LSTM

Recurrent NN - RNN

Internal State

h

x_t

A

h_0 h_1 h_2 h_3 h_t

A → A → A → A → ... → A

x_0 x_1 x_2 x_3 x_t

tanh

h x_t

h_{t-1}

h_t

A

tanh
[-1, 1]

x_{t-1}

x_t

input

C_{t-1}

h_{t-1}

x_t

h_t

c_t

tanh

tanh

σ   σ   σ

(0,1)

output

สัดส่วนของ output จาก node ก่อนหน้า

สัดส่วนที่เสริมจาก input ปัจจุบัน

บวก h ก่อนหน้า

x4
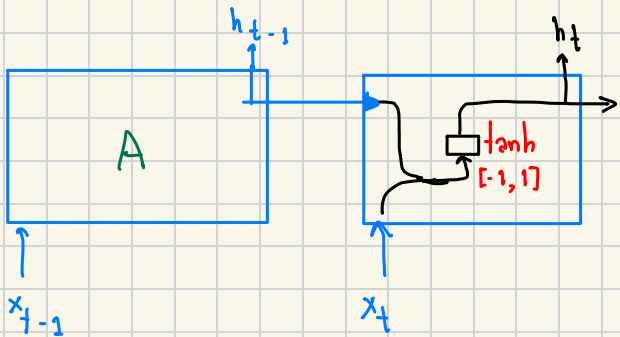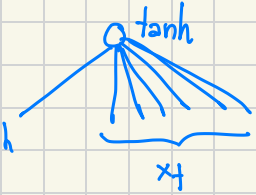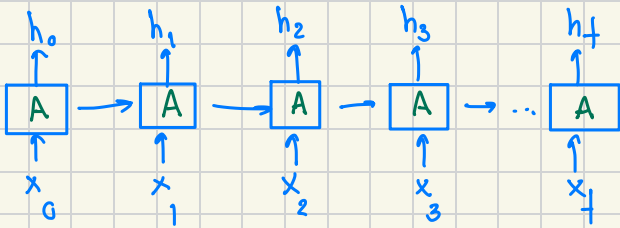
Text classification with LSTM Source code

```
1 import pandas as pd
2
3 df=pd.read_csv('text_data.csv')
4 df['class'].value_counts()
```
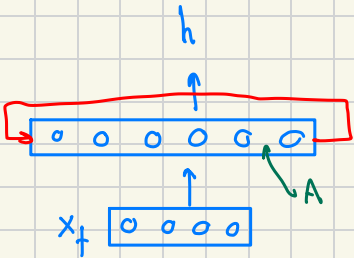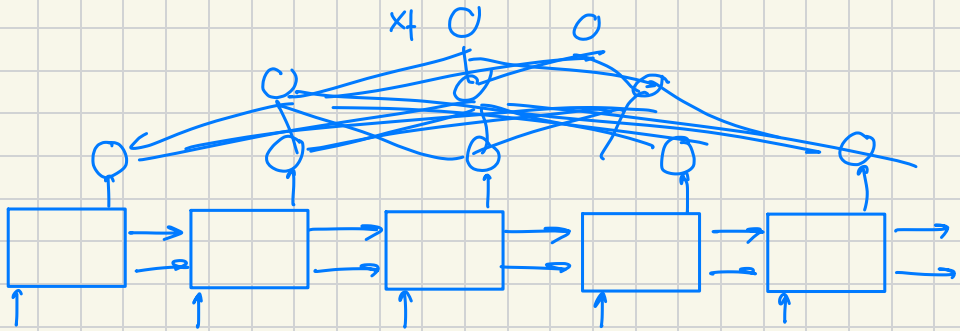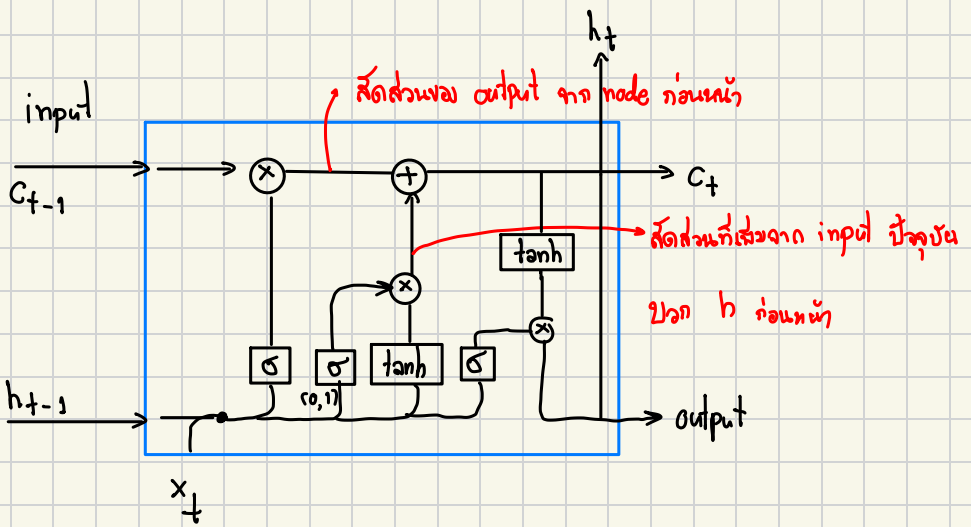
```
punpromotion.csv    2302
CH3Thailand.csv     1934
ejan2016.csv         319
Name: class, dtype: int64
```

```
1 df['word_length'] = df['message'].str.split()
2 df['word_length'] = df['word_length'].str.len()
3 df.dropna(inplace=True)                    I
```

```
1 df['word_length'].sort_values(ascending=False)
```

```
2574    355.0
4089    337.0
2113    282.0
2021    233.0
1943    219.0
```

```
1 df.message.values
```

```
array([' อย่า โกรธ แม่ ขาย 😊 😊 ดู ย้อนหลัง ที่นี่',
       ' แอคชั่น เต็ม แถม ซึ้ง จาก หาวน ดู เบื้องหลัง ลิขิต จันทร์ ฟัง นักแสดง เล่าเรื่อง สนุก กอง ถ่าย ติดตาม กอง วิก  ลิขิต จันทร์ วันพรุ่งนี้ เวลา 09.15 น. ',
       ' ร้องให้ ไหว ทด กลายเป็น ทาส เด็ก ดู ย้อนหลัง ที่นี่', ...,
       ' บัตรเครดิต ส่วนลด ไม่ต้อง เสียใจ ปัน โปร >> ขั้นตอน ปัน โปร <<  รายละเอียด สถานที่ ส่วนลด บัตรเครดิต ยอด ชำระเงิน facebook message >> รายละเอียด แบ่งปัน
         สถานที่ Siam Paragon Siam Center Siam SQ 1, CentralWorld  เวลา ศุกร์ เสาร์ อาทิตย์ เวลา 11.00 น.  21.00  ชำระ ค่า อาหาร สินค้า ส่วนลด เงินสด ค่าธรรมเนียม
         แบ่งปัน >> หมายเหตุ <<  แบ่งปัน ส่วนลด การแลก คะแนน  สงวนสิทธิ์ การเปลี่ยนแปลง แบ่งปัน ส่วนลด แจ้งให้ทราบ ล่วงหน้า',
       ' เมื่อ ใหม่ พลาด ส่วนลด บัตรเครดิต ปัน โปร',
       ' แบ่งปัน ส่วนลด โปรโมชั่น บัตรเครดิต ร้านอาหาร ร้านค้า เพื่อน ติดตาม รายละเอียด ส่วนลด ช่วงเวลา สถานที่ แบ่งปัน รายสัปดาห์ ที่นี่ ค่าธรรมเนียม แบ่งปัน ^^'],
       dtype=object)
```

```
1  from keras.preprocessing.text import Tokenizer
2  from sklearn.model_selection import train_test_split
3  from keras.utils import pad_sequences
4  from keras.models import Sequential
5  from keras.layers import Dense
6  from keras.layers import Flatten
7  from keras.layers import LSTM
8  from keras.layers import Embedding
9  from keras.layers import SpatialDropout1D
10
11 MAX_WORDS = 2500
12 MAX_SEQUENCE_LENGTH = 355    ความยาวสูงสุดของ Post
13 EMBEDDING_DIM = 100
15 tokenizer = Tokenizer(num_words=MAX_WORDS, filters='!"#$%&()*+,-./:;<=>?@[\]^_`{|}~')
16 tokenizer.fit_on_texts(df.message.values)
17 word_index = tokenizer.word_index
18 X = tokenizer.texts_to_sequences(df.message.values)
```

```
1  df.message.values[0]
```

' อย่า โกรธ แม่ ขาย 😡 😡 ดู ย้อนหลัง ที่นี่ '

' อย่า โกรธ แม่ ขาย 😡 😡 ดู ย้อนหลัง ที่นี่ ' 🗗

```
1  word_index
```

convert to integer

```
1  X[0]
```

[52, 2129, 115, 108, 401, 1, 22, 102]

```
1  X = pad_sequences(X, maxlen=MAX_SEQUENCE_LENGTH)
2  Y = pd.get_dummies(df['class']).values    ⟶ ออกเป็น 3 node
3  X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.10, random_state = 42)
```

```
1  X[0]
```

```
array([  0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
         0,    0,    0,    0,    0,    0,   52, 2129,  115,  108,  401,
         1,   22,  102], dtype=int32)
```

```
1  model = Sequential()
2  model.add(Embedding(MAX_WORDS, EMBEDDING_DIM, input_length=X_train.shape[1]))
3  model.add(SpatialDropout1D(0.2))
4  model.add(LSTM(100, dropout=0.2, recurrent_dropout=0.2))
5  model.add(Dense(3, activation='softmax'))   activate ตัวที่ Prob สูงสุด
6  model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
7
8  epochs = 5
9  batch_size = 64
10
11 history = model.fit(X_train, Y_train, epochs=epochs, batch_size=batch_size,validation_split=0.1)
```

```
1 posts=['เชิญ ดู ละคร ย้อนหลัง วันนี้ ทุ่ม','อย่า พลาด อาหาร เด็ด ราคา โดน']
2 X_new_test = tokenizer.texts_to_sequences(posts)
3 X_new_test = pad_sequences(X_new_test, maxlen=MAX_SEQUENCE_LENGTH)
4 results = model.predict(X_new_test)

1/1 [==============================] - 0s 321ms/step

1 results

array([[9.96929944e-01, 7.16787355e-04, 2.35336041e-03],
       [1.00705385e-01, 9.06365365e-02, 8.08658063e-01]], dtype=float32)
```

```
df['class'].unique()   CH3Thailand.csv', 'ejan2016.csv', 'punpromotion.csv']
```

LSTM   Regression

```
1 X_train[:5]

array([[0.          ],
       [0.00117376],      ชุดอันนึ่
       [0.00213016],
       [0.00278225],
       [0.00691214]])


1 Y_train[:5]

array([[0.00117376],
       [0.00213016],     next day
       [0.00278225],
       [0.00691214],
       [0.00965092]])
```

```
3 from keras.layers import SpatialDropout1D
4 from tensorflow.keras import activations
5
6 model = Sequential()
7 model.add(LSTM(16, input_shape=(1,1)))
8 model.add(Dense(4))
9 model.add(Dense(1))   I
10 model.compile(loss='mean_squared_error', optimizer='adam')
```
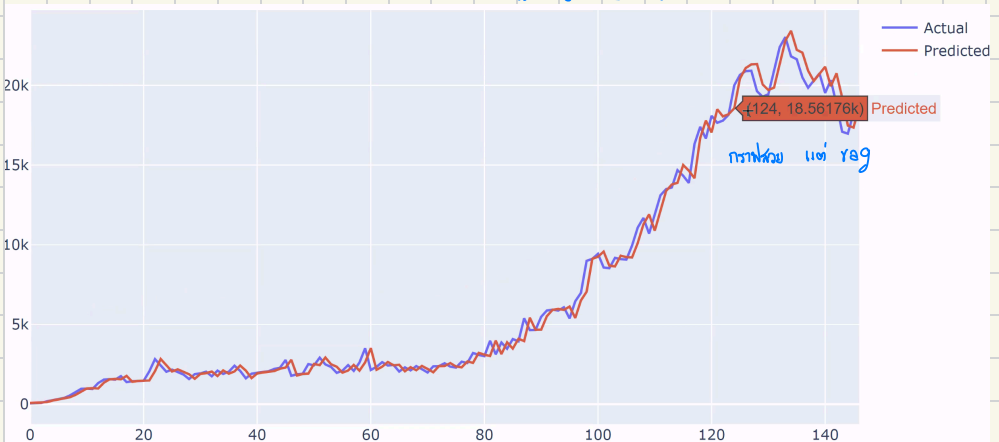
```
1 epochs = 100
2 batch_size = 5
3
4 history = model.fit(X_train, Y_train, epochs=epochs, batch_size=batch_size,validation_split=0.1)
5
```

กราฟกรณี 1 - 1   ความแปลงเป็น m - 1


```
(124, 18.56176k)  Predicted
```

Actual
Predicted

กราฟสวย แต่ rag

```
1 X_train.shape

(143, 1, 5) — 5 วัน
```

ถ้าเป็น > 1 คือสี attribute อื่น ๆ นั่น
อุณหภูมิ กับความชื้น