datascience_assignment_16.3

July 13, 2018

0.1 In a class on 100 students, 80 students passed in all subjects, 10 failed in one subject, 7 failed in two subjects and 3 failed in three subjects. Find the probability distribution of the variable for number of subjects a student from the given class has failed in.

0.2 Solution:

- Calculate Probability Mass Function (pmf) for number of subjects failed. If Number of subjects failed is 0, then probablility is 80/100 If Number of subjects failed is 1, probability is 10/100 If number of subjects failed is 2, probability is 7/100 If number of subjects failed is 2, probability is 3/100
- Create a dataframe df with two columns: no_of_failed_subjects which will have number of subjects failed probability which will be probability of number of subjects failed
- Plot the probability distribution using matplotlib

In [1]:
```
# Function pmf calculates probability mass function
# with no_of_failed_subjects as input and returns probability
def pmf(no_of_failed_subjects):
    if no_of_failed_subjects==0 :
        return 80/100
    elif no_of_failed_subjects==1:
        return 10/100
    elif no_of_failed_subjects==2:
        return 7/100
    elif no_of_failed_subjects==3:
        return 3/100
```

In [2]:
```
import numpy as np
import pandas as pd

# Create a dataframe df with two columns:
#  no_of_failed_subjects which will have number of subjects failed
# probability which will be probability of number of subjects failed
df = pd.DataFrame({'no_of_failed_subjects': np.arange(4)})
df['probability'] = df['no_of_failed_subjects'].apply(pmf)
df
```

Out[2]:

	no_of_failed_subjects	probability
0	0	0.80
1	1	0.10
2	2	0.07
3	3	0.03

In [3]:
```
# Plot the probability distribution using matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
fig, ax = plt.subplots(1, 1)
ax.plot(df['no_of_failed_subjects'], df['probability'], 'ro', ms=8, mec='r')
ax.vlines(df['no_of_failed_subjects'], 0, df['probability'], colors='r', linestyles='-
plt.title('Probability Distribution of Number of Subjects Failed')
plt.xlabel('Number of failed subjects')
plt.ylabel('Probability')
plt.show()
```