Mark Mersnik
ID: 010455185

EE257 Homework 5

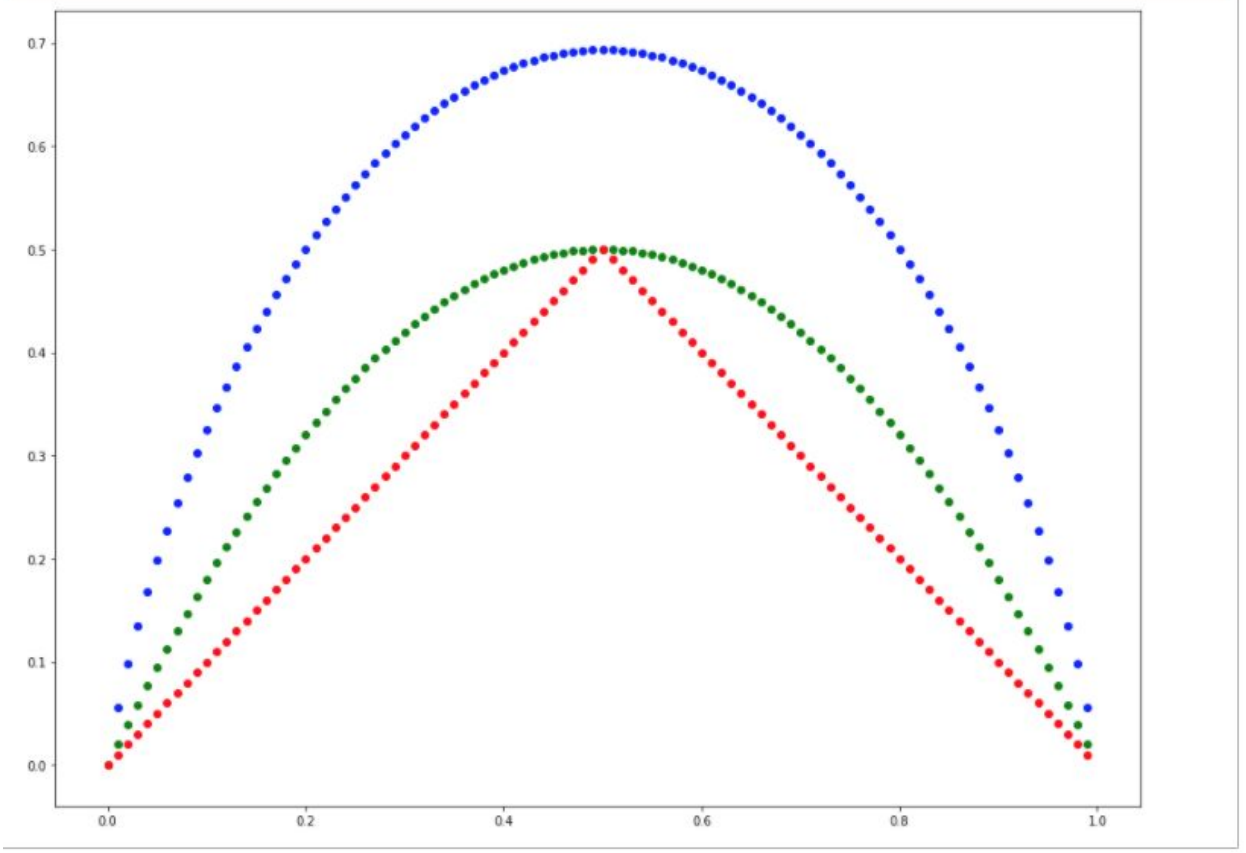
Problem 1:

a) Problem 3 (pg 332):

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

p = np.stack(i*0.01 for i in range(100))
```

```
gini = []
class_error = []
entropy = []

for i in p:
    gini.append(2*i*(1-i))
    class_error.append(1-max(i, 1-i))
    entropy.append((-i*np.log(i)-(1-i)*np.log(1-i)))
plt.figure(figsize=(16,12))
plt.scatter(p, gini, color='green')
plt.scatter(p, class_error, color='red')
plt.scatter(p, entropy, color='blue')
plt.show()
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:8: RuntimeWarning: divide by zero encountered in log

/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:8: RuntimeWarning: invalid value encountered in double_s
calars
```



b) Problem 5 on page 332/333

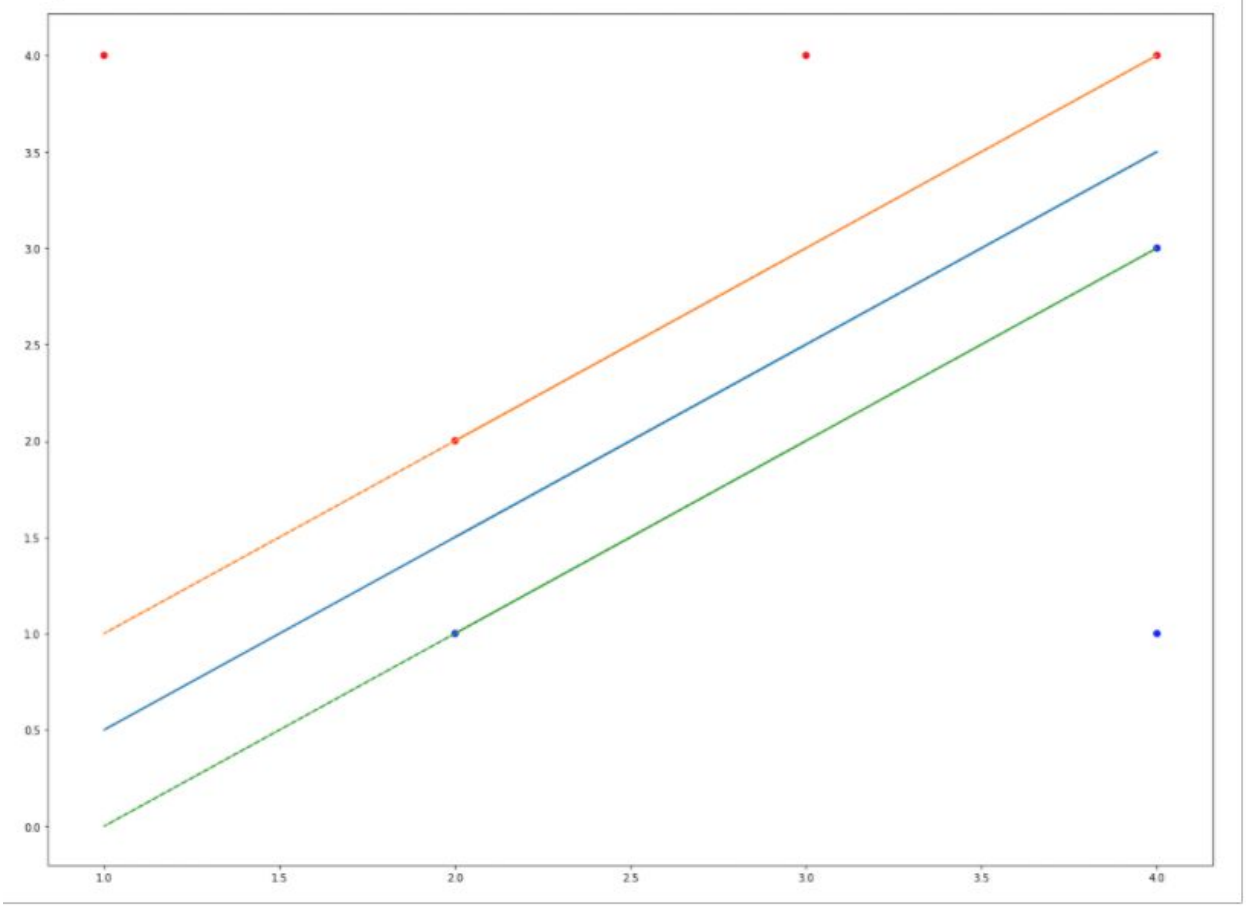
For the majority vote approach the final classification would be red since it is the majority of the dataset. For the average probability approach the final classification would be green since the average of the values is 0.45.

c) Problem 3 on page 368/369

a & b)

```
#Problem 3 on page 368/369
import matplotlib.lines as mlines

x1 = [3, 2, 4, 1, 2, 4, 4]
x2 = [4, 2, 4, 4, 1, 3, 1]
line = []
dashed = []
y = ['red', 'red', 'red', 'red', 'blue', 'blue', 'blue']
for i in x1:
    line.append(i-0.5)
    dashed.append(i-1)
plt.figure(figsize=(20,15))
plt.plot(x1, line)
plt.plot(x1,x1,'--')
plt.plot(x1, dashed, '--')
plt.scatter(x1, x2, c=y)
```

```
<matplotlib.collections.PathCollection at 0x128210450>
```

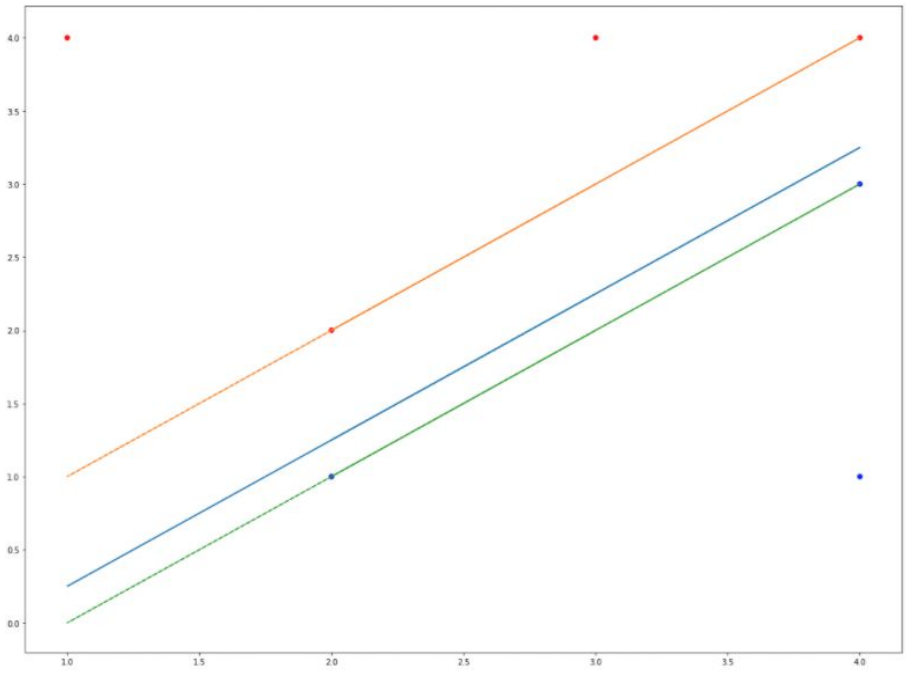


c) Equation of hyperplane: -0.5 + X1 - X2 = 0

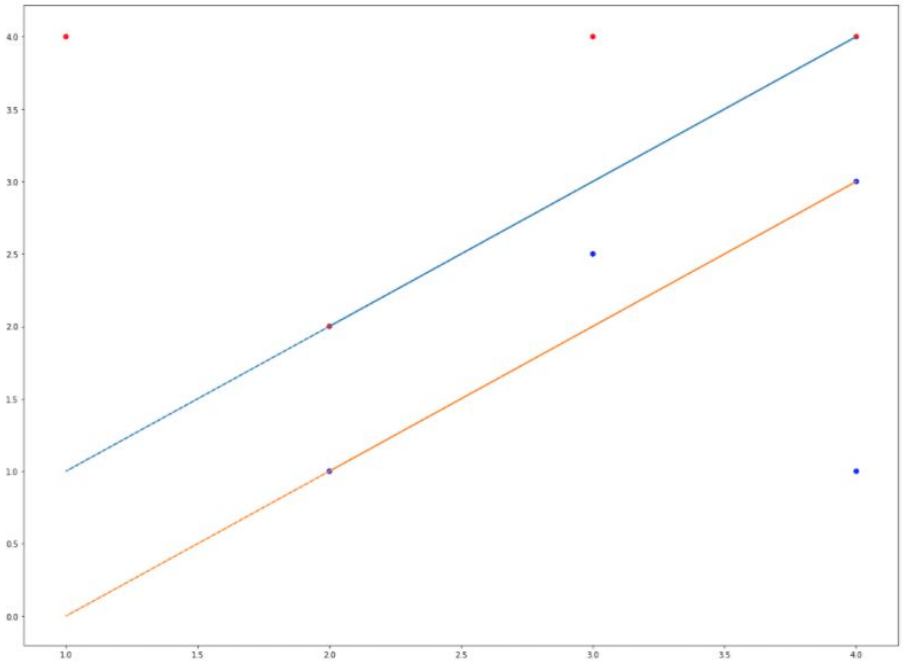
d) Margin for maximal margin hyperplane = 0.25

e) Support vectors: (2,1), (2,2), (4,3), (4,4)
f) A slight movement of (4,1) would not change the

g)



h)

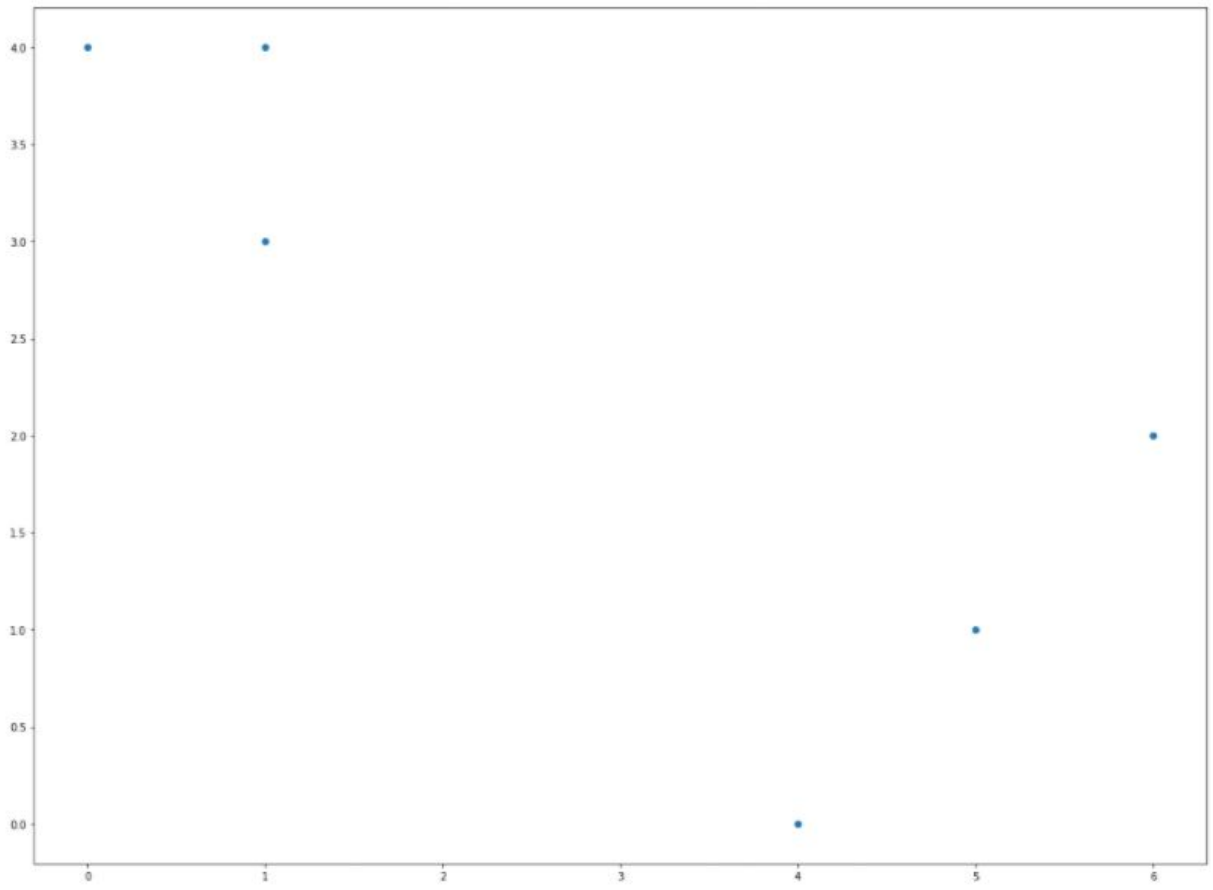


d) Problem 3 on page 414

a)

```
#Problem 3 on page 414
x1 = [1, 1, 0, 5, 6, 4]
x2 = [4, 3, 4, 1, 2, 0]

plt.figure(figsize=(20,15))
plt.scatter(x1, x2)
```

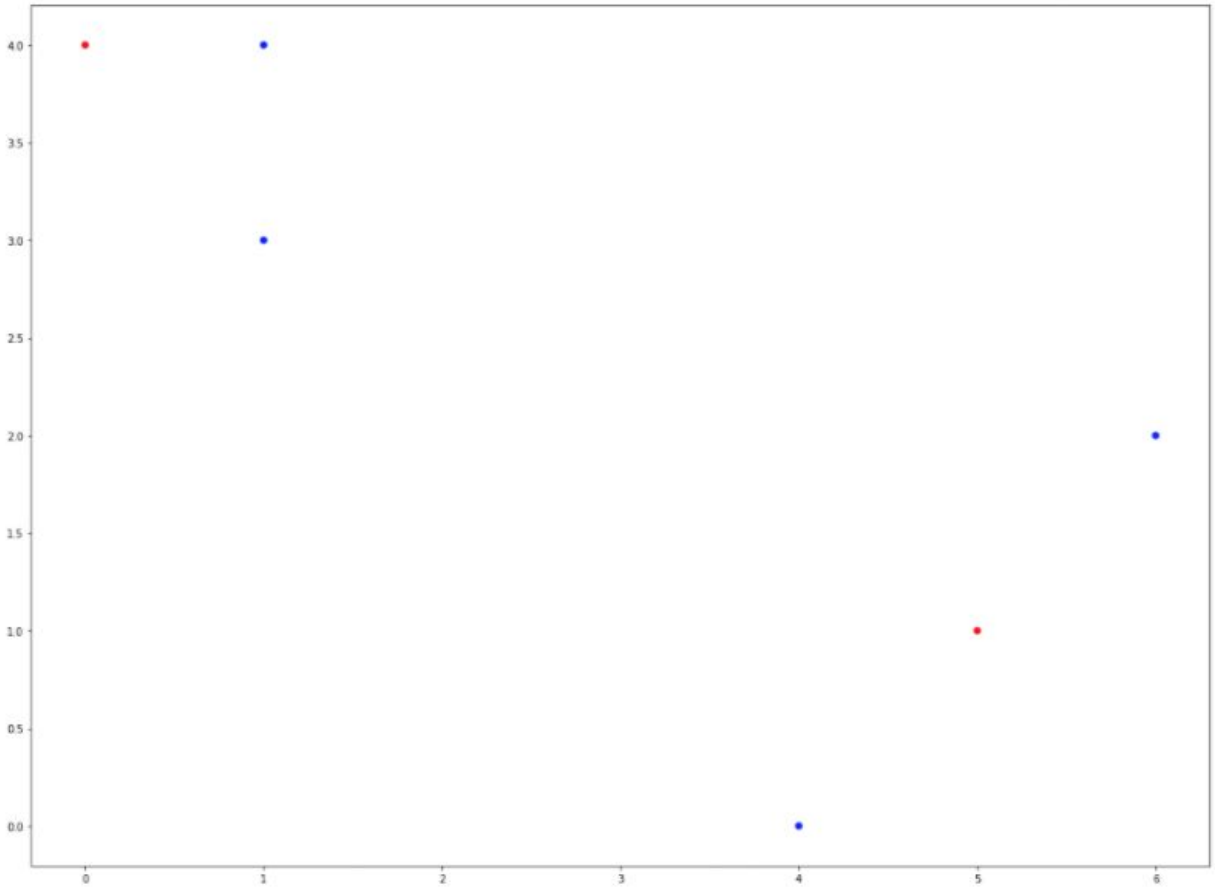
```
<matplotlib.collections.PathCollection at 0x127cc5710>
```



b)

```
#Problem 3 on page 414
x1 = [1, 1, 0, 5, 6, 4]
x2 = [4, 3, 4, 1, 2, 0]

#cluster label
df = pd.DataFrame({'colors': ['red', 'blue']})
y = df['colors'].sample(n=6, random_state=1, replace=True)

plt.figure(figsize=(20,15))
plt.scatter(x1, x2, c=y)
```

```
<matplotlib.collections.PathCollection at 0x12e162490>
```

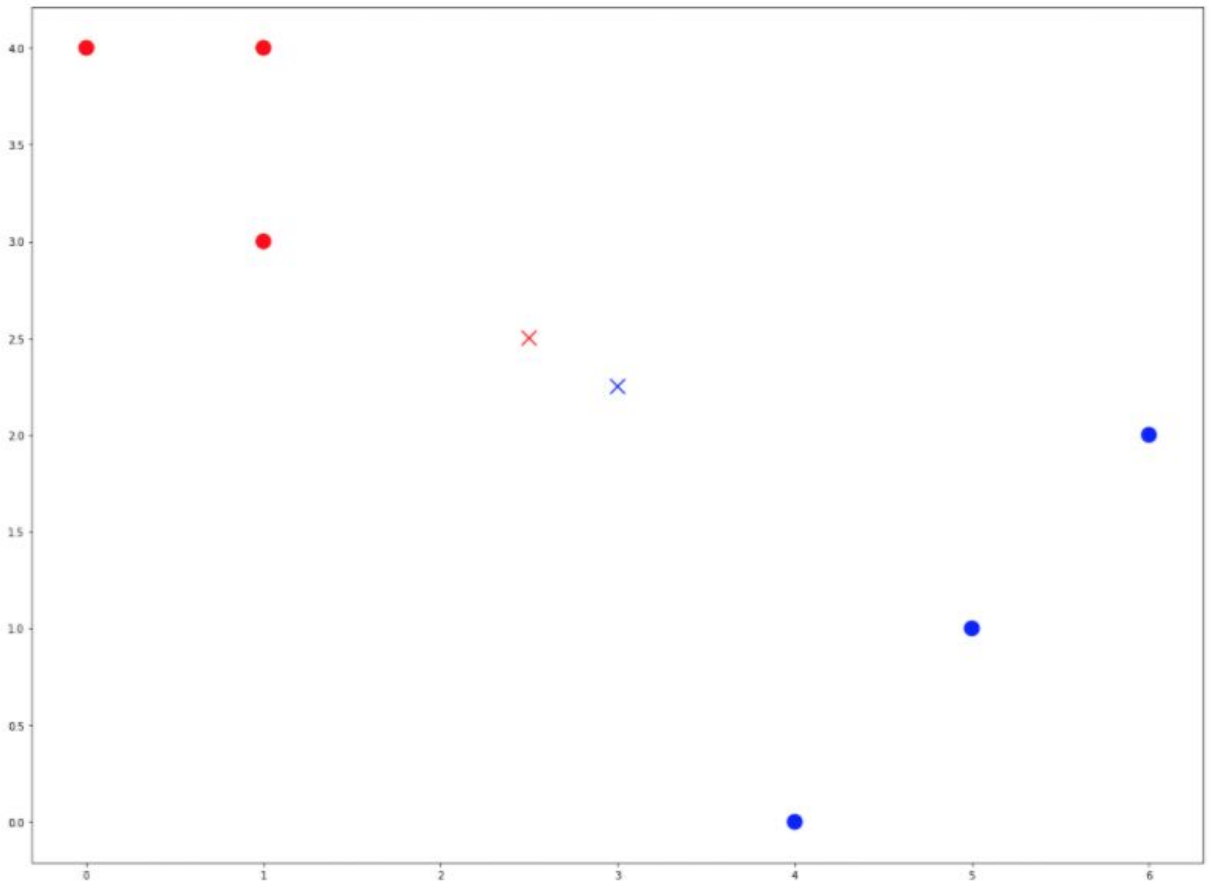


c) Centroid (Blue) : X11=¼(1+1+4+6)=3, X12=¼(4+3+2+0) = 2.25
Centroid (Red) : X21=½(0+5)=2.5, X22=½(1+4)=2.5

d)

```
#Problem 3 on page 414
x1 = [1, 1, 0, 5, 6, 4]
x2 = [4, 3, 4, 1, 2, 0]
y = ['red', 'red', 'red', 'blue', 'blue', 'blue']

cx1 = [3, 2.5]
cx2 = [2.25, 2.5]
c_colors = ['blue', 'red']

#cluster label
df = pd.DataFrame({'colors': ['red', 'blue']})
#y = df['colors'].sample(n=6, random_state=10, replace=True)

plt.figure(figsize=(20,15))
plt.scatter(x1, x2, s=200, c=y)
plt.scatter(cx1, cx2, s=200, c=c_colors, marker='x')
```

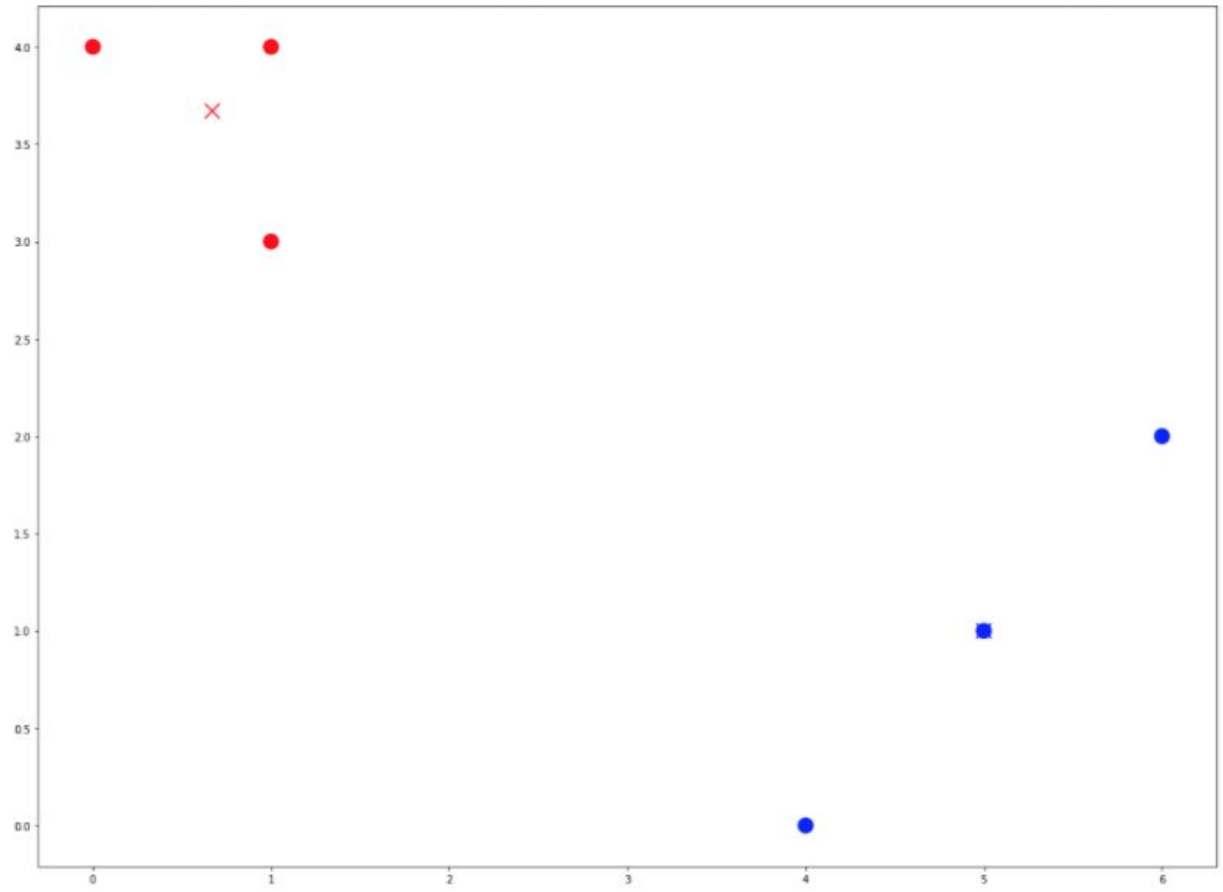
```
<matplotlib.collections.PathCollection at 0x13183ce90>
```



e & f) Centroid (Blue) : X11=⅓ (4+5+6)=5, X12=⅓(0+1+2) = 1
Centroid (Red) : X21=⅓(0+1+1)=0.67, X22=⅓(3+4+4)=3.67

```
#Problem 3 on page 414
x1 = [1, 1, 0, 5, 6, 4]
x2 = [4, 3, 4, 1, 2, 0]
y = ['red', 'red', 'red', 'blue', 'blue', 'blue']

cx1 = [5, 0.67]
cx2 = [1, 3.67]
c_colors = ['blue', 'red']

#cluster label
df = pd.DataFrame({'colors': ['red', 'blue']})
#y = df['colors'].sample(n=6, random_state=10, replace=True)

plt.figure(figsize=(20,15))
plt.scatter(x1, x2, s=200, c=y)
plt.scatter(cx1, cx2, s=200, c=c_colors, marker='x')
```

```
<matplotlib.collections.PathCollection at 0x1318ac9d0>
```



Problem 2:

a)

```
#Problem 9 page 334
from sklearn.model_selection import train_test_split

df = pd.read_csv('OJDataset.csv')
train_set = df[:800]
test_set = df[800:]
df.describe()
```

	WeekofPurchase	StoreID	PriceCH	PriceMM	DiscCH	DiscMM	SpecialCH	SpecialMM	LoyalCH	SalePriceMM	SalePriceCH
count	1070.000000	1070.000000	1070.000000	1070.000000	1070.000000	1070.000000	1070.000000	1070.000000	1070.000000	1070.000000	1070.000000
mean	254.381308	3.959813	1.867421	2.085411	0.051860	0.123364	0.147664	0.161682	0.565782	1.962047	1.815561
std	15.558286	2.308984	0.101970	0.134386	0.117474	0.213834	0.354932	0.368331	0.307843	0.252697	0.143384
min	227.000000	1.000000	1.690000	1.690000	0.000000	0.000000	0.000000	0.000000	0.000011	1.190000	1.390000
25%	240.000000	2.000000	1.790000	1.990000	0.000000	0.000000	0.000000	0.000000	0.325257	1.690000	1.750000
50%	257.000000	3.000000	1.860000	2.090000	0.000000	0.000000	0.000000	0.000000	0.600000	2.090000	1.860000
75%	268.000000	7.000000	1.990000	2.180000	0.000000	0.230000	0.000000	0.000000	0.850873	2.130000	1.890000
max	278.000000	7.000000	2.090000	2.290000	0.500000	0.800000	1.000000	1.000000	0.999947	2.290000	2.090000

b)