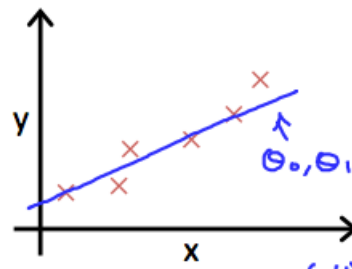
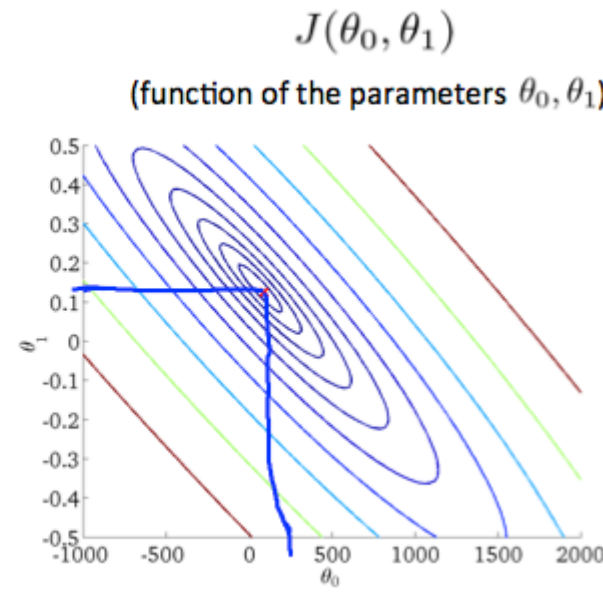
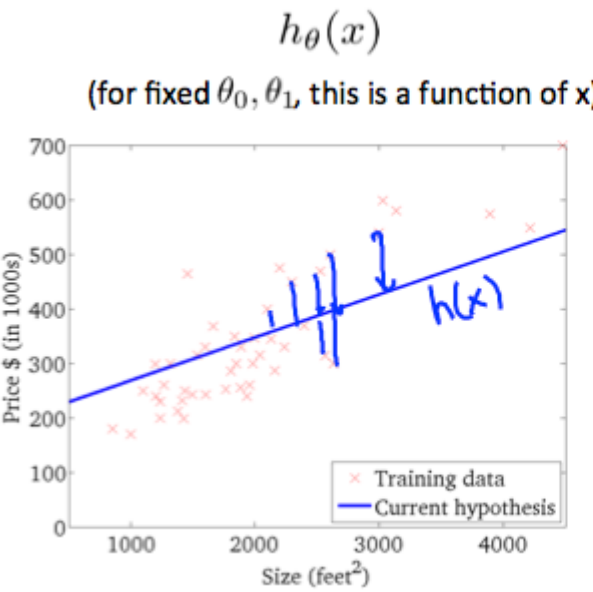
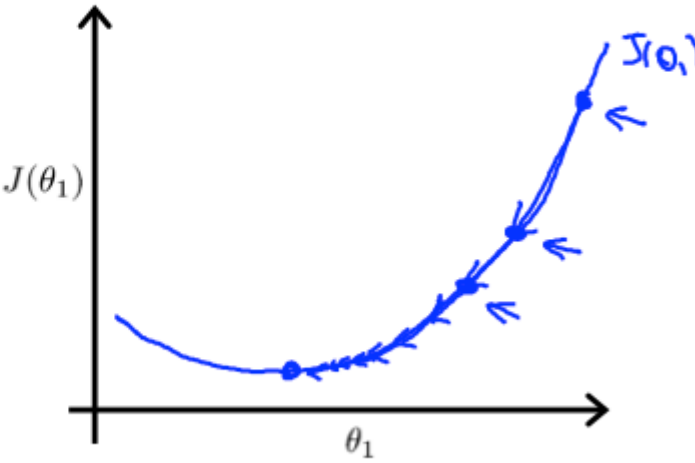


Gradient descent can converge to a local minimum, even with the learning rate α fixed.

$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

As we approach a local minimum, gradient descent will automatically take smaller steps. So, no need to decrease α over time.



Idea: Choose θ_0, θ_1 so that $h_\theta(x)$ is close to y for our training examples (x, y)

minimize $\frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$

$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$

$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$

Cost function
Squared error function

单变量回归

梯度下降直观

代价方程直观2

代价方程

线性回归的梯度下降

梯度下降

代价方程直观1

模型表示

$$\begin{aligned} \frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2 \\ &= 2 \cdot \frac{1}{2} (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (h_\theta(x) - y) \\ &= (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} \left(\sum_{i=0}^n \theta_i x_i - y \right) \\ &= (h_\theta(x) - y) x_j \end{aligned}$$

Correct: Simultaneous update

$\rightarrow \text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$

$\rightarrow \text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$

$\rightarrow \theta_0 := \text{temp0}$

$\rightarrow \theta_1 := \text{temp1}$

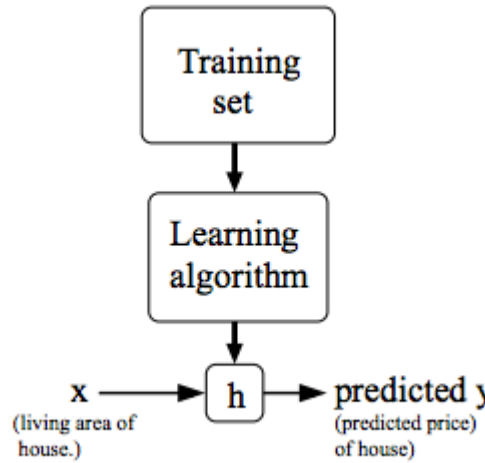
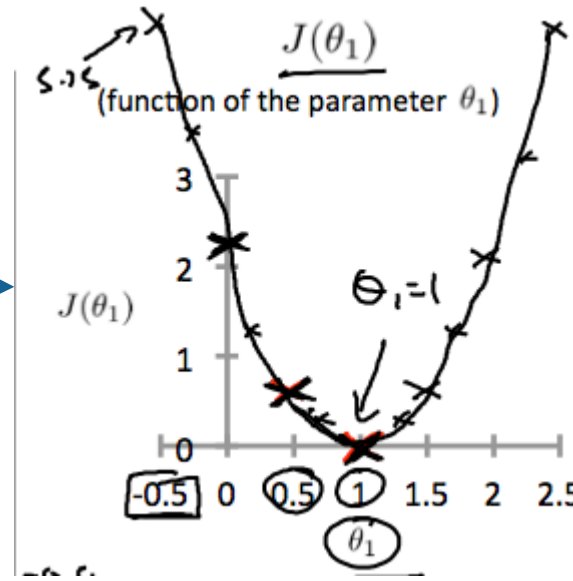
Incorrect:

$\rightarrow \text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$

$\rightarrow \theta_0 := \text{temp0}$

$\rightarrow \text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$

$\rightarrow \theta_1 := \text{temp1}$



Matrix Transpose

Example: $A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 4 & 9 \end{bmatrix}_{2 \times 3}$ $B = A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 0 & 9 \end{bmatrix}_{3 \times 2}$

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

Identity Matrix

Denoted I (or $I_{n \times n}$).
Examples of identity matrices:

1×1 $\begin{bmatrix} 1 \end{bmatrix}$ 2×2 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 3×3 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 4×4 $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

For any matrix A ,
 $A \cdot I = I \cdot A = A$

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

$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix}_{3 \times 2} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \end{bmatrix}_{2 \times 2} = \text{error}$

Matrix Elements (entries of matrix)

$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 949 & 1437 \\ 147 & 1448 \end{bmatrix}$

A_{ij} = " i, j entry" in the i^{th} row, j^{th} column.

线性代数复习

向量

Vector: An $n \times 1$ matrix.

$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$

$n = 4$
 $\leftarrow 4 - 1$

$y_i = i^{th}$ element

逆矩阵

Not all numbers have an inverse. 0 (0^{-1}) undefined

Matrix inverse: square matrix ($\# \text{ rows} = \# \text{ columns}$) A^{-1}
If A is an $m \times m$ matrix, and if it has an inverse,
 $A(A^{-1}) = A^{-1}A = I$

E.g. $\begin{bmatrix} 3 & 4 \\ 2 & 16 \end{bmatrix}_A \begin{bmatrix} 0.4 & -0.1 \\ -0.05 & 0.075 \end{bmatrix}_{A^{-1}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_{2 \times 2}$

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

矩阵乘法

$A * B * C = D * C = A * E$ ($D = A * B, E = B * C$)

$A * B \neq B * A$

矩阵scalar

$\begin{bmatrix} \text{rows} \\ \text{columns} \end{bmatrix}_A \times \begin{bmatrix} \text{columns} \\ \text{rows} \end{bmatrix}_x = \begin{bmatrix} \text{rows} \\ \text{columns} \end{bmatrix}_y$

$m \times n$ matrix (m rows, n columns) \times $n \times 1$ matrix (n -dimensional vector) = m -dimensional vector

Details:

$\begin{bmatrix} \text{rows} \\ \text{columns} \end{bmatrix}_A \times \begin{bmatrix} \text{columns} \\ \text{rows} \end{bmatrix}_B = \begin{bmatrix} \text{rows} \\ \text{columns} \end{bmatrix}_C$

$m \times n$ matrix (m rows, n columns) \times $n \times o$ matrix (n rows, o columns) = $m \times o$ matrix

The i^{th} column of the matrix C is obtained by multiplying A with the i^{th} column of B . (for $i = 1, 2, \dots, o$)

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

$\begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix}_{2 \times 2} / 4 = \frac{1}{4} \begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix}_{2 \times 2}$