff,mean,variance,device)
```

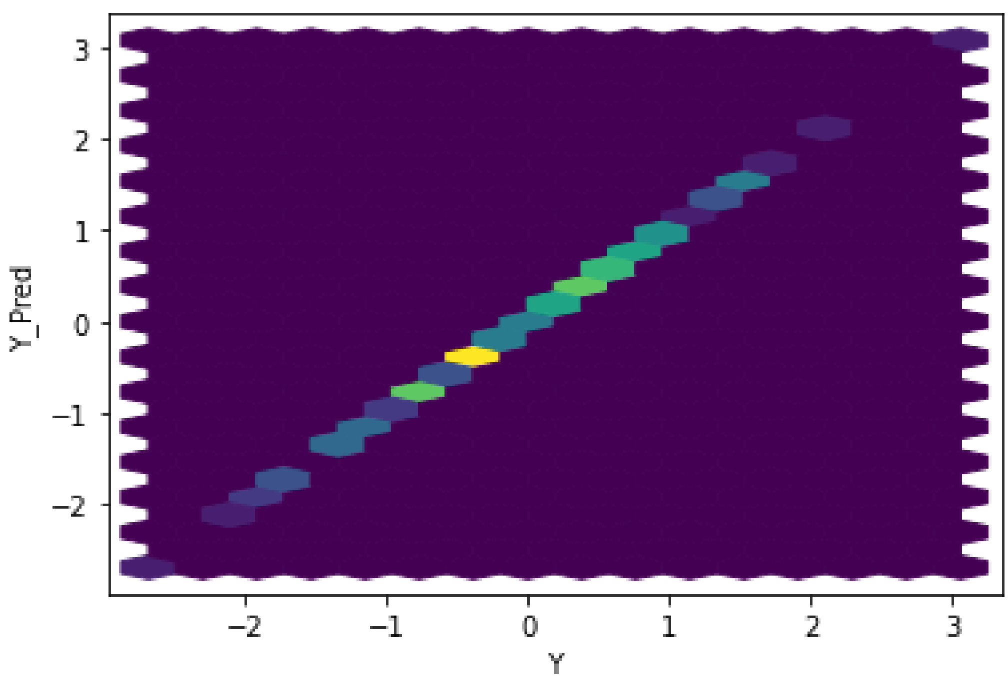



Y_Pred
3
2
1
0
−1
−2
−2
−1
0
1
2
3
Y

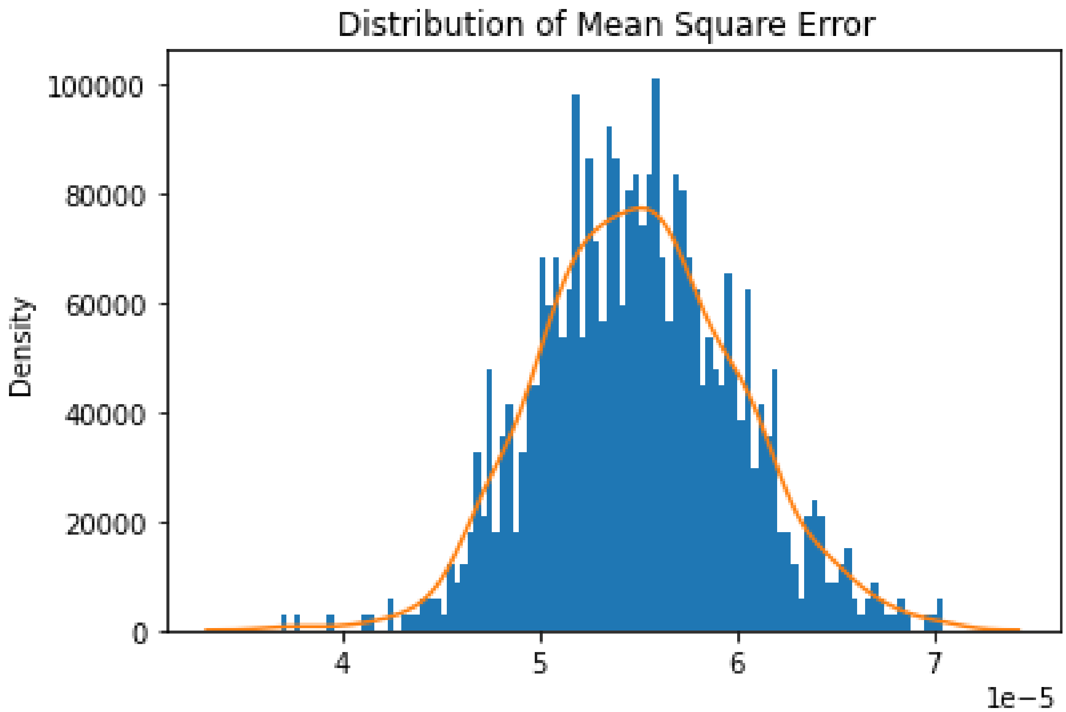

Y_Pred
3
2
1
0
−1
−2
−2
−1
0
1
2
3
Y



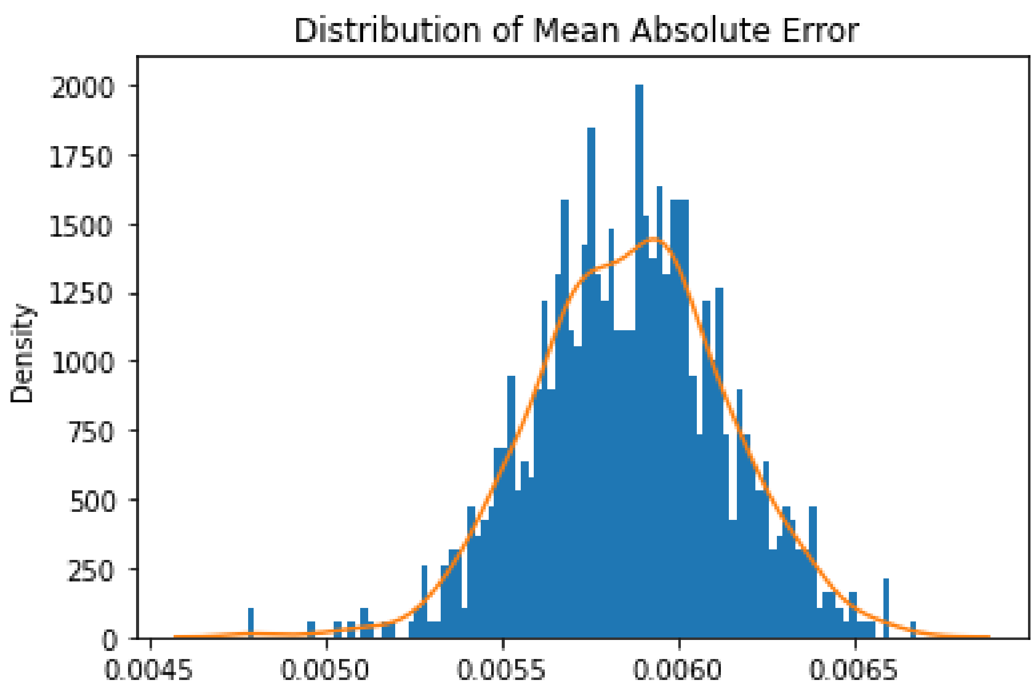
3
2
1
0
−1
−2
Y_Pred
−2
−1
0
1
2
3
Y



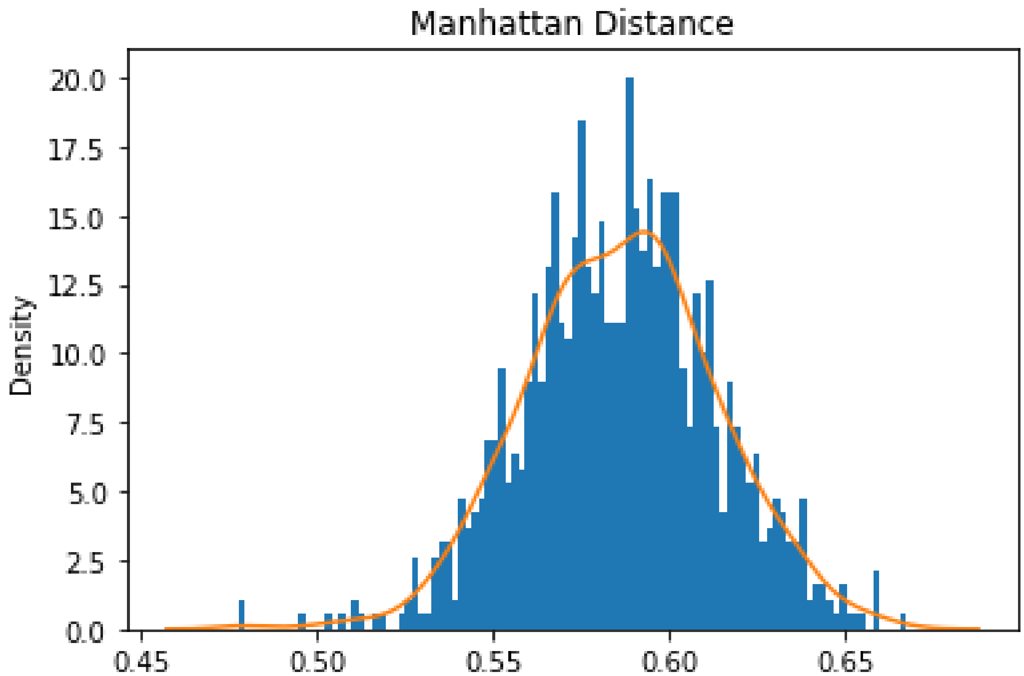
3
2
1
0
−1
−2
Y_Pred
−2
−1
0
1
2
3
Y



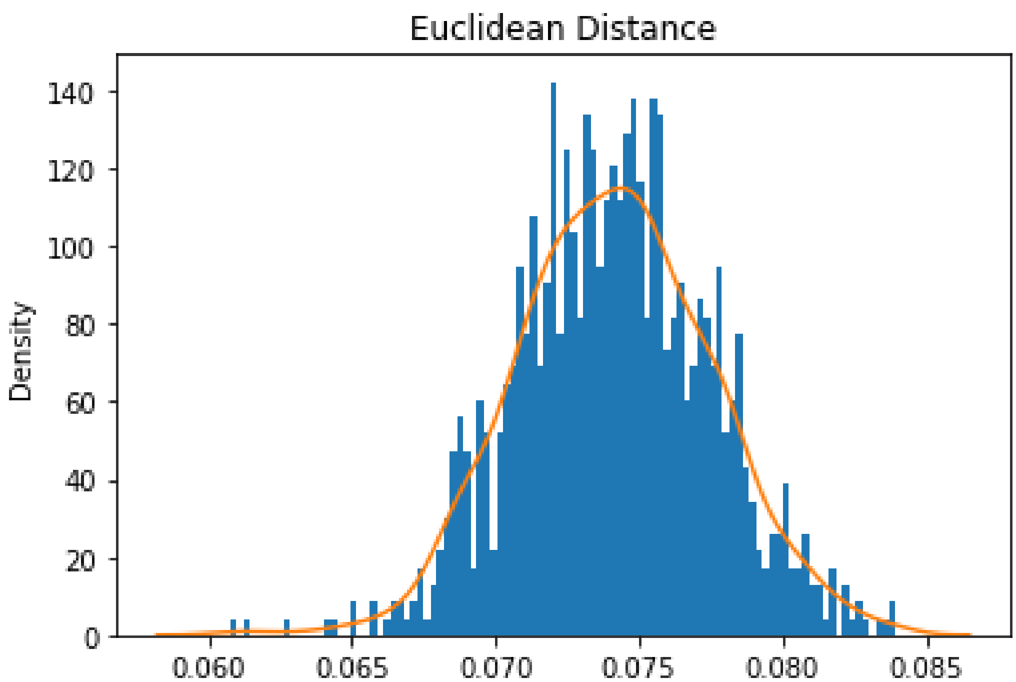

```
Mean Square Error: 5.5055287864622954e-05
```



```
Mean Absolute Error: 0.005864408159479499
Mean Manhattan Distance: 0.5864408159479498
```

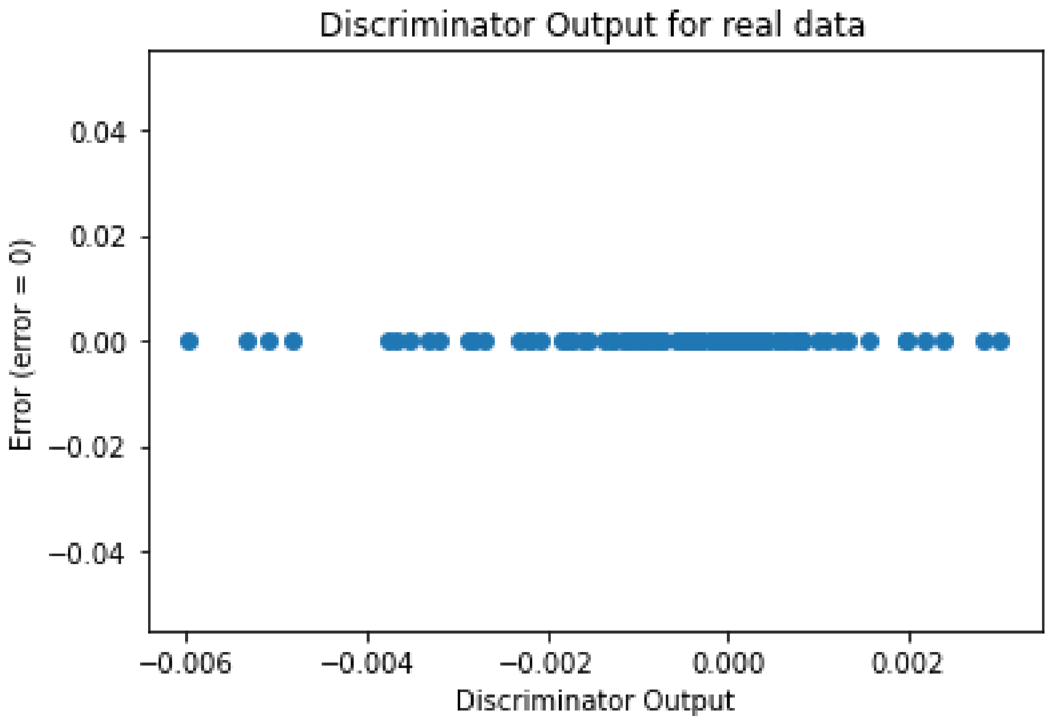


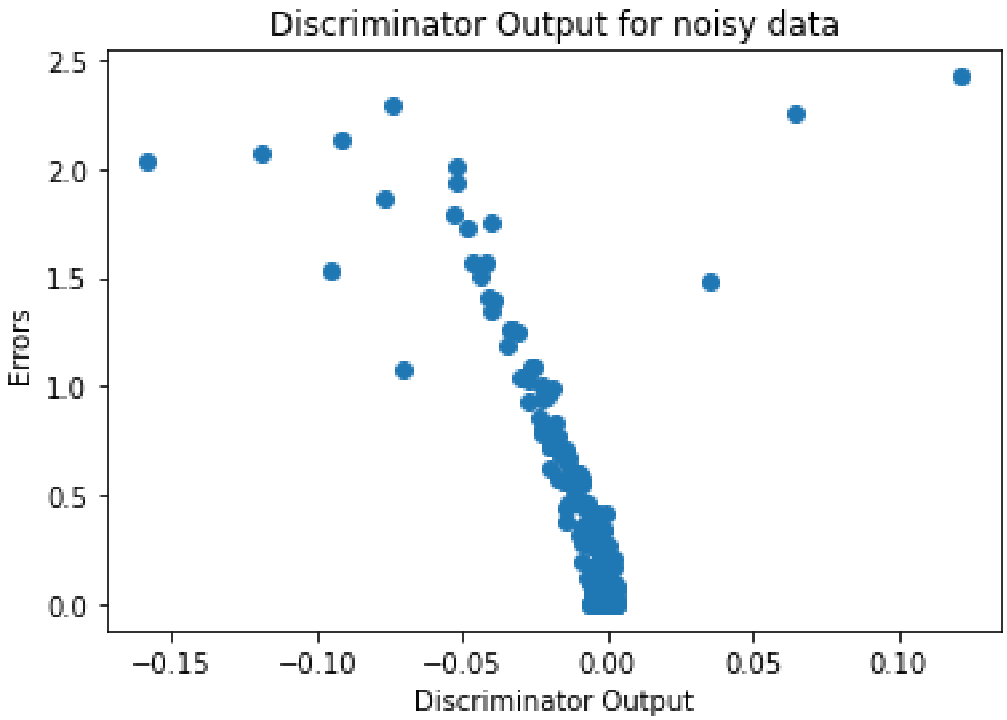

```
Mean Euclidean Distance: 0.07412229648492048
```



Sanity Checks

```
[19]: sanityChecks.discProbVsError(real_dataset,disc,device)
```





4.1 Visualization of trained GAN generator

[20]:
```
for name, param in gen.named_parameters():
    print(name,param)
```

```
output.weight Parameter containing:
tensor([[-0.0263,  0.2705,  0.3066,  0.1004,  0.3393,  0.2433,  0.1051,  0.2570,
          0.2217,  0.0931,  0.1740,  0.3189]], requires_grad=True)
output.bias Parameter containing:
tensor([0.0237], requires_grad=True)
```