



# Project Summary

Github: <https://github.com/richwellp/CS547>



# Project Overview

## Project 22: UIUC GPA

### Team Members

- Yiping Li - [yipingl4@illinois.edu](mailto:yipingl4@illinois.edu)
- Leo Yang - [junjey3@illinois.edu](mailto:junjey3@illinois.edu)
- Shijie Sun - [shijies5@illinois.edu](mailto:shijies5@illinois.edu)
- Richwell Perez - [richwell@illinois.edu](mailto:richwell@illinois.edu)

### Problem Summary

The purpose of this project is to implement deep learning concepts and techniques on a real dataset: UIUC GPA. The general questions that will require the application of deep learning is predicting the GPA/grade distribution of UIUC courses in the future. The project will provide some visualization of the data and descriptive statistics, implement linear or logistic regression, and recurrent neural networks.

### License

Dataset is obtained from Professor Ulmschneider's [uiuc-gpa-dataset](#). Project curated by Jared Canty (Summer 2022 Blackwell Program). All rights are reserved.



# Milestone 1

```
[ ] file_url = "https://raw.githubusercontent.com/wadefagen/datasets/master/gpa/uiuc-gpa-dataset.csv"

gpa_data = pd.read_csv(file_url, header=0)
gpa_data

[ ] gpa_scale = {
    'A+' : 4.0,
    'A'  : 4.0,
    'A-' : 3.67,
    'B+' : 3.33,
    'B'  : 3.0,
    'B-' : 2.67,
    'C+' : 2.33,
    'C'  : 2.0,
    'C-' : 1.67,
    'D+' : 1.33,
    'D'  : 1.0,
    'D-' : 0.67,
    'F'  : 0.0,
} # defined from https://registrar.illinois.edu/courses-grades/explanation-of-grades/

letterGrades = list(gpa_scale.keys())
gpa_data['Students_Completed'] = gpa_data[letterGrades].sum(axis=1) # Student pop. per class without W

for i in gpa_scale:
    gpa_data[i + 'asNum'] = gpa_data[i] * gpa_scale[i]

newLetterGrades = [i + 'asNum' for i in letterGrades]
gpa_data['GPA'] = gpa_data[newLetterGrades].sum(axis=1) / gpa_data['Students_Completed'] # Label

letterGrades.append('W')
gpa_data['Students'] = gpa_data[letterGrades].sum(axis=1) # Student pop. per class including with W

gpa_data
```

	Year	Term	YearTerm	Subject	Number	Course Title	Sched Type	A+	A	A-	...	B-asNum	C+asNum	CasNum	as
0	2022	Spring	2022-sp	AAS	100	Intro Asian American Studies	LCD	6	13	0	...	2.67	0.00	6.0	
1	2022	Spring	2022-sp	AAS	100	Intro Asian American Studies	DIS	0	11	5	...	5.34	2.33	0.0	
2	2022	Spring	2022-sp	AAS	100	Intro Asian American Studies	DIS	0	10	7	...	2.67	0.00	0.0	



# Linear Regression

Features:

Term (one-hot)

Year

Student population per section

Subject/Department name (one-hot)

Course Number

Label:

Average GPA of each section of each class

**R2 Score: 0.3108353231079505**

**Mean Squared Error: 0.11274402725443079**



# Deep Learning Model

Features:

- Term (one-hot)
- Year
- Student population per section
- Subject/Department name (one-hot)
- Course Number
- Course Title (one-hot)
- Primary Instructor (one-hot)

Label:

Average GPA of each section of each class

```
model summary:
Model: "sequential"

Layer (type)                Output Shape                Param #
=====
dense (Dense)                (None, 256)                 3741184
dense_1 (Dense)              (None, 256)                 65792
dense_2 (Dense)              (None, 1)                   257
=====
Total params: 3,807,233
Trainable params: 3,807,233
Non-trainable params: 0
```



# Hyperparameters

- \* Train-valid-test split ratio: 0.7 : 0.15 : 0.15
- \* Optimizer: Adam
- \* Initial learning rate: 0.001
- \* Activation function: ReLu
- \* Loss Function: Mean squared error (MSE)
- \* Epochs: 50
- \* Batch size: 64

---

# Different Optimizers



# SGD

```
Testing set:
141/141 [=====] - 1s 5ms/step - loss: 0.1560 - r_square: -1.1253e-04
1/1 [=====] - 0s 60ms/step

First five testing data points:
labels: [3.90939394 3.01621359 3.73956522 3.82          2.92307692]
predictions: [[3.3574088]
[3.3574088]
[3.3574088]
[3.3574088]
[3.3574088]]

model summary:
Model: "sequential"
```

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	3026176
dense_1 (Dense)	(None, 256)	65792
dense_2 (Dense)	(None, 2)	514
dense_3 (Dense)	(None, 1)	3

```
=====
Total params: 3,092,485
Trainable params: 3,092,485
Non-trainable params: 0
=====
```





# Adagrad

Testing set:  
141/141 [=====] - 1s 5ms/step - loss: 0.0953 - r\_square: 0.3877  
1/1 [=====] - 0s 91ms/step

First five testing data points:  
labels: [2.89134146 3.4725 3.6 3.7608 3.42282051]  
predictions: [[3.0486987]  
[3.5815594]  
[3.3927758]  
[3.57574 ]  
[3.3431113]]

model summary:  
Model: "sequential\_1"

Layer (type)	Output Shape	Param #
=====		
dense_4 (Dense)	(None, 256)	3005440
dense_5 (Dense)	(None, 256)	65792
dense_6 (Dense)	(None, 1)	257

=====

Total params: 3,071,489  
Trainable params: 3,071,489  
Non-trainable params: 0

=====



# Nadam

```
Testing set:
141/141 [=====] - 1s 4ms/step - loss: 0.0633 - r_square: 0.5934
1/1 [=====] - 0s 54ms/step

First five testing data points:
labels: [2.89134146 3.4725      3.6          3.7608      3.42282051]
predictions: [[2.9482253]
[3.620613 ]
[3.507869 ]
[3.6148648]
[3.5587149]]

model summary:
Model: "sequential_2"
```

Layer (type)	Output Shape	Param #
dense_7 (Dense)	(None, 256)	3005440
dense_8 (Dense)	(None, 256)	65792
dense_9 (Dense)	(None, 1)	257

```
=====
Total params: 3,071,489
Trainable params: 3,071,489
Non-trainable params: 0
=====
```



# RMSProp

```
Testing set:
141/141 [=====] - 1s 4ms/step - loss: 0.0790 - r_square: 0.4926
1/1 [=====] - 0s 54ms/step
```

```
First five testing data points:
labels: [2.89134146 3.4725      3.6          3.7608      3.42282051]
predictions: [[3.0347123]
[3.6743803]
[3.4486861]
[3.66894  ]
[3.3406034]]
```

```
model summary:
Model: "sequential_3"
```

Layer (type)	Output Shape	Param #
dense_10 (Dense)	(None, 256)	3005440
dense_11 (Dense)	(None, 256)	65792
dense_12 (Dense)	(None, 1)	257

```
=====
Total params: 3,071,489
Trainable params: 3,071,489
Non-trainable params: 0
=====
```



## Results and Feature Importance

$\alpha(F) = R\_squared(F) \approx 0.66$

$> \alpha(F \setminus \{'Term'\}) \approx 0.66$

$> \alpha(F \setminus \{'Year'\}) \approx 0.65$

$> \alpha(F \setminus \{'Student Population'\}) \approx 0.65$

$> \alpha(F \setminus \{'Subject'\}) \approx 0.65$

$> \alpha(F \setminus \{'Course Number'\}) \approx 0.65$

$> \alpha(F \setminus \{'Course Title'\}) \approx 0.59$

$> \alpha(F \setminus \{'Primary Instructor'\}) \approx 0.57$

```
Testing set:
300/300 [=====] - 1s 4ms/step - loss: 0.0561 - r_square: 0.6568
1/1 [=====] - 0s 77ms/step
```