$x_i = \frac{\text{Age} - 40}{20}$ $30 \le \text{Age} \le 50$

x_i ↓ input feature

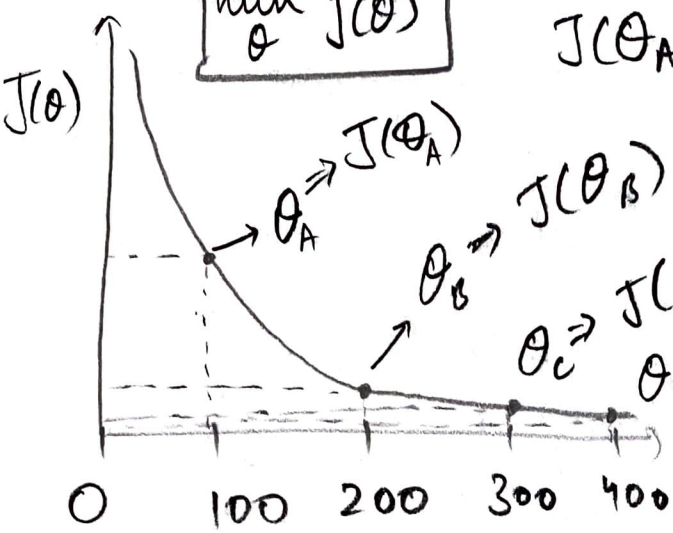
40 → avg Age

20 → range of Age

feature scaling can make gradient descent much faster

Debugging Gradient descent.

$\min_{\theta} J(\theta)$



$J(\theta_A) > J(\theta_B) > J(\theta_C) > J(\theta_D)$

$J(\theta)$ should dec↓ after every iteration

for this case $J(\theta)$ don't change significantly after 300 iterations

No. of iterations for gradient descent to get $\min_{\theta} J(\theta)$ can vary significantly

⇒ Eg of a automatic convergence test:

Declare convergence if $J(\theta)$ is less than 10^{-3} in one iterations

↳ this threshold is not definitive and plotting $J(\theta)$ v/s # iteration graph is better to declare convergence.