LINEAR REGRESSION

Isaac Lo, Ronica Gupta, Brendan Jenkins

Overview

- What is linear regression?
- Simple linear regression
- Multiple linear regression

Terms

- Target variable (Y) - value to be predicted
- Feature variables (X) - values used to predict target variables

$$Y = mX + b$$

What is Linear Regression?

- Predict target value (Y) given a set of feature variables (X)
- Explain relationship between feature (X) and target value (Y)

Simple Linear Regression

- One target variable
- One feature variable

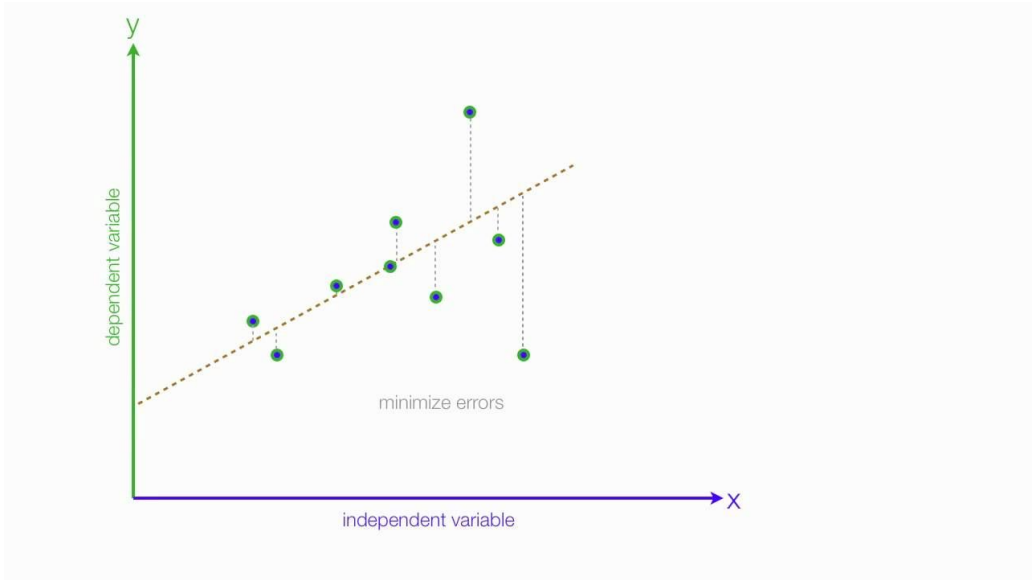
SLR: Predicting Y from X

- How can we predict Y from X, given existing data on X and Y?
- Model set up: $\hat{Y} = \beta_0 + \beta_1 X$
- Find best model parameters:
  - β_1 - slope
  - β_0 - intercept

SLR: Finding Best Model Parameters

- Minimize: $e = Y - \hat{Y}$
- Y - observed target value
- $\hat{Y}$ - predicted target value
- e - error

$$\hat{Y} = \beta_0 + \beta_1 X$$



Example: Car Price vs Horsepower

- $Price = \beta_0 + \beta_1 Horsepower$
  - $\beta_1 = 145$
  - $\beta_0 = -2500$

- $Horsepower = 150$

$$25000 = -2500 + 145(150)$$

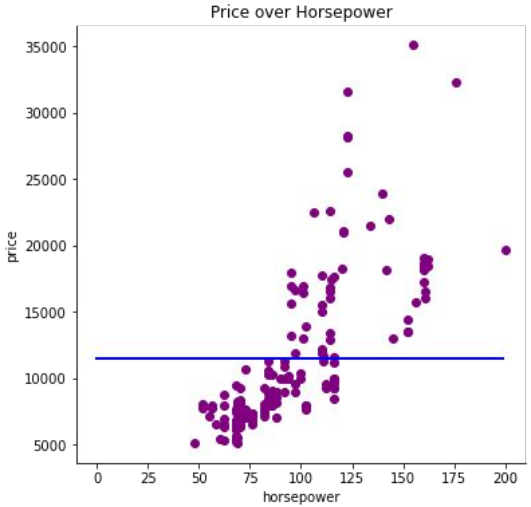
Measuring Relationship Between X and Y

- How do we know X is relevant to predicting Y?
- Hypothesis test
  - $H_0 : \beta_1 = 0$
  - $H_1 : \beta_1 \neq 0$

Example: Car Price vs Horsepower

- $H_0 : Price = \beta_0 + (0)Horsepower$
- $H_1 : Price = \beta_0 + (s)Horsepower$

H_0 :



H_1 :

Multiple Linear Regression

- One target variable
- Two or more feature variables

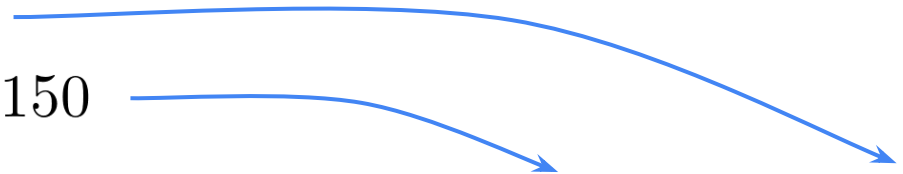
MLR: Predicting Y From Set of Xs

- Model set up: $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$
- Find best model parameters that minimizes error:
  - β_0 - intercept
  - β_1 - slope of X_1
  - β_2 - slope of X_2
  - β_p - slope of X_p

Example: Car Price vs Horsepower and Weight

- $Price = \beta_0 + \beta_1 Horsepower + \beta_2 Weight$
  - $\beta_0 = -2488$
  - $\beta_1 = 145$
  - $\beta_2 = 9.5$

- $Weight = 2500$
- $Horsepower = 150$



$$43118 = -2488 + 145(150) + 9.5(2500)$$

Example: Car Price vs Horsepower and Weight

- Hypothesis test on Horsepower:
  - $H_0 : Price = \beta_0 + (0)Horsepower + \beta_2 Weight$
  - $H_1 : Price = \beta_0 + (s)Horsepower + \beta_2 Weight$
- Hypothesis test on Weight:
  - $H_0 : Price = \beta_0 + \beta_1 Horsepower + (0)Weight$
  - $H_1 : Price = \beta_0 + \beta_1 Horsepower + (v)Weight$

Code Implementation

- Simple Linear Regression

```
# Make SLR model
slr_model = smf.ols('price ~ horsepower', data=df).fit()
```

```
slr_model.params
```

```
Intercept    -2488.731679
horsepower     145.398303
```

- Multiple Linear Regression

```
# Make mlr model
mlr_model = smf.ols('price ~ horsepower + curb_weight', data=df).fit()
```

```
mlr_model.params
```

```
Intercept     -14607.999733
horsepower        27.403983
curb_weight        9.518940
```

Summary

- Goal:
  - Predict target variable (Y) from feature variable(s) (X)
  - Explain relationship between target variable(Y) and feature variable(s) (X)
- Simple Linear Regression
- Multiple Linear Regression
- Implementation in Python