

# NB

November 20, 2019

## 1 Gaussian Naive Bayes

```
[7]: import numpy as np
import pandas as pd
import urllib.request as ur
import sklearn
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn import metrics
from sklearn.metrics import accuracy_score

url="http://archive.ics.uci.edu/ml/machine-learning-databases/spambase/spambase.
→data"

raw_data=ur.urlopen(url)

dataset=np.loadtxt(raw_data,delimiter=',')
print(dataset[0])
```

```
[ 0.      0.64    0.64    0.      0.32    0.      0.      0.      0.
 0.      0.      0.64    0.      0.      0.      0.32    0.      1.29
 1.93    0.      0.96    0.      0.      0.      0.      0.      0.
 0.      0.      0.      0.      0.      0.      0.      0.      0.
 0.      0.      0.      0.      0.      0.      0.      0.      0.
 0.      0.      0.      0.      0.      0.      0.778    0.      0.
 3.756  61.    278.    1.    ]
```

```
[12]: X=dataset[:,0:48]
y=dataset[:, -1]
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=.3,random_state=17)
BernNB=BernoulliNB(binrize=True)
BernNB.fit(X_train,y_train)
print(BernNB)

y_exp=y_test
y_pred=BernNB.predict(X_test)
```

```
print (accuracy_score(y_exp,y_pred))
```

```
BernoulliNB(alpha=1.0, binarize=True, class_prior=None, fit_prior=True)  
0.8616944243301955
```

```
[14]: MultiNB=MultinomialNB()  
MultiNB.fit(X_train,y_train)  
print(MultiNB)  
  
y_pred=MultiNB.predict(X_test)  
print(accuracy_score(y_exp,y_pred))
```

```
MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)  
0.8747284576393918
```

```
[15]: # Gaussian classifier  
GausNB=GaussianNB()  
GausNB.fit(X_train,y_train)  
print(GausNB)  
y_pred=GausNB.predict(X_test)  
print(accuracy_score(y_exp,y_pred))
```

```
GaussianNB(priors=None, var_smoothing=1e-09)  
0.8110065170166546
```

## 2 Analysis

**2.1** First we need to form a dataset containing documents with words and classes to which they belong. For a gaussian classifier, data taken considers the frequency of a word occurring in a document. In order to do this we preprocess the dataset with a preprocessing tool like Weka. After preprocessing, make use of GaussianNB() model on training data.