Python for Data Science

Machine Learning in Python: Regression Analysis

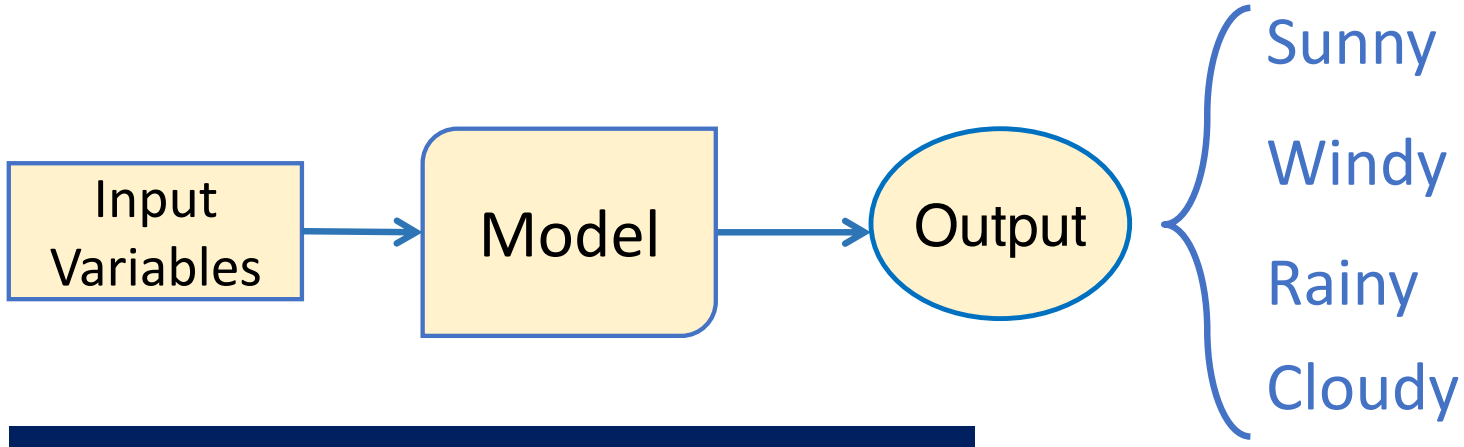
Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

By the end of this video, you should be able to:

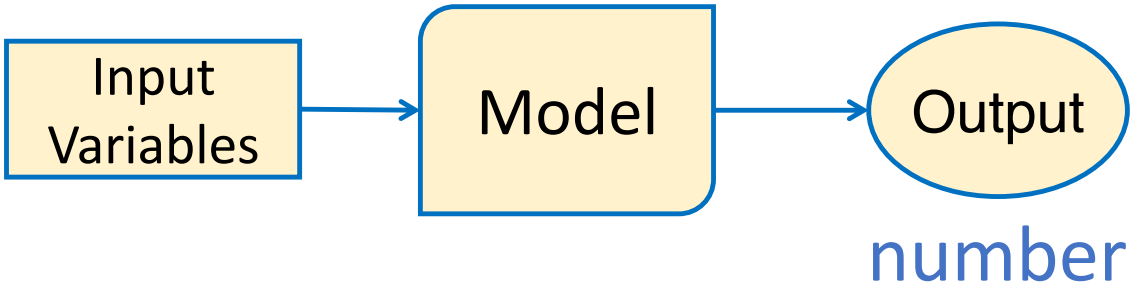
- Define what regression is
- Explain the difference between regression and classification
- Name some applications of regression

Classification Review



Classification:
Given input variables, predict category

Regression

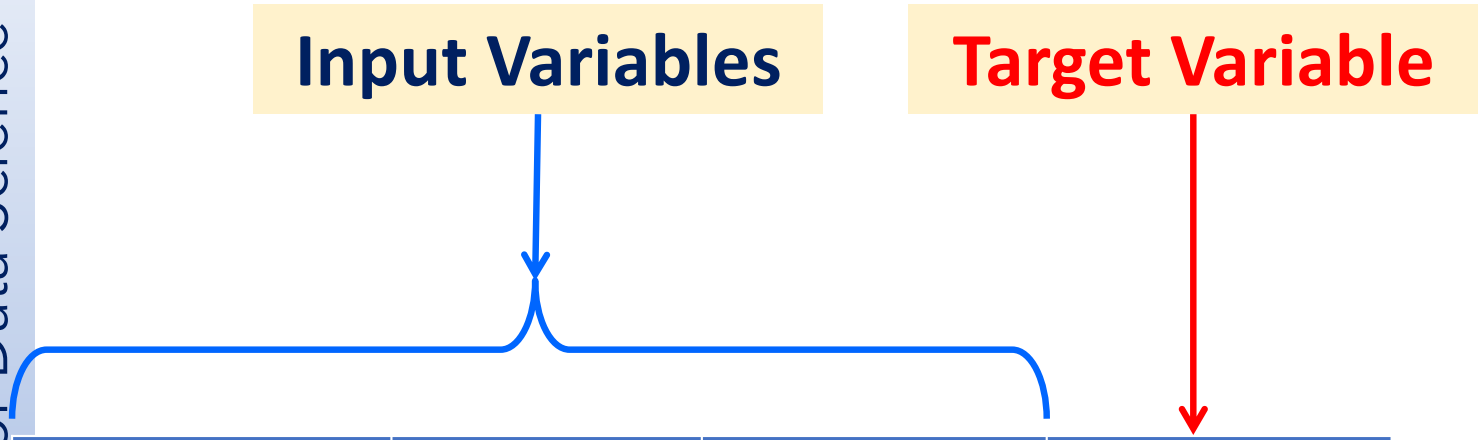


Regression:
Given input variables, predict numeric value

Regression Examples

- Forecast high temperature for next day
- Estimate average house price for a region
- Determine demand for a new product
- Predict power usage

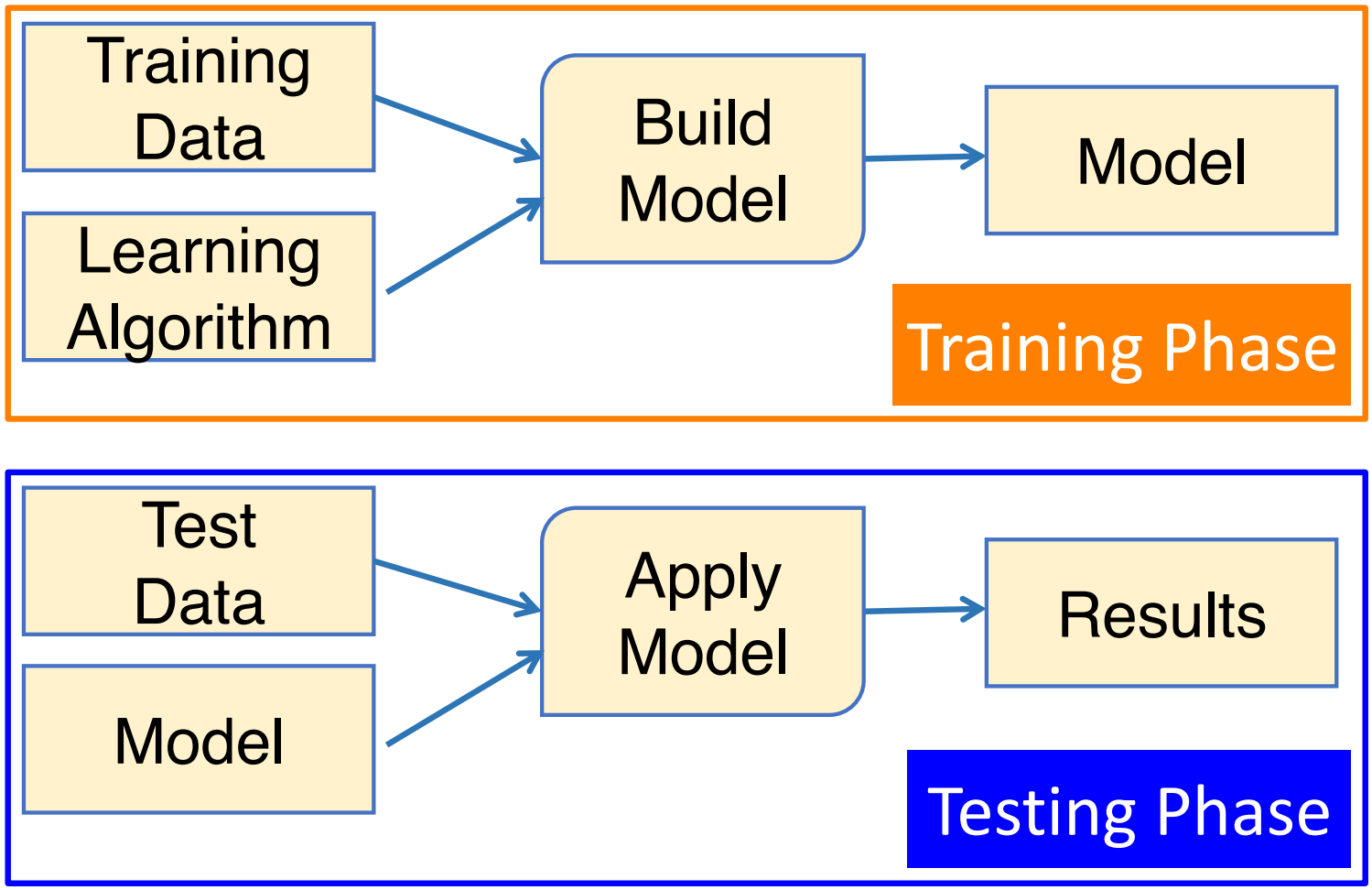
Regression is Supervised



Target is provided

Today's High	Today's Low	Month	Tomorrow's High
79	64	July	81
60	45	October	58
68	49	May	65
57	47	January	54

Training vs. Testing Phases



Datasets

Training Data

Adjust model parameters

Validation Data

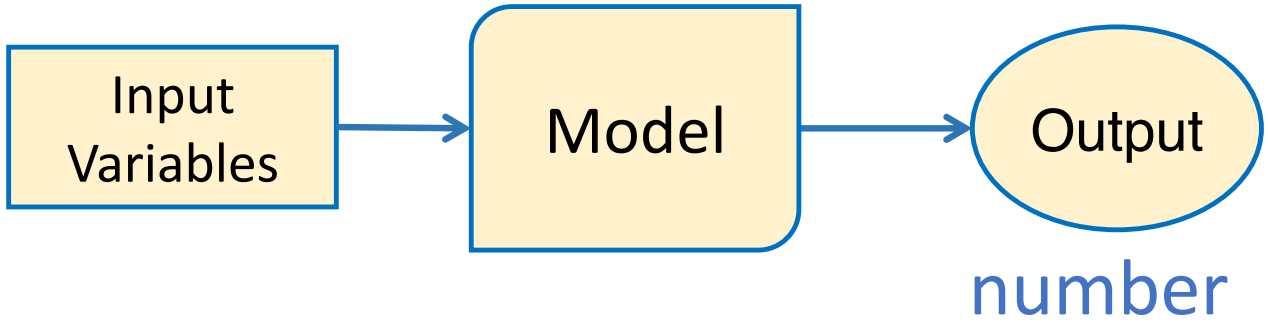
Determine when to stop training (avoid overfitting)

Estimate generalization performance

Test Data

Evaluate performance on new data

- Predict number from input variables
- Regression is a supervised task
- Target variable is numerical



Python for Data Science

Machine Learning in Python: Linear Regression

Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

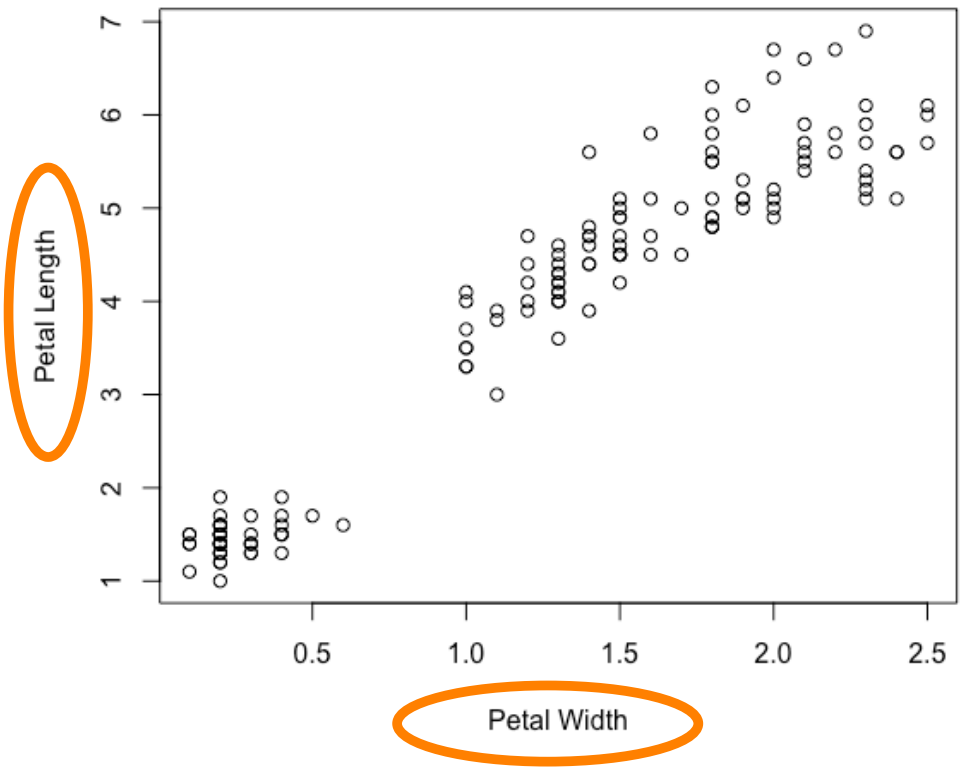
By the end of this video, you should be able to:

- Describe how linear regression works
- Discuss how least squares is used in linear regression
- Define simple and multiple linear regression

Linear Regression

- Captures relationship between numerical output and input variables
- Relationship is modeled as linear

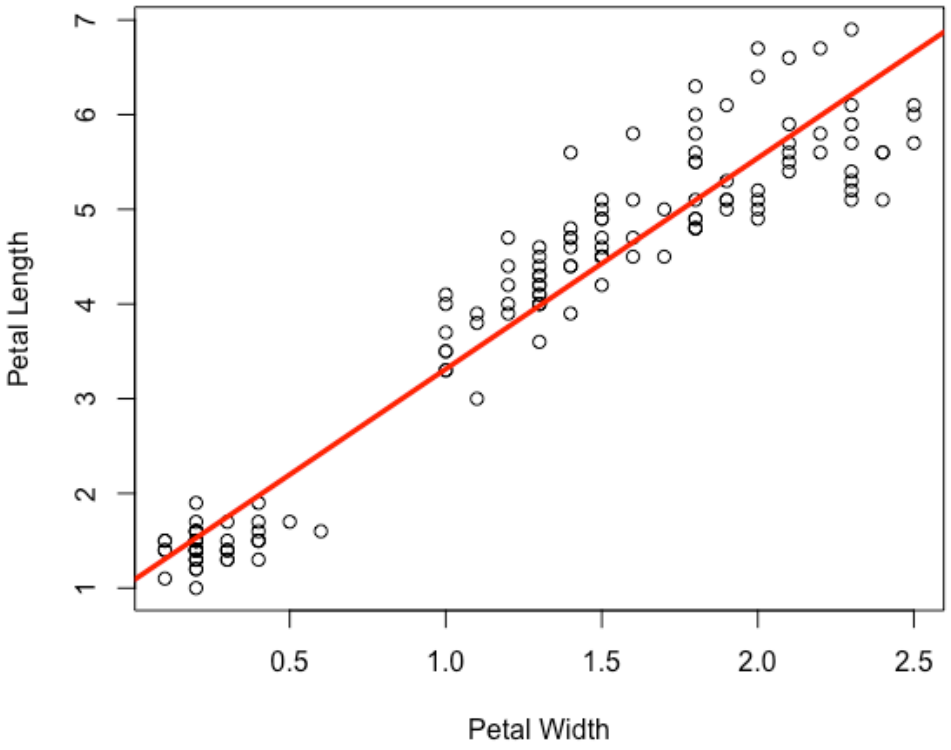
Linear Regression Model



Regression Task:

Given petal width, predict petal length.

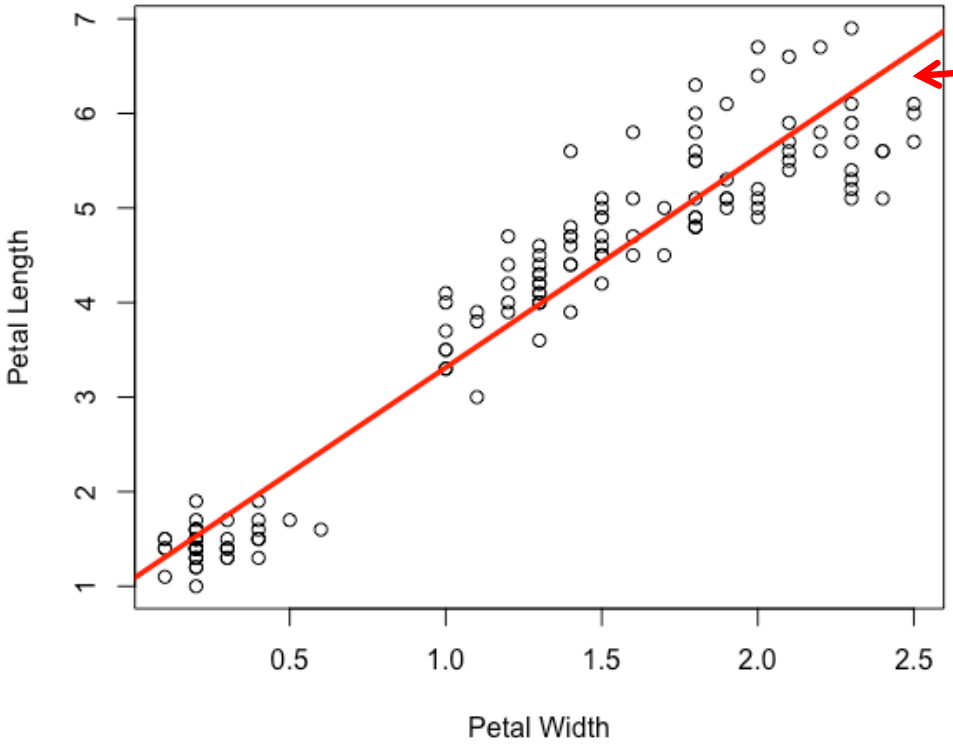
Linear Regression Model



Regression Task:

Given petal width, predict petal length.

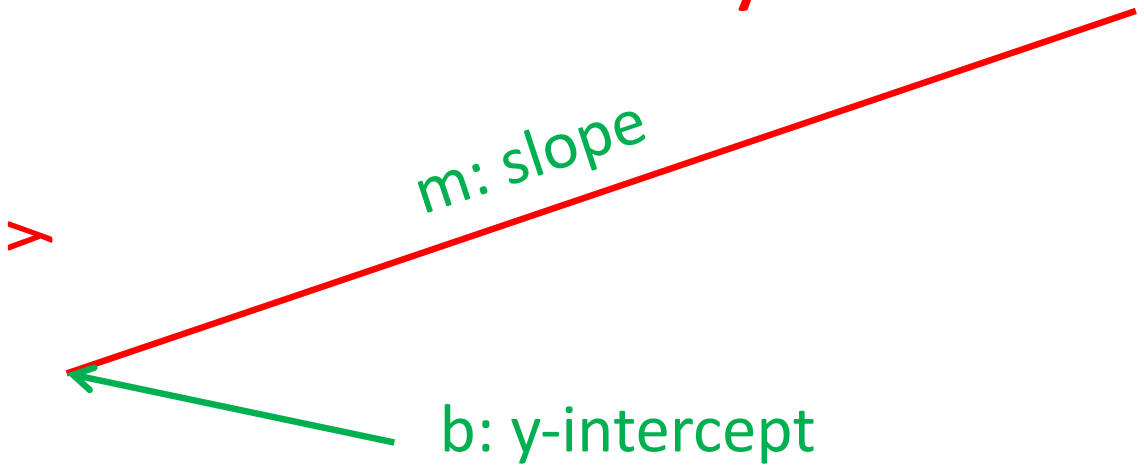
Linear Regression Model



regression line

Least Squares Algorithm

$y = mx + b$

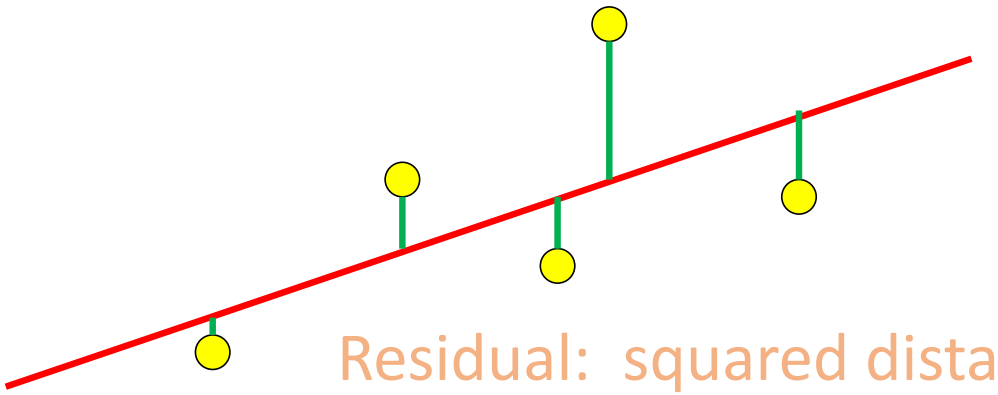


regression line

m, b: model parameters

Training linear regression model adjusts model parameters to fit samples

Least Squares Method

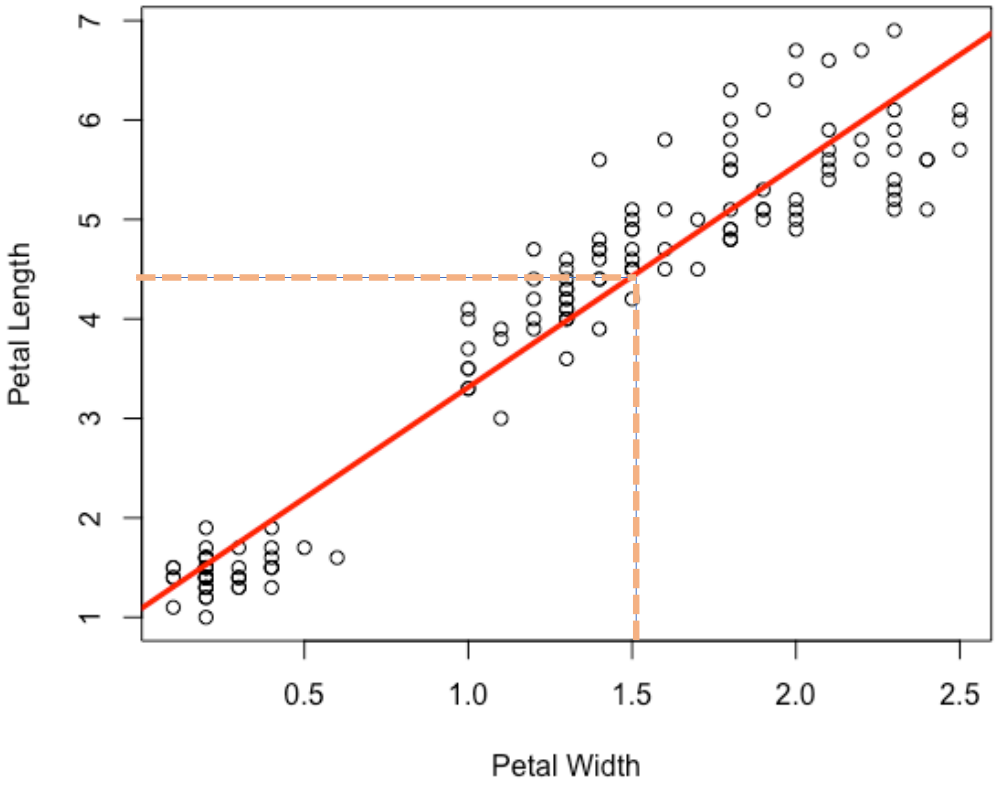


regression line
sample
distance from regression line (error)

Residual: squared distance from regression line

Goal: Find regression line that makes sum of residuals as small as possible

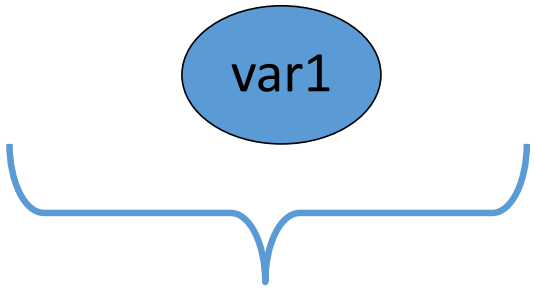
Linear Regression Model



Applying model:
Given petal width =1.5,
prediction is
petal length = 4.5

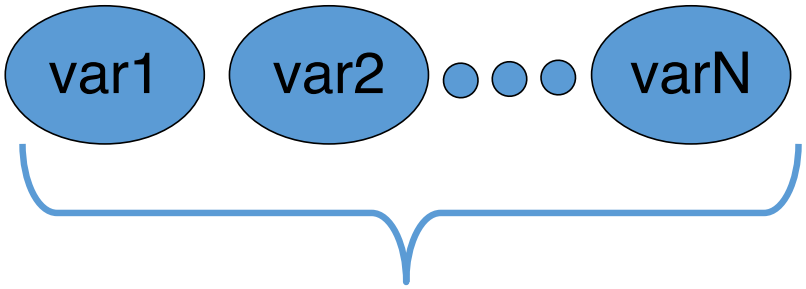
Types of Linear Regression

Simple Linear Regression



Input has one variable

Multiple Linear Regression



Input has >1 variables

Linear Regression Summary

- Captures linear relationship between numerical output and input variables
- Model can be fitted using least squares