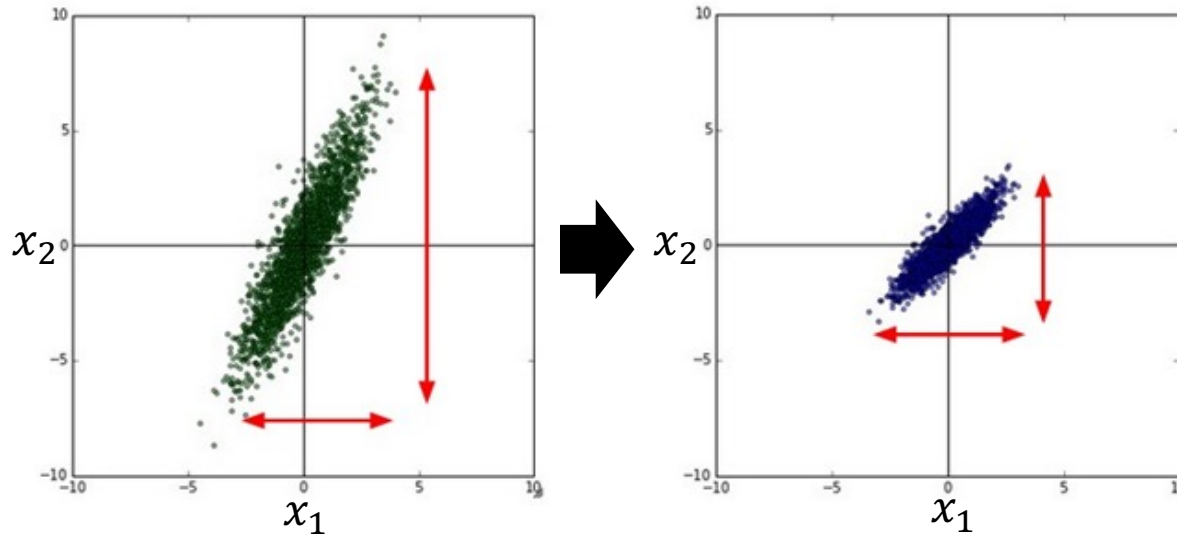# Gradient Descent Tips

Tip Three: Feature Scaling

Feature Scaling: Part I
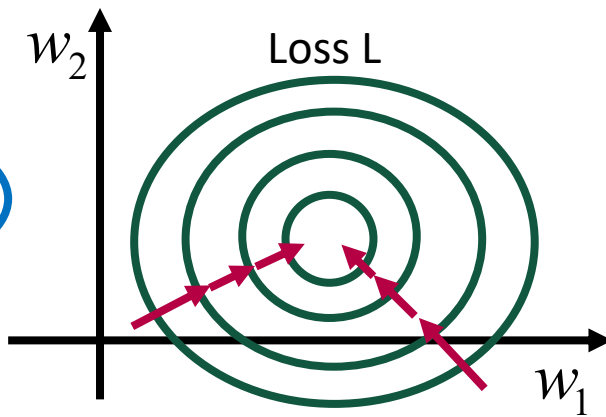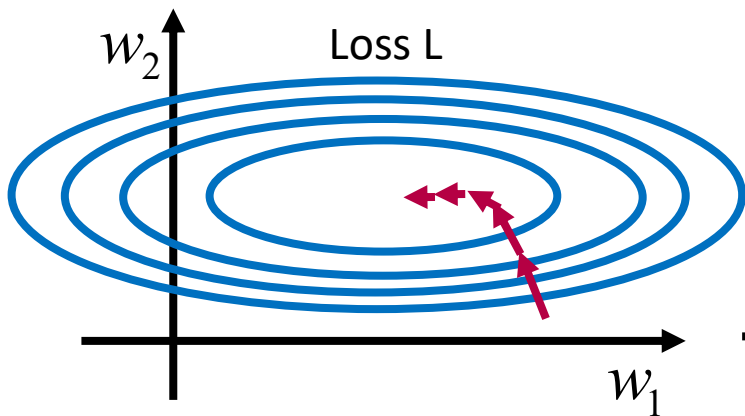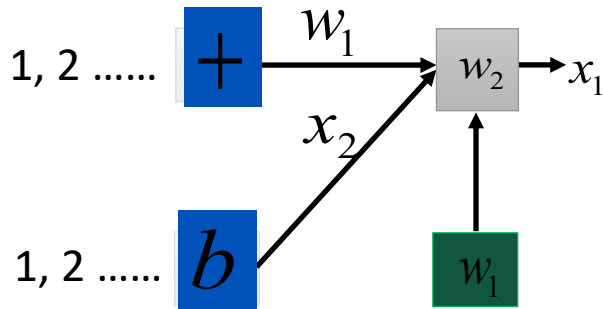
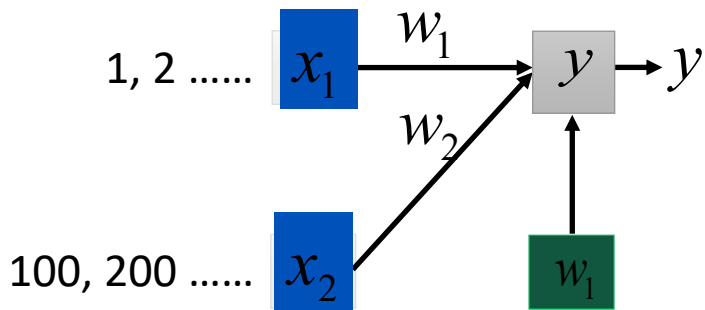$$y = b + w_1 x_1 + w_2 x_2$$

Make different features have the same scaling

$$y = b + w_1 x_1 + w_2 x_2$$

Feature Scaling: Part II

Feature Scaling: Part III

x^1 x^2 x^3 x^r x^R

x_1^1 x_1^2

x_2^1 x_2^2

$\cdots$ $\cdots$

$\cdots$

$\cdots$ $\cdots$
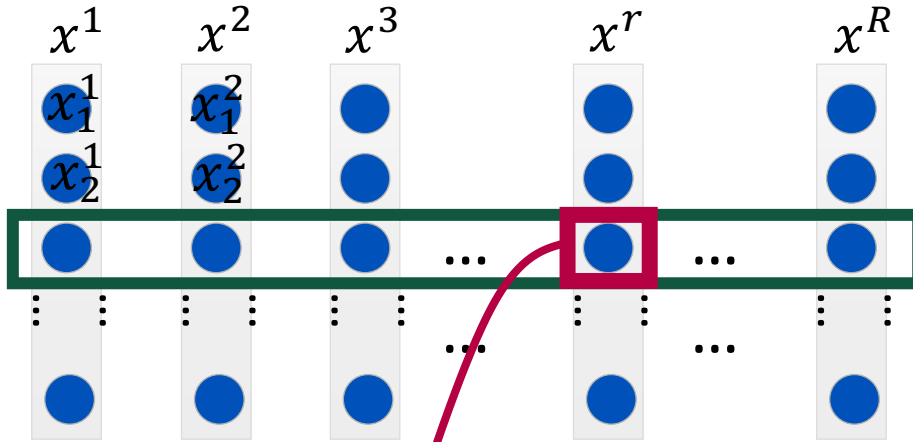
For each dimension i:

mean: m_i

standard deviation: σ_i

$$x_i^r \leftarrow \frac{x_i^r - m_i}{\sigma_i}$$

The means of all dimensions are 0, and the variances are all 1.