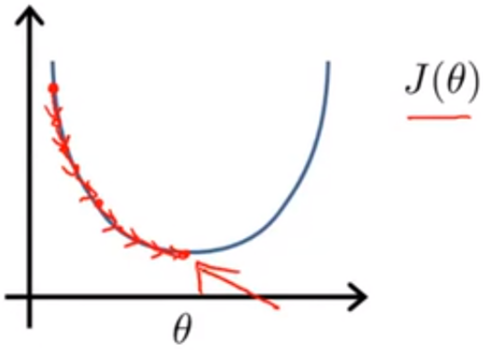
使用标准方程法求解参数 θ 的最优值，使用迭代法可以一次性求解 θ 的最优值，可以一步得到最优值。

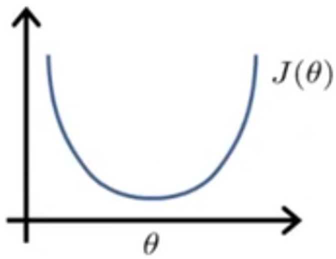
Gradient Descent



举例

Intuition: If 1D $(\theta \in \mathbb{R})$

$$J(\theta) = a\theta^2 + b\theta + c$$



我们假设有一个非常简单的代价函数 J(θ)，假设 θ 是一个数字不是向量，J 是这个实参数 θ 的二次函数，如何最小化一个二次函数？---求导，并将导数置零，对 J 关于 θ 的导数，求出使得 J(θ)最小的 θ 值。

当 Θ 不是实数的时候，它是一个 n+1 维的参数向量,代价函数 J 是这个向量的函数，也就是 θ0 到 θm 的函数，像下图右边这个平方代价函数。

$$\theta \in \mathbb{R}^{n+1} \qquad J(\theta_0, \theta_1, \ldots, \theta_m) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$\frac{\partial}{\partial \theta_j} J(\theta) = \cdots = 0 \quad \text{(for every } j\text{)}$$

Solve for $\theta_0, \theta_1, \ldots, \theta_n$

Theta m. And, a cost

一个代价函数看起来是这样

通过对每个参数 θ 求 J 得偏导数，然后把它们全部置零，求出 θ0，θ1 一直到 θn 的值，能够最小化代价函数 J 的 θ 值。等价的这个 θ 能够使得代价函数 J(θ)最小化。

举例：

Examples: $m = 4$.

Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) y
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

假设有一个 m=4 个训练样本，假设这两个训练样本就是我的所有数据，在我的训练数集中加上一列对应额外特征变量的 x0

Examples: $m = 4.$

	Size (feet²)	Number of bedrooms	Number of floors	Age of home (years)	Price ($1000)
x_0	x_1	x_2	x_3	x_4	y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

接下来构建矩阵，矩阵包含了训练样本的所有特征变量

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix}$$

这里有我所有的特征变量 x，我们将这些数字全部放到矩阵中，对 y 构建一个向量 y

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

所以 X 会是一个 M*(n+1)维矩阵,y 会是一个 m 维向量，其中 m 是训练样本数量，n 是特征变量数,n+1 是因为我加的这个额外的特征变量 x0,如果用矩阵 X 和向量 y 来计算这个，θ 等于 x 转置乘以 X 的逆乘以 X 的转置乘以 y。这样就能够使得代价函数最小化 θ

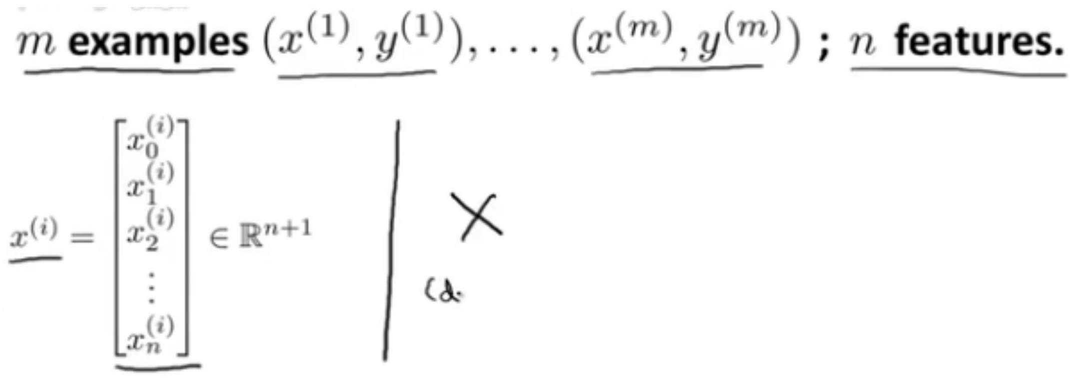
Examples: $m = 4.$

	Size (feet²)	Number of bedrooms	Number of floors	Age of home (years)	Price ($1000)
x_0	x_1	x_2	x_3	x_4	y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

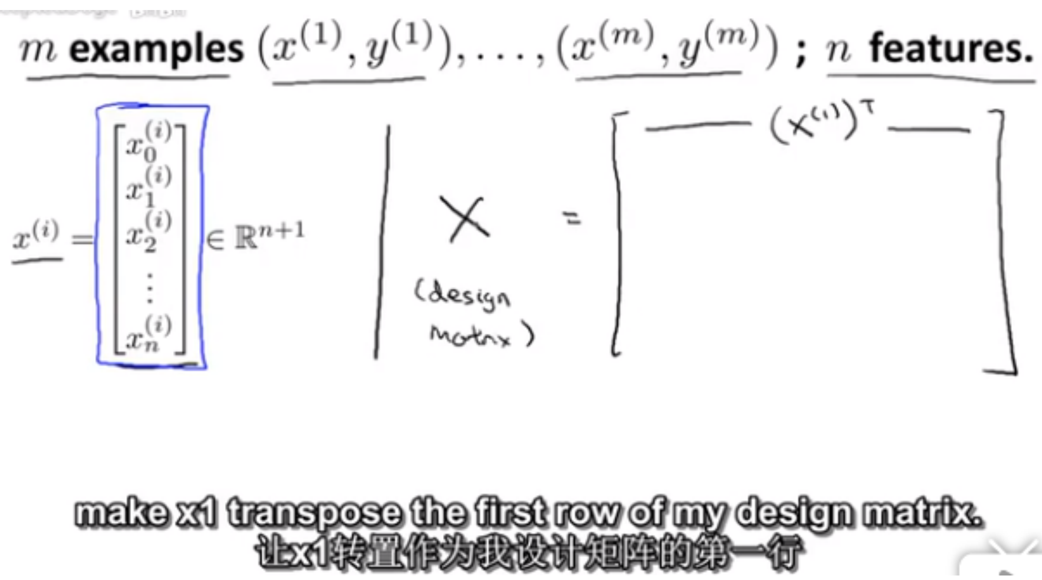
$$X = \underbrace{\begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix}}_{m \times (n+1)} \qquad y = \underbrace{\begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}}_{m\text{-dimensional vect}}$$

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

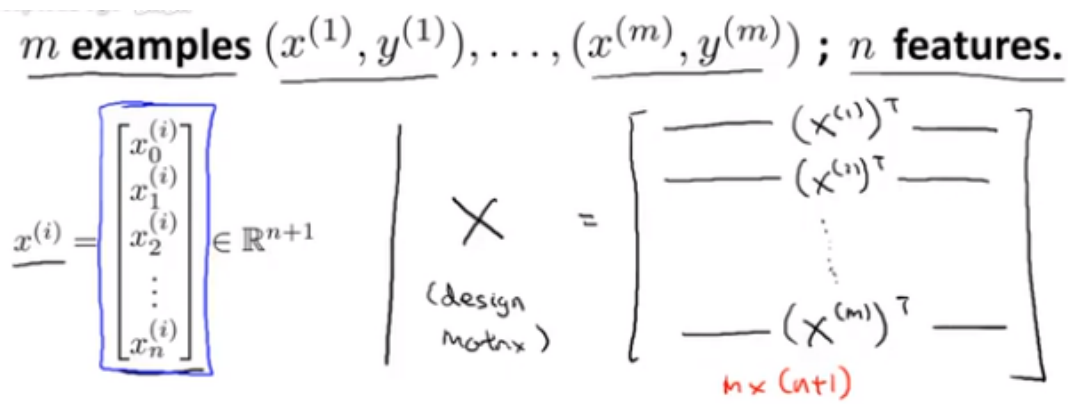
假如我们有 M 个训练样本，x1y1 直到 xnyn，n 个特征变量,每一个训练样本 xi 可能看起来像一个向量,像这样一个 n+1 维特征向量，我们要构建矩阵 X 的方法，也叫设计矩阵。



每个训练样本给出一个这样特征向量，这样的 n+1 维向量，构建设计矩阵 • X 的方法，取第一个训练样本，一个向量，取他的转置，让 x1 的转置作为我设计矩阵的第一行



然后将 x2 的转置作为 X 的第二行，以此类推直到最后一个训练样本 xn，取它的转置作为矩阵 X 的最后一行，这个矩阵就是一个 m*(n+1)维矩阵。



24.4-6