

GRADIENT DESCENT

RULE OF THUMB

Gradient descent will take longer if features are not similarly scaled.

Andrew Ng's rule is that if feature values range from -3 to 3 or $-\frac{1}{3}$ to $\frac{1}{3}$ then it is okay.

Chris Albon