Plot the training and testing datasets

September 26, 2019

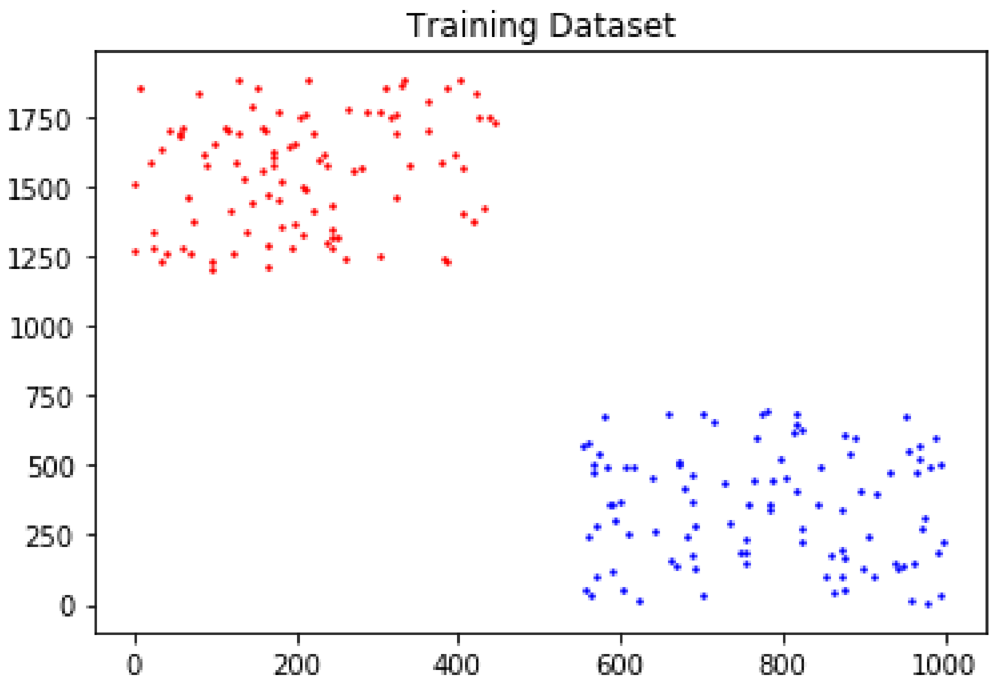
```
In [2]: import matplotlib.pyplot as plt
        import numpy as np
```

$i = 1, 2, \cdots, n$
$\hat{y}_i = \sigma(z_i)$
$z_i = w^T x_i + b$
$\sigma(z) = \frac{1}{1+\exp(-z)}$
$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} f_i(w, b)$
$f_i(w, b) = -y_i \log \hat{y}_i - (1 - y_i) \log(1 - \hat{y}_i)$

1. Plot two clusters of points for training dateset

```
In [52]: # u^prime = u - 10^(-5) * x
         # v^prime = v - 10^(-5) * y
         # b^prime = b - 10^(-5)

         x_lim = 1000

         X = np.empty(200, dtype=float)
         Y = np.empty(200, dtype=float)
         L = np.empty(200, dtype=float)

         # Training Dataset

         x_1 = np.random.randint(0, 450, 100)
         y_1 = np.random.randint(1200, 1900, 100)

         x_2 = np.random.randint(550, 1000, 100)
         y_2 = np.random.randint(0, 700, 100)

         for i in range(100):
             X[i] = x_1[i]
             X[100 + i] = x_2[i]
             Y[i] = y_1[i]
             Y[100 + i] = y_2[i]
```

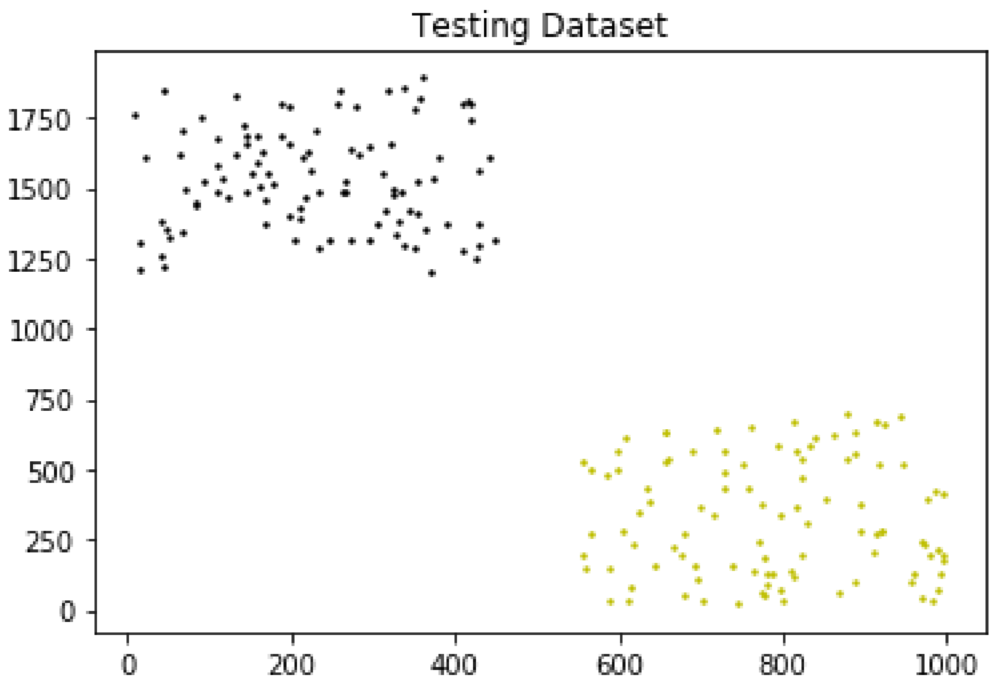
```
    L[i] = 0
    L[100 + i] = 1

plt.title('Training Dataset')
for x in range(200):
    if x < 100:
        plt.scatter(X[x], Y[x], c='r', s=2)
    else:
        plt.scatter(X[x], Y[x], c='b', s=2)
```



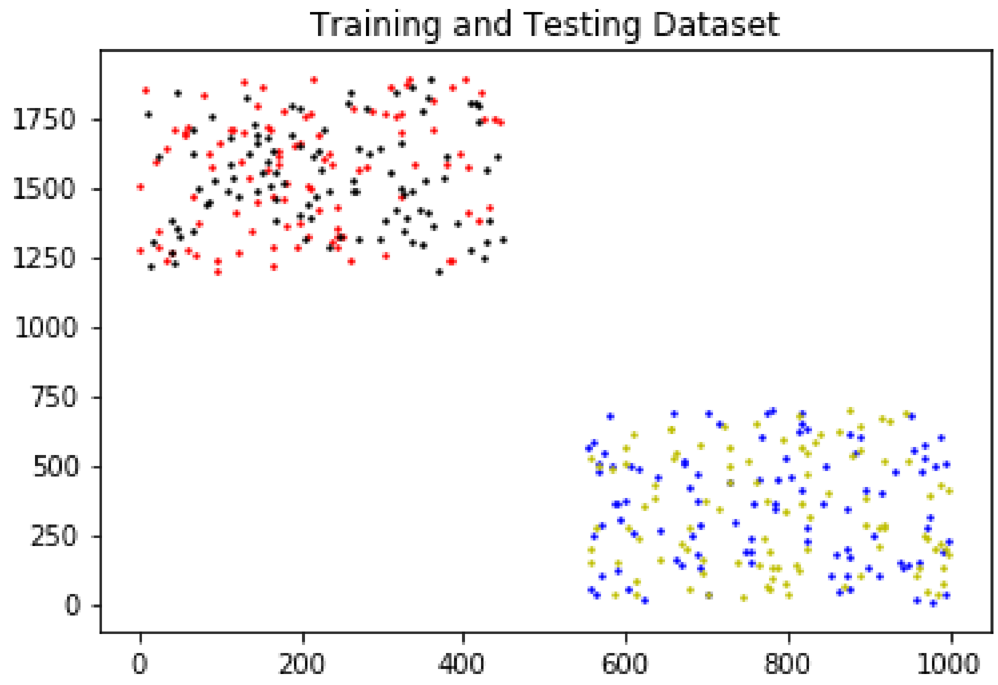
2. Plot two clusters of points for testing dataset

```
In [53]: # Testing Dataset

         tX = np.empty(200, dtype=float)
         tY = np.empty(200, dtype=float)

         tx_1 = np.random.randint(0, 450, 100)
         ty_1 = np.random.randint(1200, 1900, 100)

         tx_2 = np.random.randint(550, 1000, 100)
         ty_2 = np.random.randint(0, 700, 100)
```

```
        for i in range(100):
            tX[i] = tx_1[i]
            tX[100 + i] = tx_2[i]
            tY[i] = ty_1[i]
            tY[100 + i] = ty_2[i]

        plt.title('Testing Dataset')
        for x in range(200):
            if x < 100:
                plt.scatter(tX[x], tY[x], c='k', s=2)
            else:
                plt.scatter(tX[x], tY[x], c='y', s=2)
```



```
In [60]: plt.title('Training and Testing Dataset')

         for x in range(200):
             if x < 100:
                 plt.scatter(X[x], Y[x], c='r', s=2)
             else:
                 plt.scatter(X[x], Y[x], c='b', s=2)
         for x in range(200):
             if x < 100:
                 plt.scatter(tX[x], tY[x], c='k', s=2)
             else:
                 plt.scatter(tX[x], tY[x], c='y', s=2)
```


Training and Testing Dataset
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
1000

```
In [ ]:
```