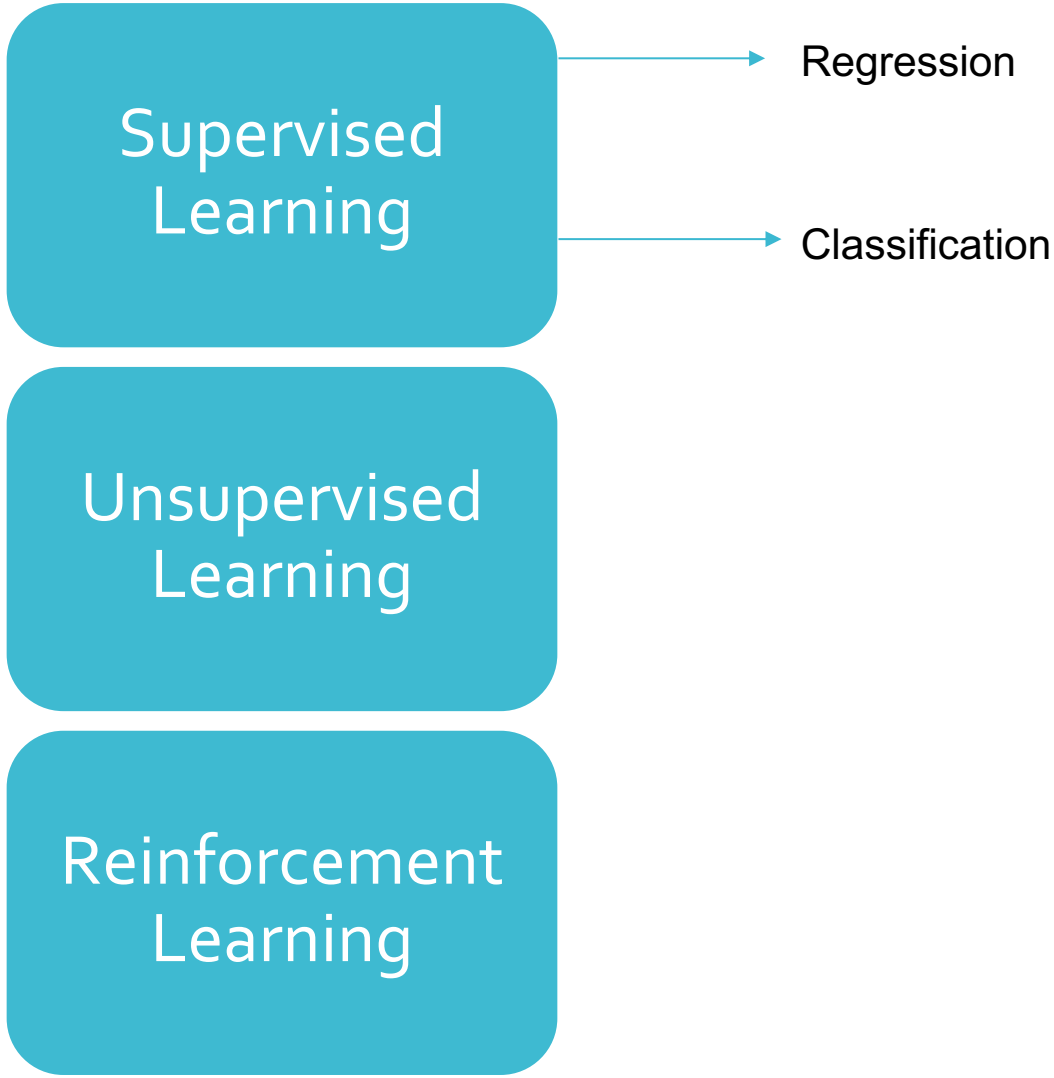
Intro to Simple Linear Regression

Sucheta Sadalgekar

Agenda

- Introduction
- Types of Learning
- Supervised Learning – Simple Linear Regression
- Use case in Python

Types of Learning



What is Regression?

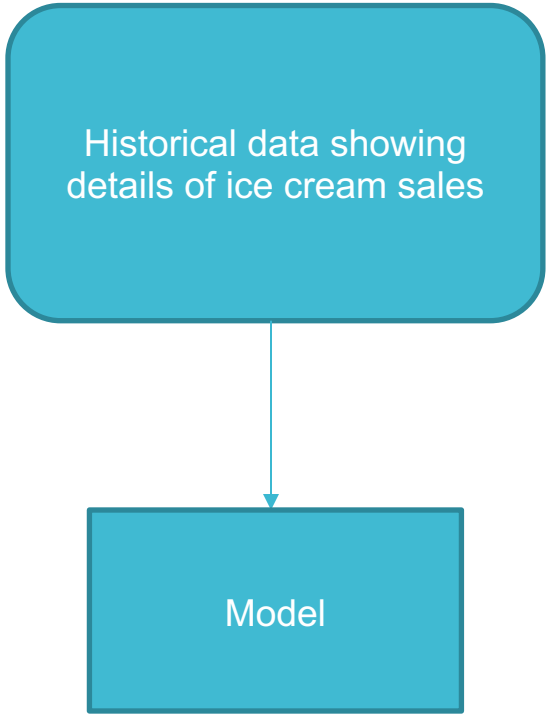
X: Independent Variable

Y: Dependent Variable

Temperature	Price	Ice Cream Sales
57.56	0.27	215
61.52	0.282	325
53.42	0.277	185
59.36	0.28	332
65.3	0.272	406
71.78	0.262	522
66.92	0.275	412
77.18	0.267	614
74.12	0.265	544
64.58	0.277	421
72.68	0.282	445
62.96	0.27	

Regression is the process of predicting a continuous value

Regression Model



Types of Regression models

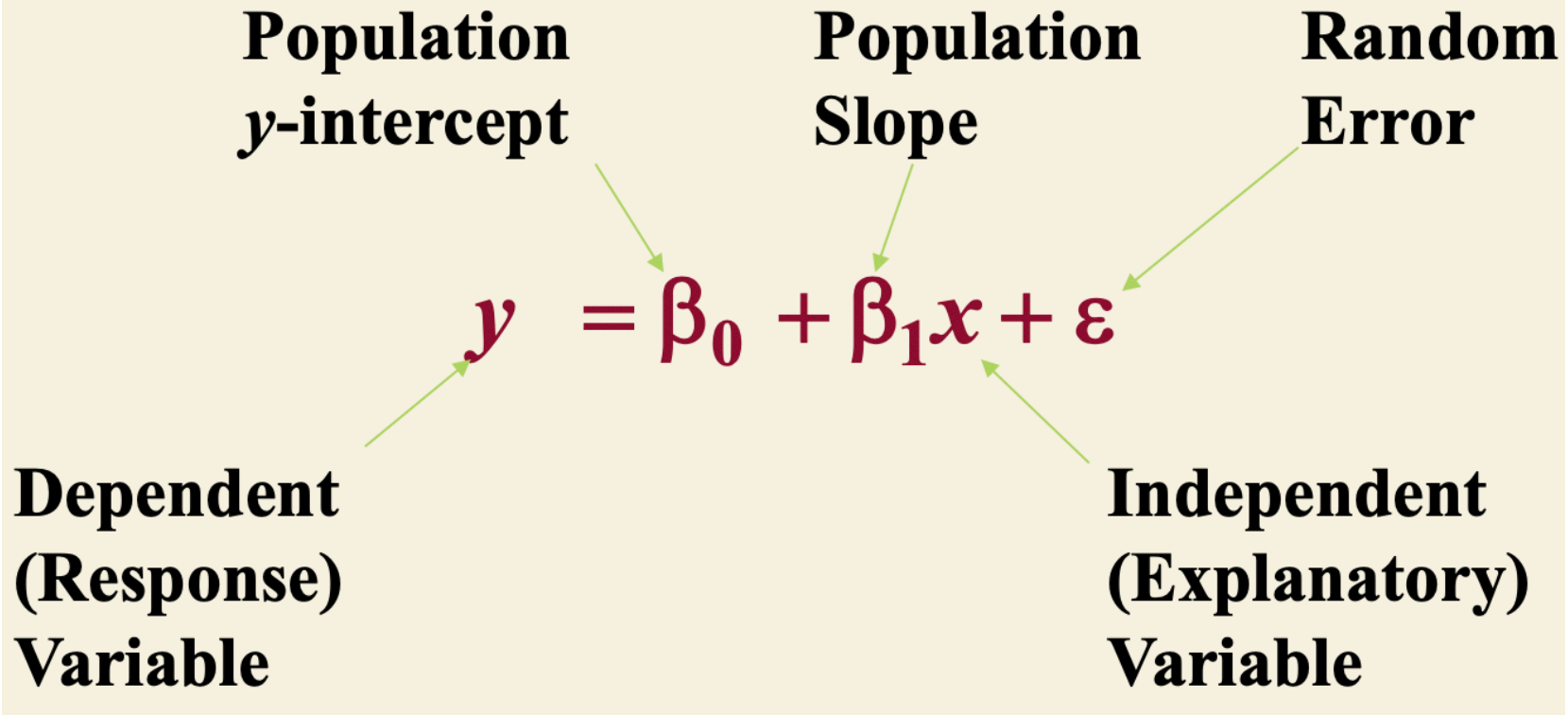
- Simple Regression
- Simple Linear Regression
- Simple Non- linear Regression

Example: Predict ice-cream sales based on temperature

- Multiple Regression
- Linear Multiple Regression
- Non-Linear Multiple Regression

Example: Predict ice-cream sales based on temperature and price

Simple Linear Regression



Linear Regression Assumptions

- Linear relation between dependent and independent variable
- Assumption II: errors are independent
- Assumption III: errors are normally distributed

- QQ plot

- Shapiro-wilk test

- Assumption IV: errors have mean o and std sigma (homoscedastic)

Metric for Regression models

- RMSE
- MSE
- R Squared
- MAE

Pros of linear regression

- Fast
- No parameter tuning
- Easy to understand, highly interpretable

- Train / Split


Test on a portion of train set
Entire Dataset
Training Set
Testing Set
• Test-set is a portion of the train-set
• High "training accuracy"
• Low "out-of-sample accuracy"
Train/Test Split
Entire Dataset
Training Set
Testing Set
• Mutually exclusive
• More accurate evaluation on out-of-sample accuracy

Thank you!