fake news detection

June 27, 2023

[19]:
```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import LinearSVC
```

[20]:
```
data = pd.read_csv("news.csv")
```

[21]:
```
data['fake']= data['label'].apply(lambda x: 0 if x == "REAL" else 1)
```

[22]:
```
data = data.drop("label",axis=1)
```

[23]:
```
X, y = data['text'], data['fake']
```

[24]:
```
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size = 0.2)
```

[25]:
```
vectorizer = TfidfVectorizer(stop_words = "english",max_df = 0.7)
X_train_vectorized = vectorizer.fit_transform(X_train)
X_test_vectorized = vectorizer.transform(X_test)
```

[26]:
```
clf = LinearSVC()
clf.fit(X_train_vectorized,y_train)
```

[26]:
```
LinearSVC()
```

[27]:
```
clf.score(X_test_vectorized, y_test)
```

[27]:
```
0.9408050513022889
```

[28]:
```
len(y_test)*0.9344
```

[28]:
```
1183.8848
```

[29]:
```
len(y_test)
```

[29]:
```
1267
```

[30]:
```
with open("mytest.txt", "w", encoding = "utf-8")as f:
    f.write(X_test.iloc[10])
```

[31]:
```
with open("mytest.txt","r",encoding="utf-8") as f:
    text = f.read()
```

[32]:
```
vectorized_text=vectorizer.transform([text])
```

[33]:
```
clf.predict(vectorized_text)
```

[33]:
```
array([1], dtype=int64)
```

[36]:
```
y_test.iloc[10]
```

[36]:
```
1
```

[ ]:
```

```