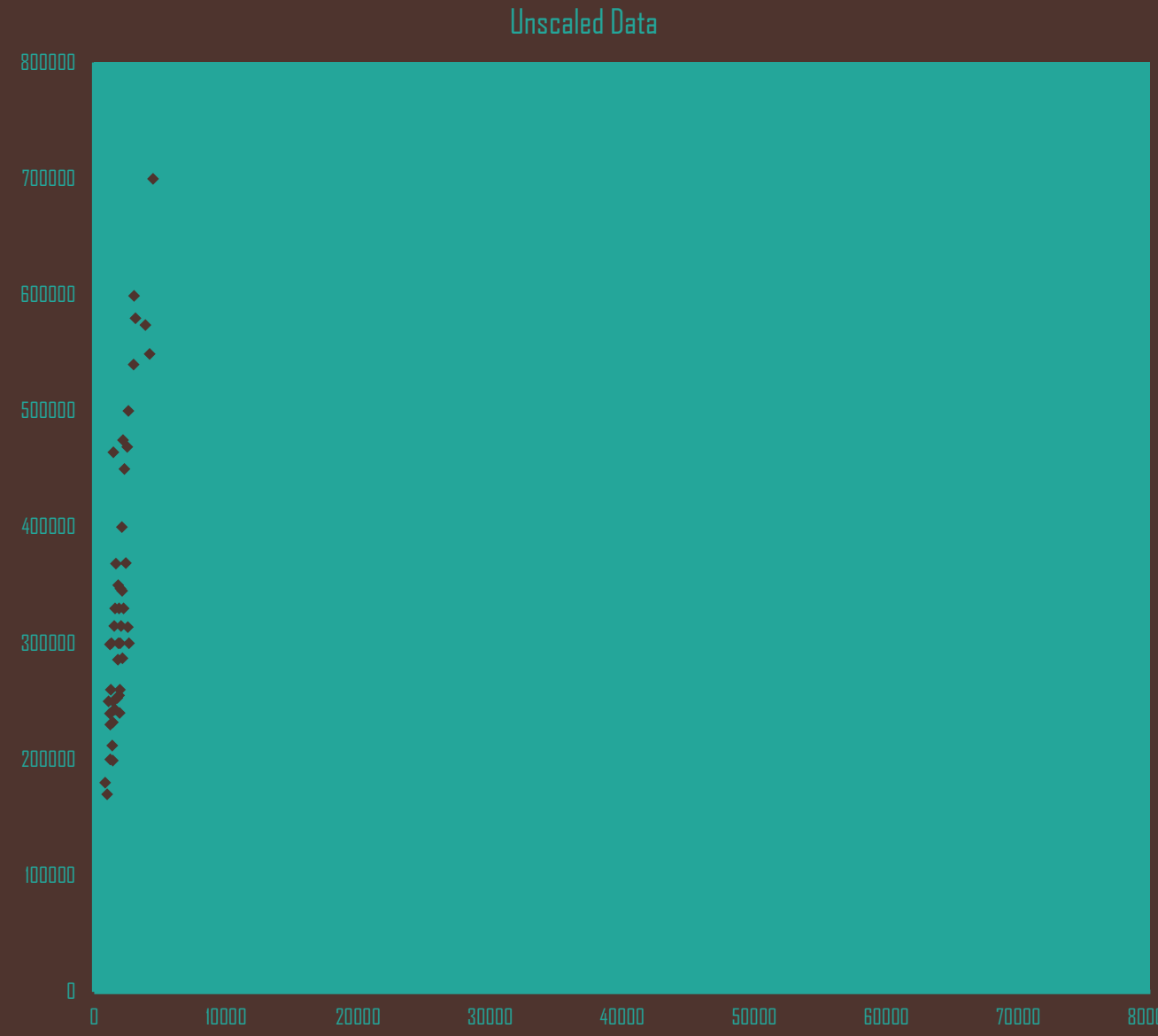
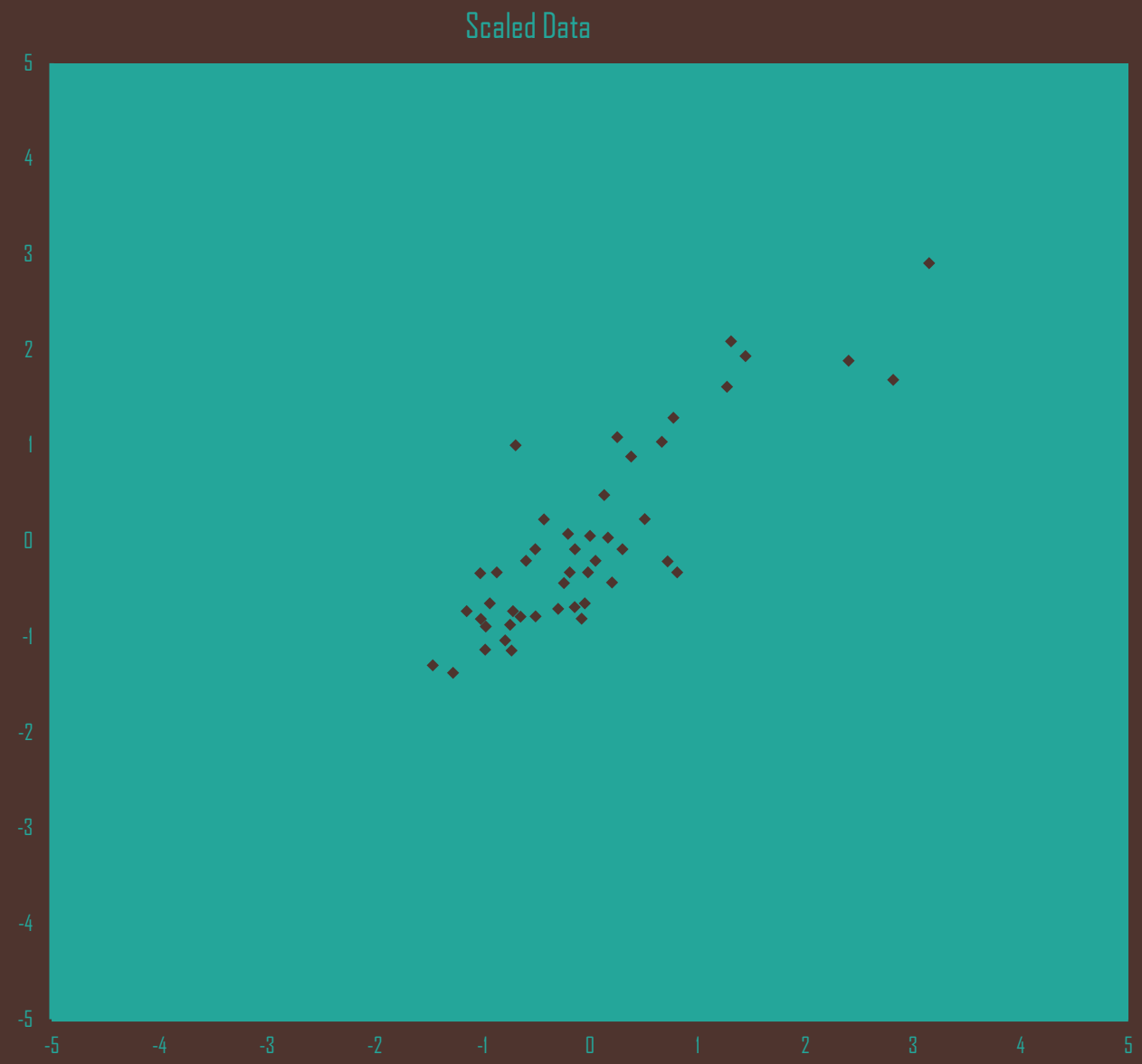
Feature Scaling


Unscaled Data
800000
700000
600000
500000
400000
300000
200000
100000
0
0
10000
20000
30000
40000
50000
60000
70000
80000


Scaled Data
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

One of the main requirements of most ML algorithms is feature scaling

Feature Scaling ensures that the each feature and output are uniformly scaled

$$z_j^{(i)} = \frac{x_j^{(i)} - \mu_j}{\sigma_j}$$

Where:

μ_j = is the mean of all values for feature j and

σ_j = is the standard deviation of all values for feature j

$$z_y^{(i)} = \frac{y^{(i)} - \mu_y}{\sigma_y}$$

Where:

μ_y = is the mean of all output values

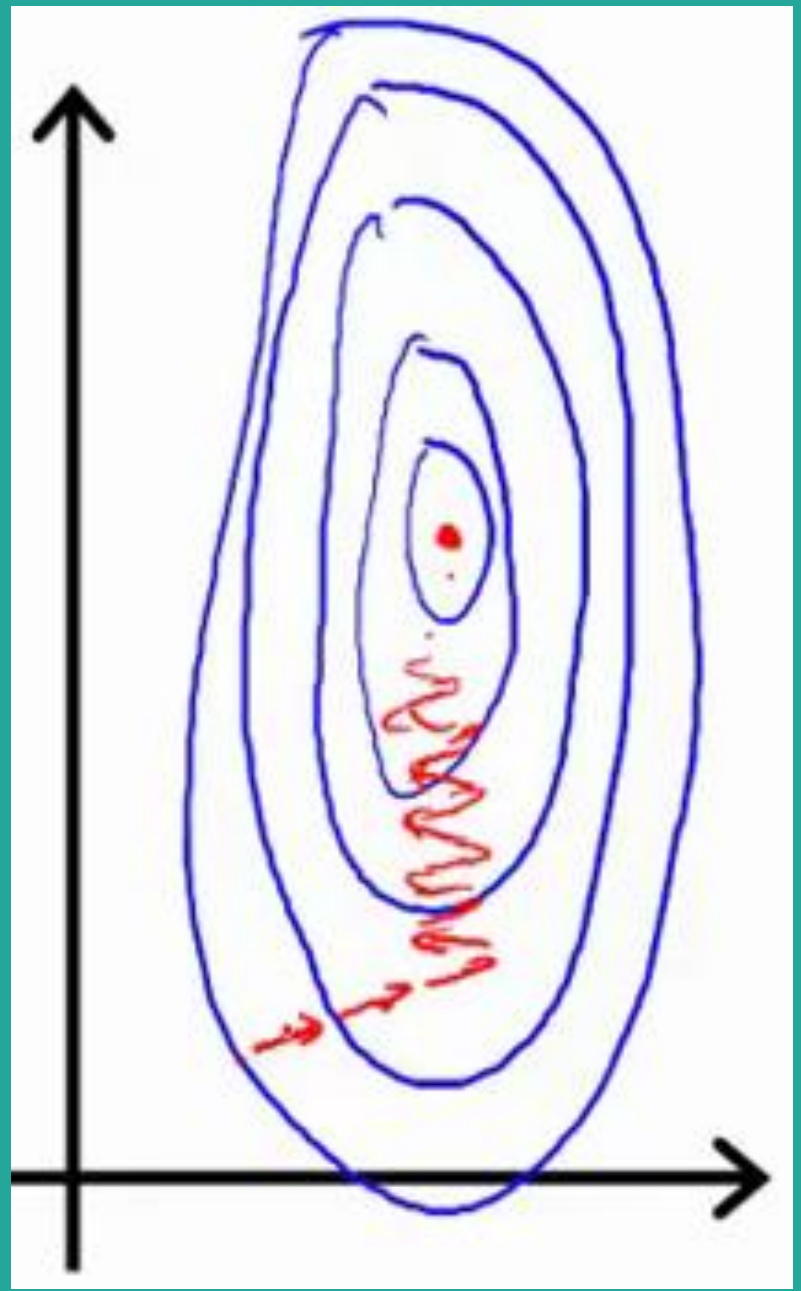
σ_y = is the standard deviation of all output values

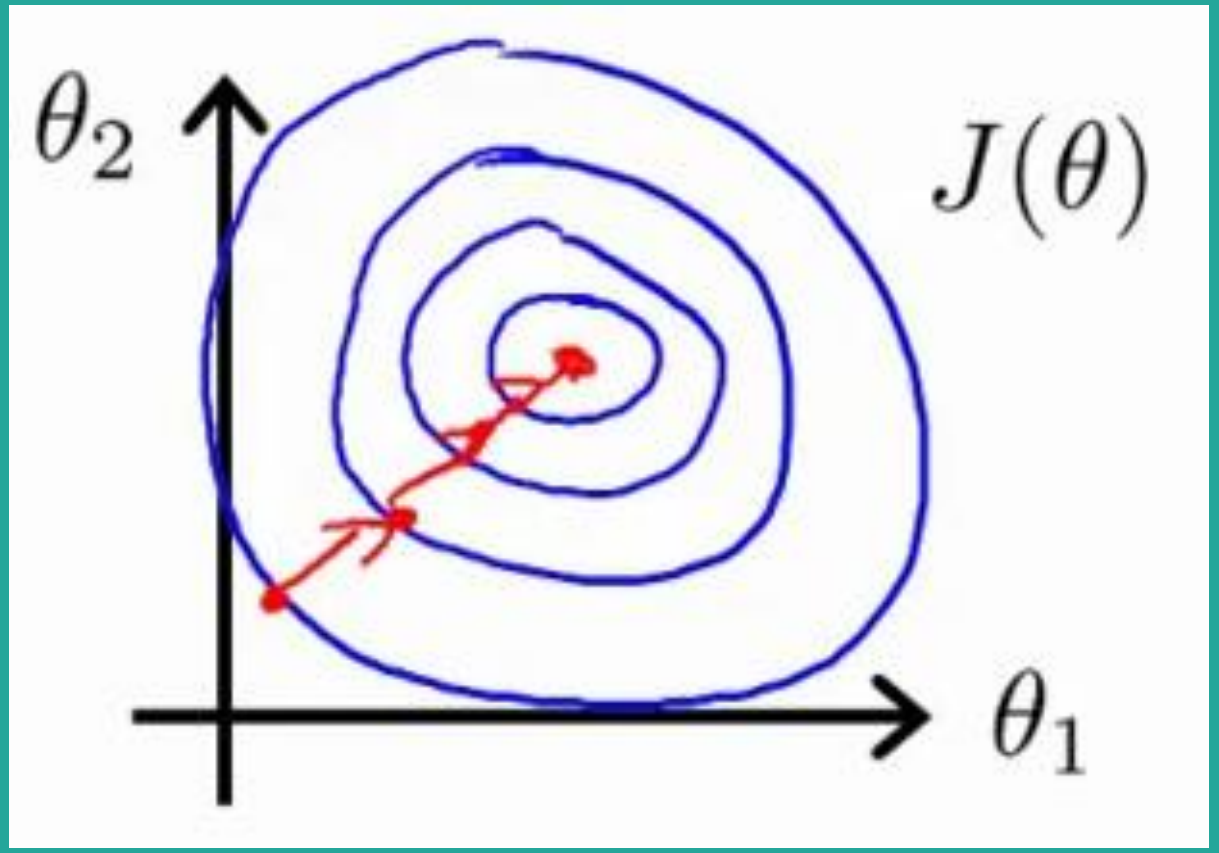
$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} x_j^{(i)}$$

$$\sigma_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(x_j^{(i)} - \mu_j\right)^2}$$

Applying z-score to all the values in the dataset ensures that the mean of each group is always 0 and the variance is always 1

Feature Scaling optimizes gradient descent by shortening the path to the minima




θ_2
$J(\theta)$
θ_1