$\Rightarrow$ 6

→ Gradient Descent

- Feature Scaling

→ Max Division: $x_j = \dfrac{x_j}{\max(x_j)}$

$\mu \equiv$ mean

$\sigma \equiv$ standard dev

→ Mean Normalization : $x_j = \dfrac{x_j - \mu}{\max(x_j) - \min(x_j)}$

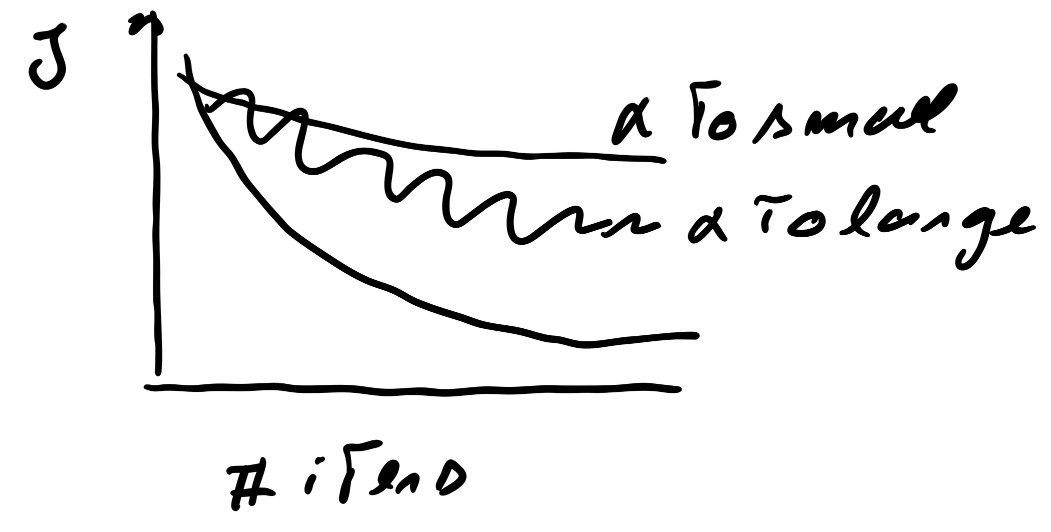
→ Z-Score Normalization : $x_j = \dfrac{x_j - \mu}{\sigma}$

- Aim for about $-1 \leq x_j \leq 1$ for each feature x_j

- Checking Gradient Descent for Convergence

$$\left.\begin{aligned} w_j &= w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w}, b) \\ b &= b - \alpha \frac{\partial}{\partial b} J(\vec{w}, b) \end{aligned}\right\} \text{Gradient Descent}$$

- Plot $J(\vec{w}, b) \times$ # iterations



- Automatic convergence test
  - Let $\varepsilon = 10^{-3}$
  - If $J(\vec{w}, b)$ decreases by $\leq \varepsilon \Rightarrow$ declare convergence

    $\Rightarrow$ Found parameters $\vec{w}, b$ close to global minimum.

- Choosing the Learning Rate
  - Learning rate to large $\Rightarrow$ overshoot the minimum
  - With a small ~~enough~~ enough α, $J(\vec{w}, b)$ should decrease on every iteration
  - Values α to try : 0.001 0.01 0.1 ...

- Feature Engineering
  - Using intuition to design new features, by transforming, or combining original features

$\to f_{\vec{w},b}(x) = \underbrace{w_1 x}_{\text{original}} + \underbrace{w_2 x^2 + w_3 x^3}_{\text{added}} + b$