02_ML_Multiple_linear_regression

1 Machine Learning

1.1 Multiple Linear Regression

▾ Step-1 Import dataset

```
import pandas as pd
df = pd.read_csv("ml_data_salary.csv")
df.head()
```

	age	distance	YearsExperience	Salary
0	31.1	77.75	1.1	39343
1	31.3	78.25	1.3	46205
2	31.5	78.75	1.5	37731
3	32.0	80.00	2.0	43525
4	32.2	80.50	2.2	39891

▾ Step-2 Define dependent and independent variables

```
X = df[["age","distance", "YearsExperience"]]
y = df["Salary"]
```

▾ Step-3 Fit Linear Regression Model

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model = model.fit(X, y)
model
```

```
▾ LinearRegression
LinearRegression()
```

```
model.coef_
```

```
array([-3.00216193e+15,  1.18788781e+15,  3.24424072e+13])
```

Step-4 Evaluating Model Fitness

```
# Model Fitness
print("Score for data =" ,model.score(X, y))
```

```
Score for data = 0.9569960750337954
```

Step-5 Prediction of unknown values

```
model.predict([[31.1,77.75,1.1]])
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
array([36209.375])
```

STEP-6 Linear Accuracy Score

```
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

# Assume you have your features (X) and target variable (y) defined

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Predict on the test set
y_pred = model.predict(X_test)

# Compute the accuracy score (R-squared)
accuracy = r2_score(y_test, y_pred)

print("Accuracy score:", accuracy)
```

```
Accuracy score: 0.9024461774180497
```

Colab paid products - Cancel contracts here

✓ 0s   completed at 6:52 PM   ● ✕