MSE491: Application of Machine Learning in Mechatronic Systems

Linear Regression and Gradient Descent

Mohammad Narimani, *Ph.D., P.Eng.*
Lecturer
School of Mechatronic Systems Engineering
Simon Fraser University

SFU SIMON FRASER UNIVERSITY SURREY

Outline

- Regression and Gradient Descent
- Overfitting and complexity; training set, validation set, test set

Simple Linear Regression

- Example: As an HR officer, you have a candidate with 5 years of experience. There is no solid formula for salary calculation. Then what is the best salary you should offer to him, while you have the files for some pervious employees with different years of experience?

Simple Linear Regression

YearsExperience	Salary
1.1	39,343
1.3	46,205
1.5	37,731
2.0	43,525
2.2	39,891
2.9	56,642
3.0	60,150
3.2	54,445
3.2	64,445
3.7	57,189
3.9	63,218
4.0	55,794
4.0	56,957
4.1	57,081
4.5	61,111
4.9	67,938
5.1	66,029
5.3	83,088
5.9	81,363
6.0	93,940
6.8	91,738
7.1	98,273
7.9	101,302
8.2	113,812
8.7	109,431
9.0	105,582
9.5	116,969
9.6	112,635
10.3	122,391
10.5	121,872

hypothesis : $h_\theta(x) = \theta_0 + \theta_1 x$

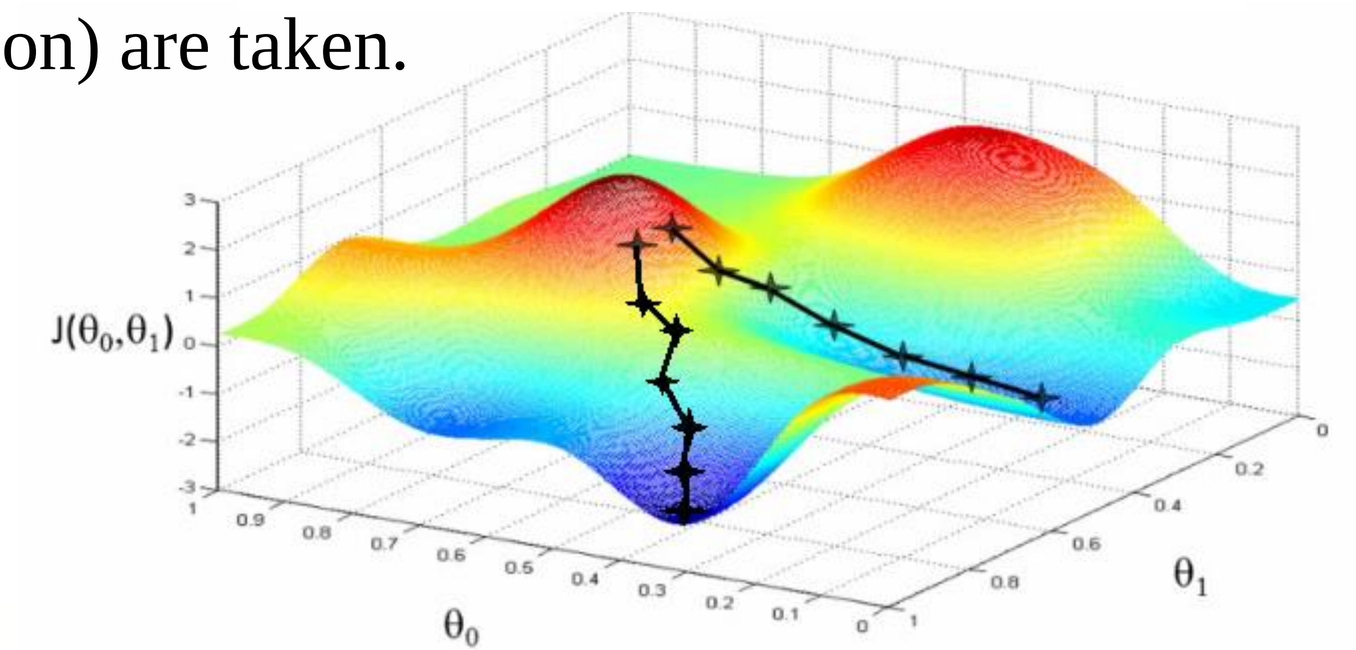
How to choose $\theta_i's$?

-

Gradient descent

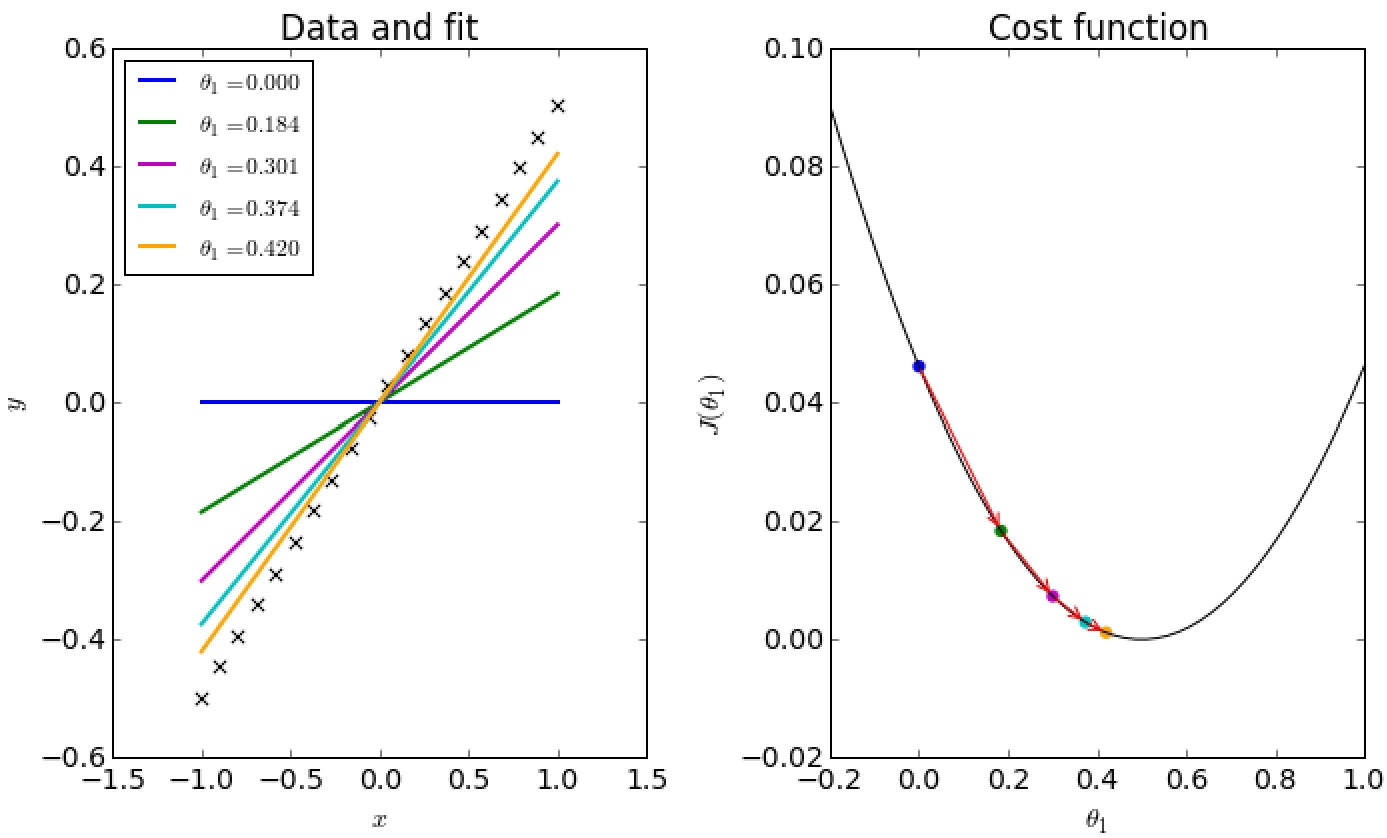
In gradient descent the step taken goes in the direction of the steepest descent/ascent, a direction given by the gradient of the function.

In the figure below, the journey starts at two locations on top of the hill and each two descent paths (minimization) are taken.



Gradient descent

Gradient descent

- 

Gradient descent

- Learning Rate:
  - Large Learning Rate value:
  - Small Learning Rate value:
  - Proper Value

Gradient descent for Linear Regression

Multiple Linear Regression

- The prediction of house prices using the size of the house (in square feet), and the number of bedrooms.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots \theta_n x_n$$

n: is the number of features

$x_j^{(i)}$: j^{th} feature of the i^{th} training example

Size (sf)	# of BR	Price
2104	3	399900
1600	3	329900
2400	3	369000
1416	2	232000
3000	4	539900
1985	4	299900
1534	3	314900
1427	3	198999
1380	3	212000
1494	3	242500
1940	4	239999
::::	::::	::::
1200	3	299000
852	2	179900
1852	4	299900
1203	3	239500

Multiple Linear Regression

- Notation:

Multiple Linear Regression

- Gradient Descent

Multiple Linear Regression

- Feature Scaling:

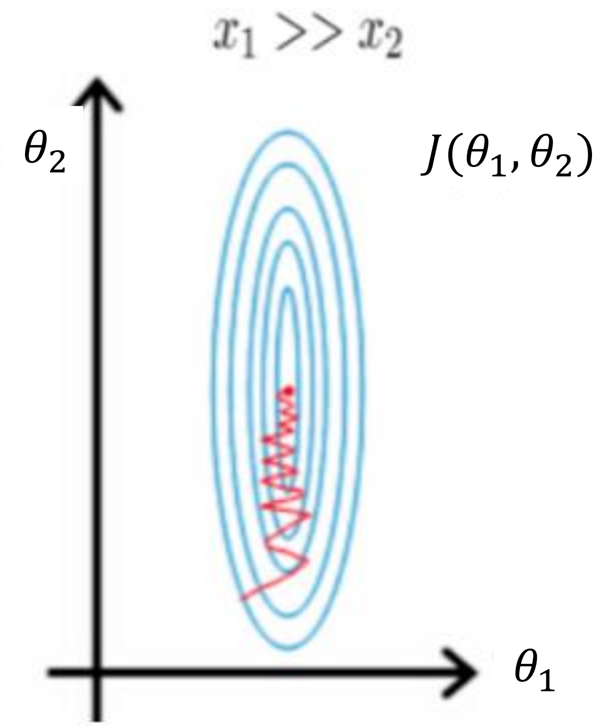
Without scaling, in some machine learning algorithms, cost functions will not work properly without normalization:

- It may takes longer time to find the optimum values

Size (sf)	# of BR	Price
2104	3	399900
1600	3	329900
2400	3	369000
1416	2	232000
3000	4	539900
1985	4	299900
1534	3	314900
1427	3	198999
1380	3	212000
1494	3	242500
1940	4	239999
::::	::::	::::
1200	3	299000
852	2	179900
1852	4	299900
1203	3	239500

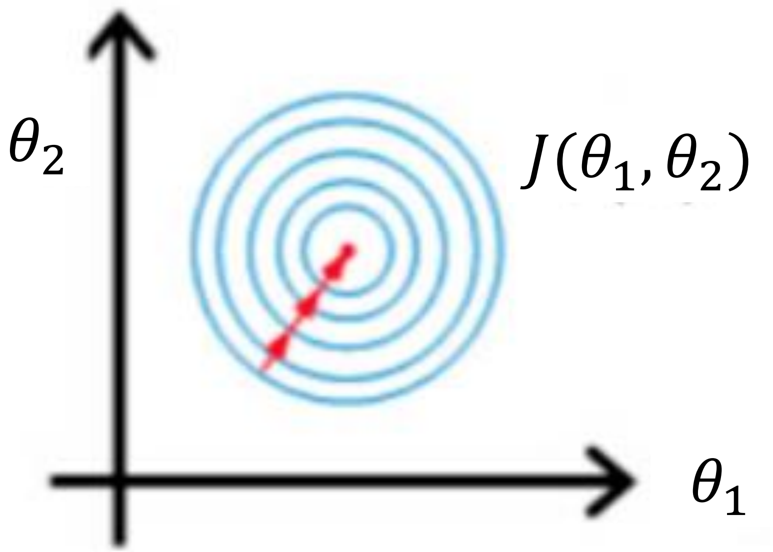
Multiple Linear Regression

- Feature Scaling:



Multiple Linear Regression

- Feature Scaling:



Multiple Linear Regression

- Feature Scaling:

Notes:

Multiple Linear Regression

- Feature Scaling: Mean Normalization

$$x_i \rightarrow \frac{x_i - x_{mean}}{x_{max} - x_{min}}$$

$$for \quad i \neq 0$$

Then,

$$-0.5 < x_i < 0.5$$

Analytical solution for linear regression problems

- Linear regression problems are represented in matrix notations as:

$$X.\theta = y$$

where X is an $m \times (n+1)$ matrix.

m: # of samples (training values)

n : # of features

Then

$$\theta = \left(X^T X\right)^{-1}.X^T.y$$