Project 5 - Admission Prediction

By Yuheng Lin

Mount google drive to the colab

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989

Enter your authorization code:
··········
Mounted at /content/drive
```

Import packages

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import cross_val_score
```

Read the data file from google drive

```
student_data = pd.read_csv('/content/drive/My Drive/Admission_Predict_Ver1.1.csv', names = ["GRE"
student_data.head()
```

	GRE	TOEFL	UR	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65

Set up our inputs and outputs

```
inputs = student_data.drop('Chance of Admit', axis=1)
outputs = student_data['Chance of Admit']
```

Randomly Select Train and Test Inputs and Outputs

```
inputs_train, inputs_test, outputs_train, outputs_test = train_test_split(inputs, outputs)
```

1. Linear Regression

Train the model by using training data

```
LR = LinearRegression().fit(inputs_train, outputs_train)
```

Use 5 fold cross-validation check the accuarcy of the linear regression model

```
five_fold_LR_accuracy = cross_val_score(LR, inputs, outputs, cv=5)
print(five_fold_LR_accuracy)
```

```
[0.67763918 0.79424809 0.86447645 0.81935698 0.89828691]
```

2. Least Angle Regression

Import the LAR from sklearn.linear_model

Train my LAR model by using training inputs and outputs

```
from sklearn.linear_model import Lars

lar = Lars().fit(inputs_train, outputs_train)
```

Show the accuracy of my logistic regression by using 5-fold cross validation

```
five_fold_lar_accuracy = cross_val_score(lar, inputs, encoded_outputs, cv=5)
print(five_fold_lar_accuracy)
```

```
[0.68062195 0.79491182 0.86685492 0.82590321 0.90038994]
```

3. Bayesian Ridge Regression

Import Bayesian Ridge package from sklearn.linear_model

Train the Bayesian Ridge Regression model with training inputs and outputs

Show the accuracy of the Bayesian Ridge Regression model by using 5-fold cross validation

```
from sklearn.linear_model import BayesianRidge

brm = BayesianRidge()
```

```
brm.fit(inputs_train, outputs_train)

five_fold_brm_accuracy = cross_val_score(brm, inputs, outputs, cv=5)
print(five_fold_brm_accuracy)
```

```
[0.67593835 0.79382633 0.86389218 0.81885088 0.89845662]
```

Conclusion

By comparing the accuracies of three different model, we can see that least angle regression has the best accuracy which is around 82%.