Multiple Linear Regression

Model: $f_{w,b}(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$

$= \vec{w} \cdot \vec{x} + b$

Gradient Descent: $w_n = w_n - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) x_n$

$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)$$

Feature Scaling:

$price = w_1 x_1 + w_2 x_2 + b$ (x_1: size, x_2: rooms)

x_1 range: 300 - 2k x_2 range: 0 - 5

Because of the disparity in feature vals, gradient descent may be slow.

solution: feature scaling

Divide by max:

$x_1 \text{scaled} = \frac{x_1}{2k} \rightarrow x_1$ range: .15 - 1

$x_2 \text{scaled} = \frac{x_2}{5} \rightarrow x_2$ range: 0 - 1

Mean normalization:

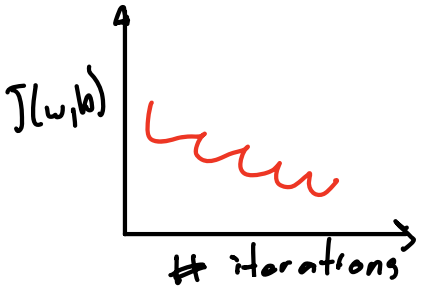
$x_1 = \frac{x_1 - \mu_1}{2000 - 300} \rightarrow -.18 - .82$

$x_2 = \frac{x_2 - \mu_2}{5 - 0} \rightarrow -.46 - .54$

Z-score normalization:

$$x_n = \frac{x_n - \mu_n}{\sigma_n}$$

Choosing the Learning Rate



learning rate might be too large

solution: use smaller α

values to try: .001 .01 0.1 1 ...

Feature Engineering

$$f_{w,b}(\vec{x}) = w_1 x_1 + w_2 x_2 + b$$

x_1: width, x_2: depth

area = width · depth

$$x_3 = x_1 \cdot x_2$$

$$f_{w,b}(\vec{x}) = w_1 x_1 + w_2 x_2 + w_3 x_3 + b$$