

Machine learning
Assignment 2
Aakanksha Darekar
202200733
A1 09

Q. Build a simple linear regression model using the given data- set (Auto.csv). Perform a linear regression with mpg as the response and all other variables as the features.

```
import pandas as pd

import statsmodels.api as sm

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

data = pd.read_csv('Auto.csv')

# Define the response variable and the features

X = data[['cylinders', 'displacement', 'horsepower', 'weight', 'acceleration', 'model_year', 'origin']]

y = data['mpg']

# Add a constant to the features (required for statsmodels)

X = sm.add_constant(X)

# Fit the linear regression model

model = sm.OLS(y, X).fit()

# (a) Print model summary
```

```
print(model.summary())
```

```
# (b) Plot residual plot
```

```
residuals = model.resid
```

```
sns.residplot(x=model.fittedvalues, y=residuals, lowess=True)
```

```
plt.xlabel('Fitted values')
```

```
plt.ylabel('Residuals')
```

```
plt.title('Residual Plot')
```

```
plt.show()
```

```
# (c) Plot scatter plot showing the fitted line
```

```
plt.scatter(y, model.fittedvalues)
```

```
plt.plot(y, y, color='red') # Line showing the perfect fit
```

```
plt.xlabel('Actual mpg')
```

```
plt.ylabel('Fitted mpg')
```

```
plt.title('Actual vs Fitted mpg')
```

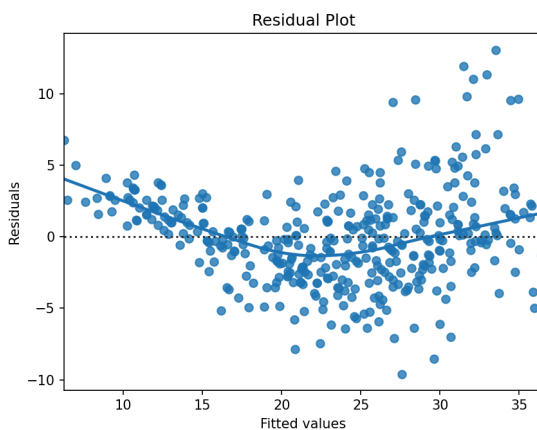
```
plt.show()
```

(a) Print model summary and

```
=====
                        OLS Regression Results
=====
Dep. Variable:          mpg      R-squared:          0.821
Model:                  OLS      Adj. R-squared:       0.818
Method:                 Least Squares  F-statistic:       252.4
Date:                   Wed, 22 Jan 2025  Prob (F-statistic):  2.04e-139
Time:                   15:49:32   Log-Likelihood:    -1023.5
No. Observations:       392      AIC:               2063.
Df Residuals:           384      BIC:               2095.
Df Model:                7
Covariance Type:        nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-17.2184	4.644	-3.707	0.000	-26.350	-8.087
cylinders	-0.4934	0.323	-1.526	0.128	-1.129	0.142
displacement	0.0199	0.008	2.647	0.008	0.005	0.035
horsepower	-0.0170	0.014	-1.230	0.220	-0.044	0.010
weight	-0.0065	0.001	-9.929	0.000	-0.008	-0.005
acceleration	0.0806	0.099	0.815	0.415	-0.114	0.275
model_year	0.7508	0.051	14.729	0.000	0.651	0.851
origin	1.4261	0.278	5.127	0.000	0.879	1.973
Omnibus:		31.906	Durbin-Watson:		1.309	
Prob(Omnibus):		0.000	Jarque-Bera (JB):		53.100	
Skew:		0.529	Prob(JB):		2.95e-12	
Kurtosis:		4.460	Cond. No.		8.59e+04	

(b) Plot residual plot



(c) Plot scatter plot showing the fitted line

