Linear Regression.

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$.

Parameters: θ_1, θ_0.

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$ or any other cost function.

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$.

Gradient descent algorithm:

repeat until convergence {

$$\theta_j = \theta_j - \underbrace{\alpha}_{\text{learning rate}} \cdot \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0, 1)$$

}.

Simultaneously update θ_0 and θ_1.

If α is too small, gradient descent can be slow;
if α is too large, gradient descent can overshoot the minimum; it may fail to converge, or even diverge.

Gradient descent can converge to a local minimum, even with the learning rate α fixed.

Convex function, is bowl-shape function, has global minimum, no local minimum

Batch Gradient descent: each step of gradient descent uses all the training examples

Linear algebra Review

Vector: an $n \times 1$ matrix.

Matrix Addition:

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0.5 \\ 4 & 10 \\ 3 & 2 \end{bmatrix}$$

They should have the same dimension.

Scalar Multiplication:

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} \times 3 = \begin{bmatrix} 3 & 0 \\ 6 & 15 \\ 9 & 3 \end{bmatrix}$$

Properties:

$A \times B \neq B \times A$. not commutative

$A \times B \times C = (A \times B) \times C = A \times (B \times C)$. Associative.

Identity Matrix, denoted I or $I_{n \times n}$. $\begin{bmatrix} 1 & & \\ & 1 & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$

For any matrix A, $A \cdot \underline{I} = \underline{I} \cdot A = A$.

I could be in different dimensions.

Matrix inverse: if A is an $m \times m$ matrix, and if it has an inverse,

$$A \cdot A^{-1} = A^{-1} \cdot A = I.$$

Matrices that don't have an inverse are "singular" or "degenerate".

Matrix Transpose: Let A be an $m \times n$ matrix, and let $B = A^T$, then B is an $n \times m$ matrix, and $B_{ij} = A_{ji}$.