Confused student EEG

許宇韶 110034563

目錄

背景介紹

現況描述

每個大學生都難免會面臨課業壓力、人際壓力、感情壓力或家庭的壓力，通常課業壓力被認為是大學生的主要壓力來源，不同科系面臨的課業壓力可能有異。其中包含了因為無法適應教師的上課方式與不敢詢問教師有關課業上的問題而造成學習障礙；而學習方法欠佳，不知如何計劃，還有學習能力的不足，則容易造成挫折感。

5W1H

WHY

協助對於學習有困難的學生從而落實快樂學習?

WHAT

困惑學生的腦電圖預測

WHO

學生

WHEN

學習時間

WHERE

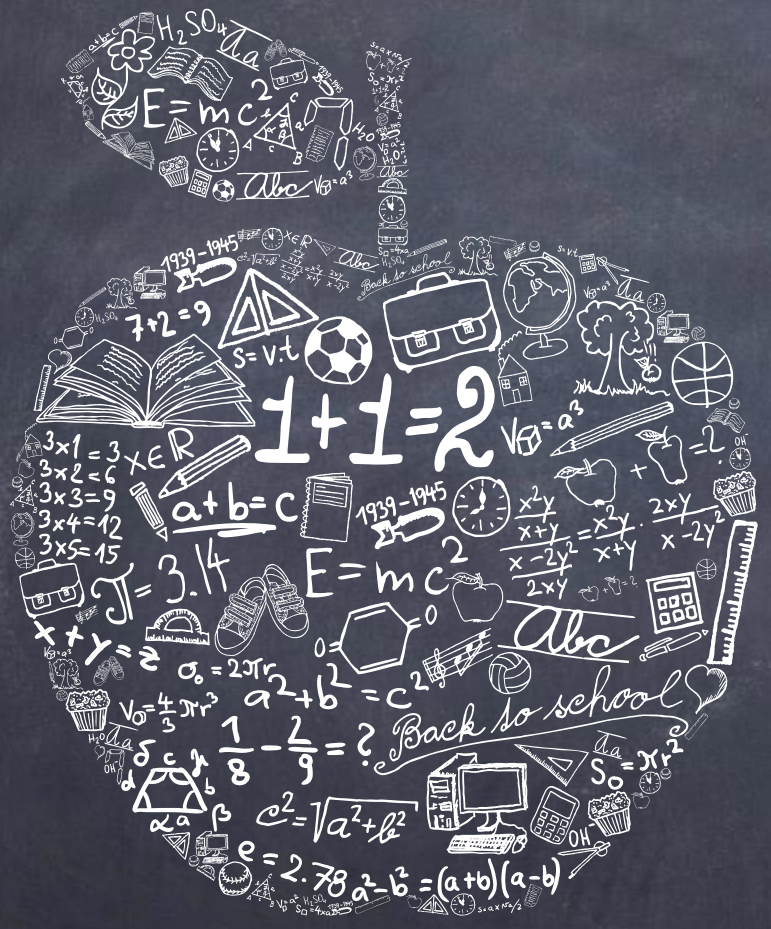
教室

HOW

LSTM、LogisticRegression

資料介紹及處理

資料內容



由Kaggle公開數據集中取得Confused Student EEG prediction資料集，資料集共分成2種類別(0&1)：

1. 數據共有12000筆，18項特徵
2. 目標應變數為：1為困惑、0為無明顯反應
3. 特徵

SubjectID	float64
VideoID	float64
Attention	float64
Mediation	float64
Raw	float64
Delta	float64
Theta	float64
Alpha1	float64
Alpha2	float64
Beta1	float64
Beta2	float64
Gamma1	float64
Gamma2	float64
predefinedlabel	float64
user-definedlabeln	float64
age	int64
ethnicity	object
gender	object

資料處理

One hot encoding

Standard Scaler

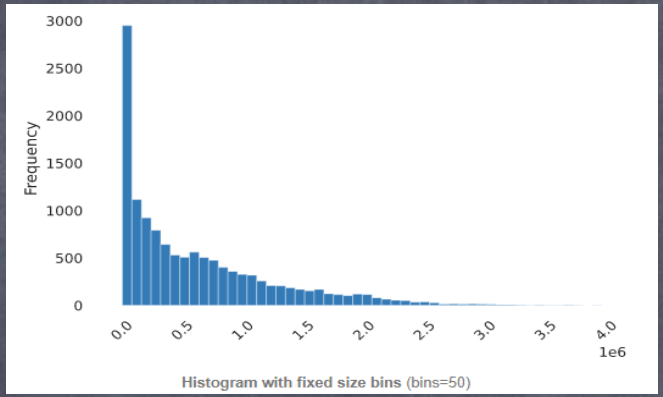
檢查缺失值

資料視覺化

Delta

Quantile statistics	
Minimum	448
5-th percentile	6086.5
Q1	98064
median	395487
Q3	916623
95-th percentile	1945866
Maximum	3964663
Range	3964215
Interquartile range (IQR)	818559

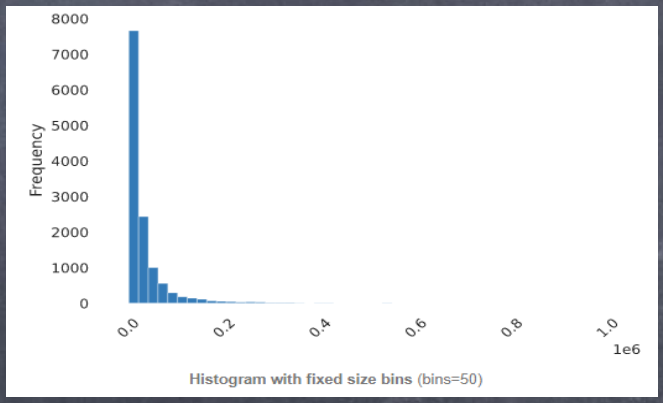
Descriptive statistics	
Standard deviation	637623.5626
Coefficient of variation (CV)	1.052557074
Kurtosis	1.84600758
Mean	605785.2617
Median Absolute Deviation (MAD)	355506
Skewness	1.395603016
Sum	7760714988
Variance	$4.065638076 \times 10^{11}$
Monotonicity	Not monotonic



Alpha2

Quantile statistics	
Minimum	2
5-th percentile	2264
Q1	6852
median	14959
Q3	34550.5
95-th percentile	125765
Maximum	1016913
Range	1016911
Interquartile range (IQR)	27698.5

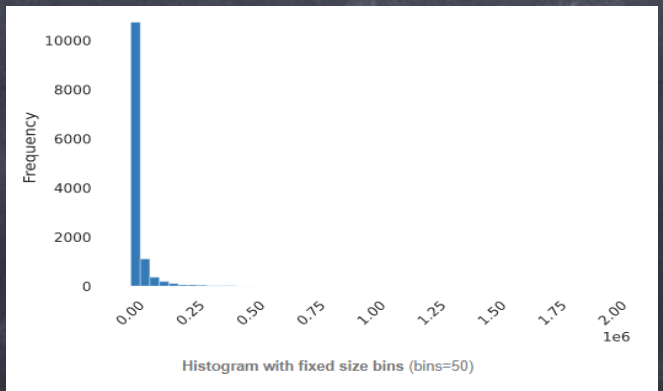
Descriptive statistics	
Standard deviation	58314.10075
Coefficient of variation (CV)	1.757327843
Kurtosis	50.35553589
Mean	33183.39318
Median Absolute Deviation (MAD)	10120
Skewness	5.719044311
Sum	425112450
Variance	3400534346
Monotonicity	Not monotonic



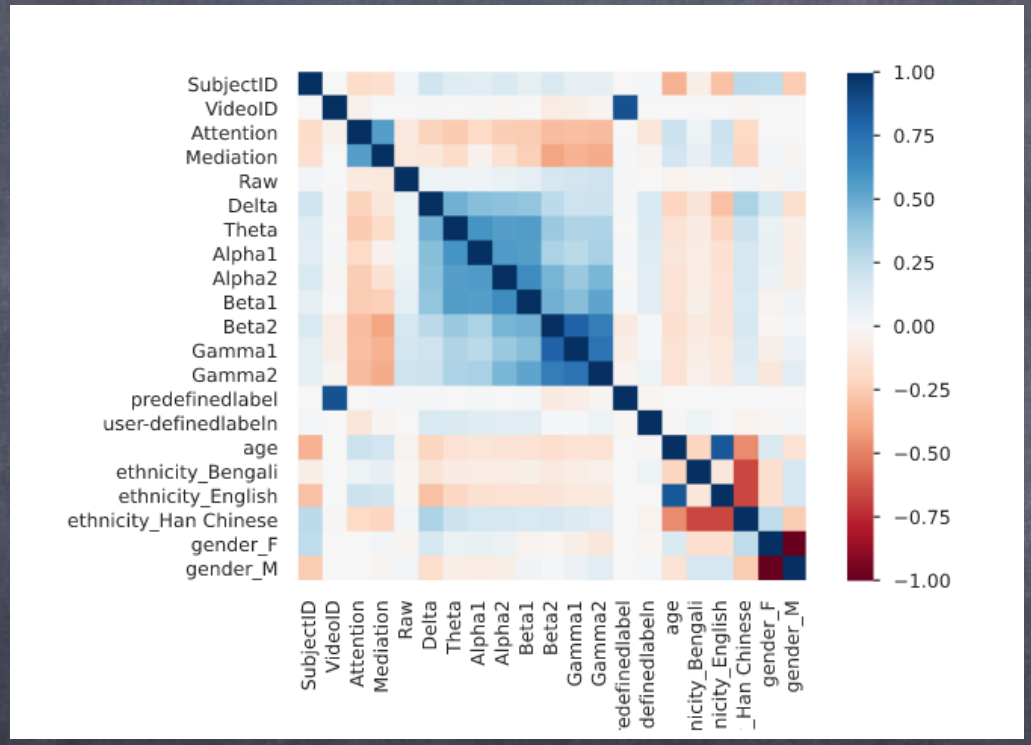
Gamma1

Quantile statistics	
Minimum	1
5-th percentile	1312.5
Q1	4058
median	9763
Q3	24888
95-th percentile	111699.5
Maximum	1972506
Range	1972505
Interquartile range (IQR)	20830

Descriptive statistics	
Standard deviation	79826.36692
Coefficient of variation (CV)	2.6975154
Kurtosis	181.8814538
Mean	29592.55281
Median Absolute Deviation (MAD)	6980
Skewness	10.89715456
Sum	379110194
Variance	6372248856
Monotonicity	Not monotonic



資料視覺化



- VideoID & Predefinedlabel
- Beta2 & Gamma1
- Gamma1 & Gamma2

特徵篩選

Index	0
VideoID	6.77143
Alpha2	2.54267
Delta	2.44771
Gamma1	2.29428
Theta	2.14002
Beta1	2.09874
Alpha1	2.03632
Gamma2	1.88297
Raw	1.57544
Beta2	1.43891
Attention	1.06214
Mediation	0.543061
predefinedlabel	0.509191
SubjectID	0.381945
age	0.195931
ethnicity_Bengali	0
ethnicity_English	0
ethnicity_Han Chinese	0
gender_F	0
gender_M	0

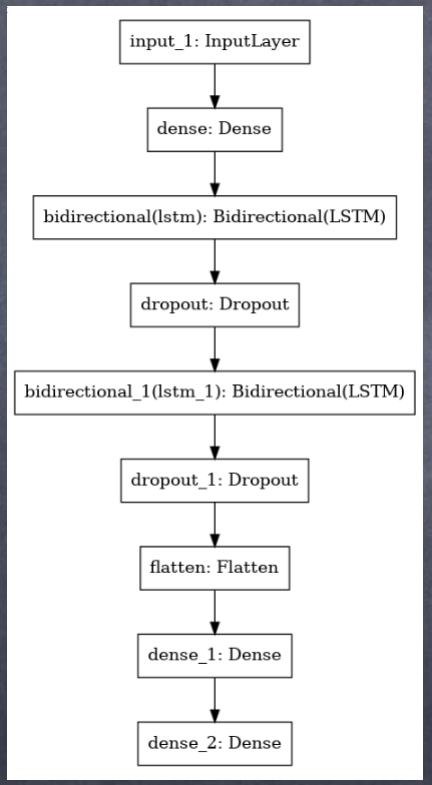
```
from sklearn.feature_selection import mutual_info_classif
mi_score=mutual_info_classif(data.drop('user-definedlabeln',axis=1),data['user-definedlabeln'])

mi_score=pd.Series(mi_score,index=data.drop('user-definedlabeln',axis=1).columns)

mi_score=(mi_score*100).sort_values(ascending=False)
```

- Drop SubjectID & VideoID
- Drop age & gender & ethnicity
- Drop SubjectID & VideoID & age & gender & ethnicity

模型選擇

Logistic Regression & LSTM

```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 17, 1)]           0
_________________________________________________________________
dense (Dense)                (None, 17, 64)            128
_________________________________________________________________
bidirectional (Bidirectional (None, 17, 512)           657408
_________________________________________________________________
dropout (Dropout)            (None, 17, 512)           0
_________________________________________________________________
bidirectional_1 (Bidirection (None, 17, 256)           656384
_________________________________________________________________
dropout_1 (Dropout)          (None, 17, 256)           0
_________________________________________________________________
flatten (Flatten)            (None, 4352)              0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               557184
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 129
=================================================================
Total params: 1,871,233
Trainable params: 1,871,233
Non-trainable params: 0
_________________________________________________________________
```



訓練過程

Logistic Regression

Drop SubjectID & VideoID

ACC : 0.5298478345688646

Drop age & gender & ethnicity

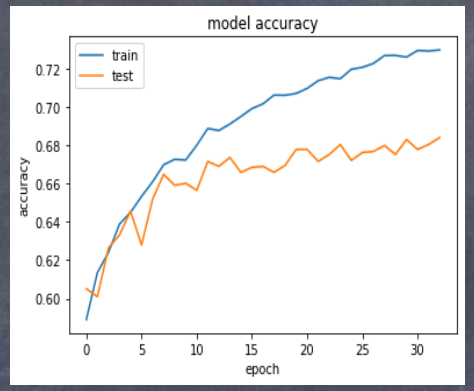
ACC : 0.5306281701131487

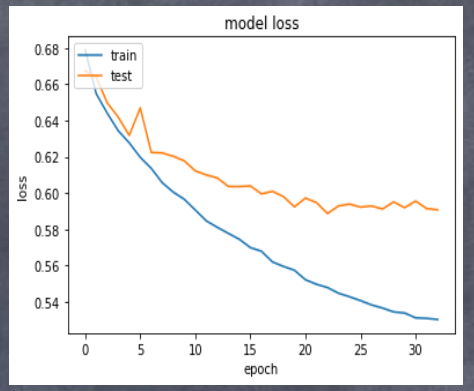
Drop SubjectID & VideoID age & gender & ethnicity

ACC : 0.5645727662895045

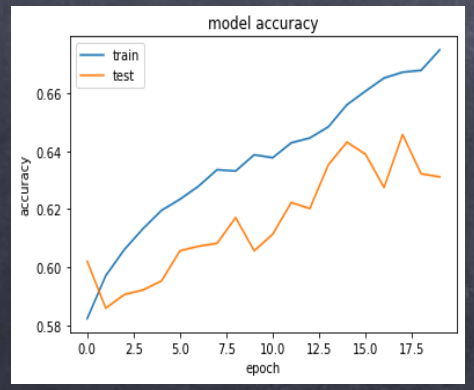
LSTM

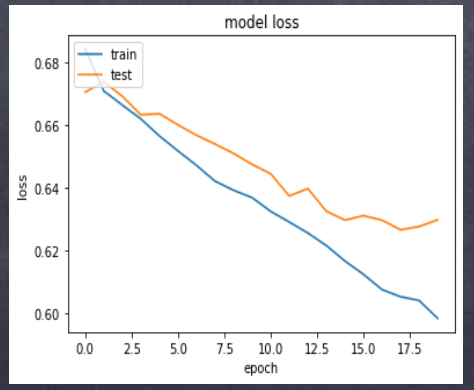
Drop SubjectID & VideoID





Drop SubjectID & VideoID & age & gender & ethnicity





LSTM

Drop SubjectID & VideoID age & gender & ethnicity

	precision	recall	f1-score	support
0.0	0.99	1.00	1.00	974
1.0	1.00	0.99	1.00	948
accuracy			1.00	1922
macro avg	1.00	1.00	1.00	1922
weighted avg	1.00	1.00	1.00	1922

超參數調整

使用的損失函數 Binary Cross Entropy，並使用 Early_Stopping 的回調函數來避免過度擬合，並使用 lr_scheduler 來改變模型訓練時的學習率。從 learning_rate = 0.001 和 batch_size = 20 開始訓練 100 個 epochs。

Q&A