Introduction to Machine Learning in Engineering Science

National Cheng Kung University Department of Engineering Science

Instructor: Chi-Hua Yu

Lab 6

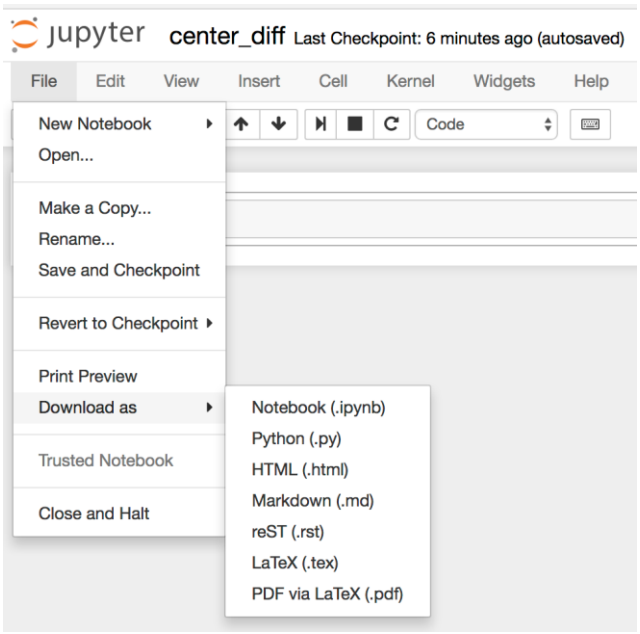
Programming, Due 11:55 am, Saturday, November 27rd, 2021

Submit by 08:00pm on 11/24 will receive a 20% bonus. Late submission before post of solution: score*0.8 (the solution will usually be posted within a week); no late submission after the post of solution

---

Lab Submission Procedure (請仔細閱讀)

1. You should submit your Jupyter notebook and Python script (`*.py`, in Jupyter, click File, Download as, Python (`*.py`)).



2. Name a folder using your student id and lab number (e.g., n96081494_lab1), put all the python scripts into the folder and zip the folder (e.g., n96081494_ lab 1.zip).

3. Submit your lab directly through the course website.

Total 100%

1. (100%) **Please download the zip file lab6.zip from Moodle.** Name your Jupyter notebook `RNN_SPAM` and Python script `RNN_SPAM.py`. Please create an RNN model to classify spam. The dataset SPAM.zip can be downloaded from Moodle.

   Data preprocessing is to convert the text in Message to a vector. You can use the function `Tokenizer` (`from keras.preprocessing.text import Tokenizer`) to complete.

Message

Text: `'Received, understood n acted upon!'`

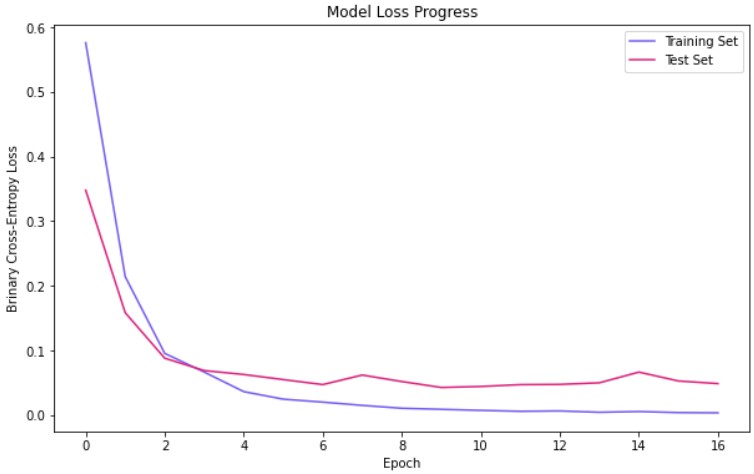
↓

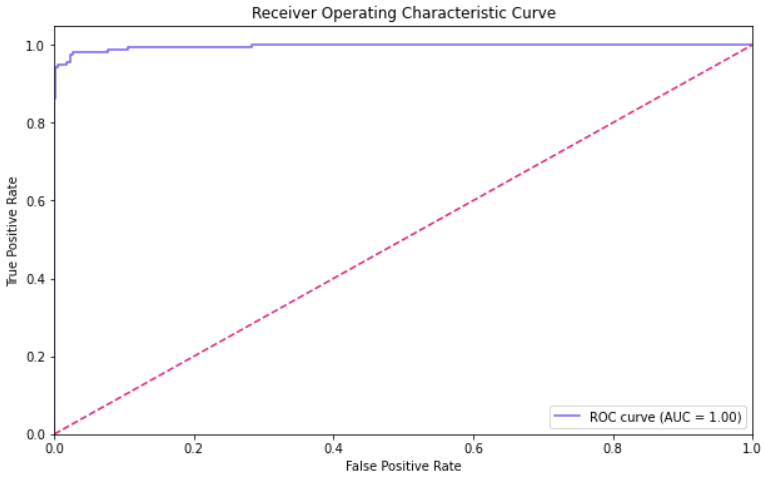
Vector: `[1903. 3620. 98. 2462. 3621.]`

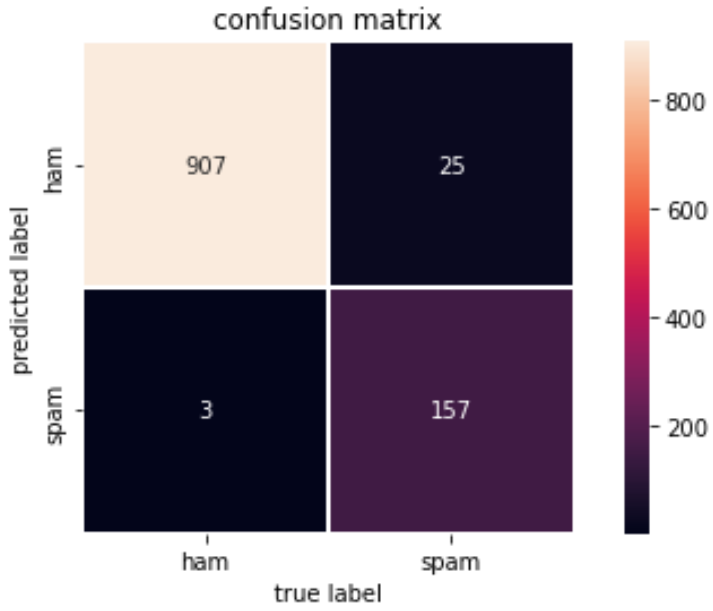
After converting the text to a vector, use `sequence.pad_sequences` (`from tensorflow.keras.preprocessing import sequence`) to fill the length of all sequences to the same length.
Below is the shape of data after pre-processing:

```
x_train shape: (4367, 200)
x_test shape: (1092, 200)
```

Use the pre-processed data to train the model, please report training history, roc curve, confusion matrix and `balanced_accuracy_score`.








```
balanced_accuracy_score: 0.975348712446352
```