

# Introduction to Data Science

## Linear Regression



**Arijit Mondal**

Dept. of Computer Science & Engineering

Indian Institute of Technology Patna

`arijit@iitp.ac.in`

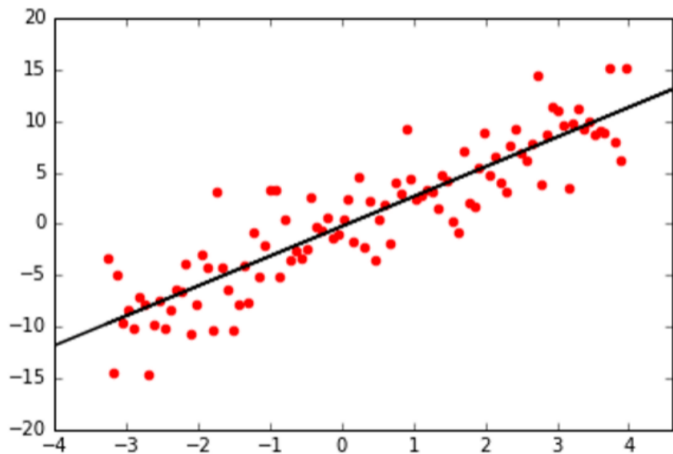
# Introduction

- Most representative machine learning method
- Easy to understand
- Generally appropriate as a default / baseline model
  - Price of a house grows linearly with area
  - Weight increases linearly with the food eaten

# Introduction

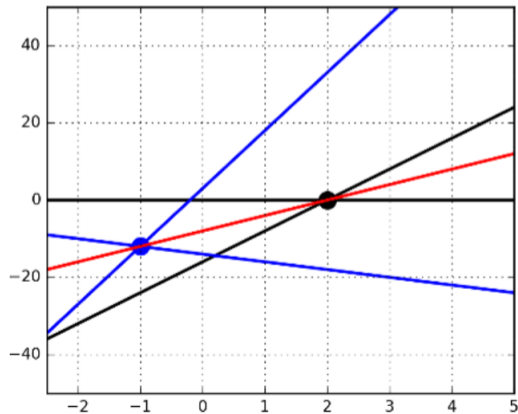
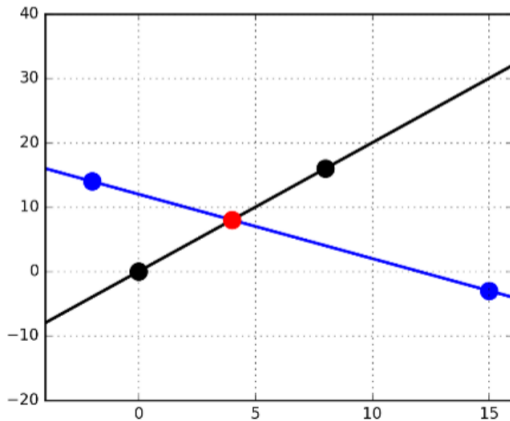
- Most representative machine learning method
- Easy to understand
- Generally appropriate as a default / baseline model
  - Price of a house grows linearly with area
  - Weight increases linearly with the food eaten
- Statistician's rule - if you want a linear function, just measure two points only!!!

# Example



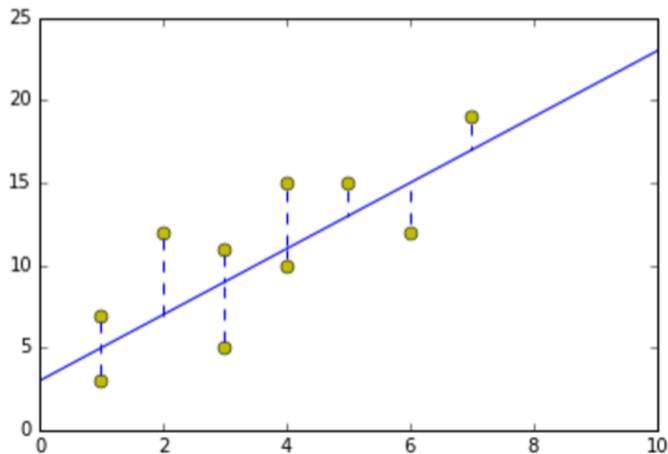
# Linear regression and duality

- For solving linear systems, we find the single point that lies on  $n$  given lines
- In regression, we are given  $n$  points and we seek the line that lies on all the points
- By duality transformation both are the same



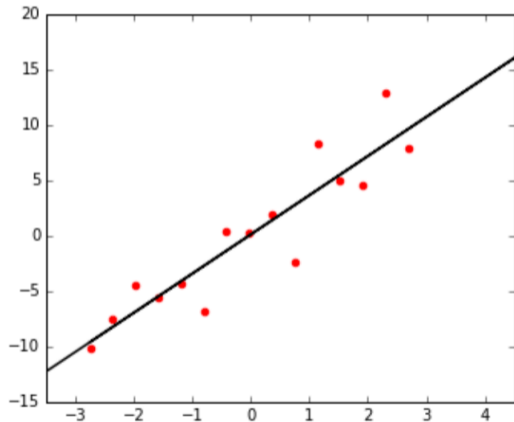
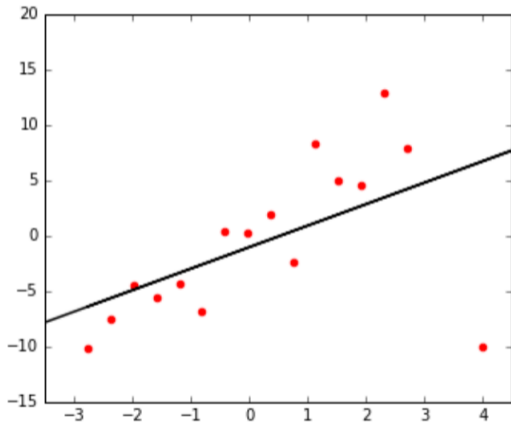
# Error in linear regression

- Residual error
- Least squares regression minimizes the sum of the squares of the residuals of all points



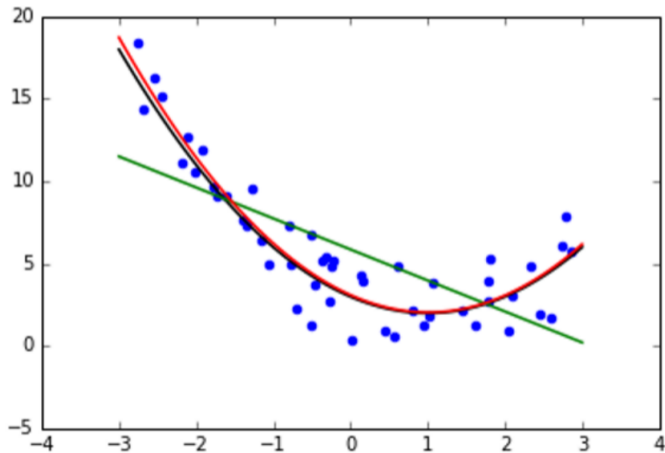
# Finding the optimal fit

# Better models: Removing outliers





# Better models: Non-linear functions



# Feature and target scaling

- Unreadable coefficient
- Numerical imprecision
- Inappropriate formulations
- Feature scaling — Z-score

# Regression as parameter fitting

# Convex parameter spaces

