ps2-6-solve

May 15, 2021

1 (a)

code implements on src/p06_spam.py

[1]:

```
import numpy as np

from src import util
from src import svm

from src.p06_spam import create_dictionary, transform_text, \
                         fit_naive_bayes_model, \
                         predict_from_naive_bayes_model, \
                         get_top_five_naive_bayes_words, \
                         compute_best_svm_radius
```

[2]:

```
train_messages, train_labels = util.load_spam_dataset(
    'data/ds6_train.tsv')
val_messages, val_labels = util.load_spam_dataset(
    'data/ds6_val.tsv')
test_messages, test_labels = util.load_spam_dataset(
    'data/ds6_test.tsv')
```

[3]:

```
dictionary = create_dictionary(train_messages)
train_matrix = transform_text(train_messages, dictionary)
val_matrix = transform_text(val_messages, dictionary)
test_matrix = transform_text(test_messages, dictionary)
```

[4]:

```
util.write_json('src/output/p06_dictionary', dictionary)
np.savetxt('src/output/p06_sample_train_matrix',
    train_matrix[:100,:])
```

2 (b)

code implements on src/p06_spam.py

[5]:

```
naive_bayes_model = fit_naive_bayes_model(
    train_matrix, train_labels)
```

[6]:
```
naive_bayes_predictions = predict_from_naive_bayes_model(
    naive_bayes_model, test_matrix)
```

[7]:
```
naive_bayes_accuracy = np.mean(
    naive_bayes_predictions == test_labels)
```

[8]:
```
print('Naive Bayes had an accuracy of {} on the testing set'.
 ↪format(naive_bayes_accuracy))
```

```
Naive Bayes had an accuracy of 0.978494623655914 on the testing set
```

[9]:
```
np.savetxt('src/output/p06_naive_bayes_predictions',
    naive_bayes_predictions)
```

3 (c)

code implements on `src/p06_spam.py`

[10]:
```
top_5_words = get_top_five_naive_bayes_words(
    naive_bayes_model, dictionary)
```

[11]:
```
print('The top 5 indicative words for Naive Bayes are: ', top_5_words)
```

```
The top 5 indicative words for Naive Bayes are:  ['claim', 'won', 'prize',
'tone', 'urgent!']
```

[12]:
```
util.write_json('src/output/p06_top_indicative_words', top_5_words)
```

4 (d)

code implements on `src/p06_spam.py`

[13]:
```
optimal_radius = compute_best_svm_radius(
    train_matrix, train_labels, val_matrix, val_labels,
    [0.01, 0.1, 1, 10])
print('The optimal SVM radius was {}'.format(optimal_radius))
```

```
The optimal SVM radius was 0.1
```

[14]:
```
util.write_json('src/output/p06_optimal_radius', optimal_radius)
```

[15]:
```
svm_predictions = svm.train_and_predict_svm(
    train_matrix, train_labels, test_matrix, optimal_radius)
svm_accuracy = np.mean(svm_predictions == test_labels)
print('The SVM model had an accuracy of {} on the testing set'.
 ↪format(svm_accuracy))
```

```
The SVM model had an accuracy of 0.9695340501792115 on the testing set
```

[ ]:

```
```