Bernoulli naive bayes cl_weka

November 20, 2019

1 Bernoulli Naive Bayes Classifier

[5]:
```
import pandas as pd
import numpy as np

w = pd.read_csv("C:/Users/kshitij/Desktop/Desktop/AML/CSV Files/bayes_weka.csv")

w
```

[5]:

	Class	Beijing	Chinese	Macau	Shanghai	Japan	Tokyo
0	c	1	1	0	0	0	0
1	c	0	1	0	1	0	0
2	c	0	1	1	0	0	0
3	j	0	1	0	0	1	1
4	j	0	1	0	0	1	1

[6]:
```
X = w.iloc[:,1:]
y = w.iloc[:,0]

X
```

[6]:

	Beijing	Chinese	Macau	Shanghai	Japan	Tokyo
0	1	1	0	0	0	0
1	0	1	0	1	0	0
2	0	1	1	0	0	0
3	0	1	0	0	1	1
4	0	1	0	0	1	1

[7]:
```
y
```

[7]:
```
0    c
1    c
2    c
3    j
4    j
Name: Class, dtype: object
```

[9]:
```
from sklearn.naive_bayes import BernoulliNB
#Using Bernoulli Naive Bayes Classifier
```

```
clf = BernoulliNB()
clf.fit(X,y)
```

[9]:
```
BernoulliNB(alpha=1.0, binarize=0.0, class_prior=None, fit_prior=True)
```

[10]:
```
y_pred = clf.predict(X)
from sklearn.metrics import accuracy_score
print("Accuracy Score : {0}".format(accuracy_score(y,y_pred)))
prediction = clf.predict(np.array([[1,1,1,1,1,1]]))
prediction
```

```
Accuracy Score : 1.0
```

[10]:
```
array(['j'], dtype='<U1')
```

2 Analysis

2.1 First we need to form a dataset containing documents with words and classes to which they belong.For a bernoulli classifier, we need data to be in the binomial or binary form.In order to do this we preprocess the dataset with a preprocessing tool like Weka. After preprocessing,make use of BernoulliNB() model on training data.