

README file

Jiangshan Luo
U89971259

1. Code File

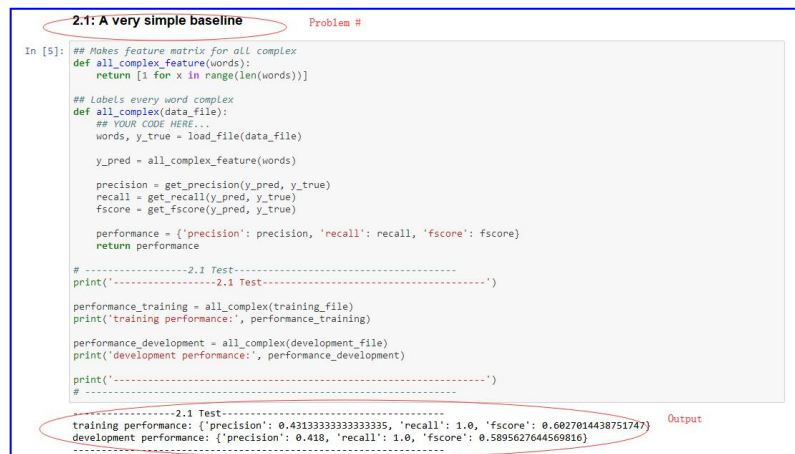
The code file contains workable code and data that it needs.

There are **3 different types of the code** (Jupyter Notebook, python source code, HTML). Please pick whichever you can run. Note that the outputs can be easily visualized in Jupyter Notebook and HTML, so it might be better to view those two.

There is also **data files** which contains training and validation data.

2. Source Code Clarification

The question **titles** are marked about each code block, and the **results** of them are right below. And also, the order of the questions are the same as which in the assignment document.



```
2.1: A very simple baseline Problem #
In [5]: ## Makes feature matrix for all complex
def all_complex_feature(words):
    return [1 for x in range(len(words))]

## Labels every word complex
def all_complex(data_file):
    ## YOUR CODE HERE...
    words, y_true = load_file(data_file)
    y_pred = all_complex_feature(words)

    precision = get_precision(y_pred, y_true)
    recall = get_recall(y_pred, y_true)
    fscore = get_fscore(y_pred, y_true)

    performance = {'precision': precision, 'recall': recall, 'fscore': fscore}
    return performance

# -----2.1 Test-----
print('-----2.1 Test-----')

performance_training = all_complex(training_file)
print('training performance:', performance_training)

performance_development = all_complex(development_file)
print('development performance:', performance_development)

print('-----2.1 Test-----')
# -----2.1 Test-----
training performance: {'precision': 0.4313333333333333, 'recall': 1.0, 'fscore': 0.6027014438751747}
development performance: {'precision': 0.418, 'recall': 1.0, 'fscore': 0.5895627644569816}
```

Image is borrowed from HW1

3. Clarification for output

There will be outputs for almost each of the cell in jupyter notebook. Those are the testing or evaluation parts for the functions. And you can still find the original output in the main program cell. Under the main program cell are the code for the discussion parts.