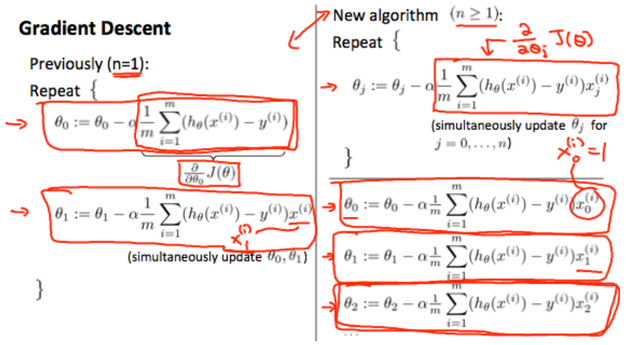
多变量线性回归

- 矩阵计算参数
  - 矩阵不可逆
    - 太多特征(m <= n) -> 正则化
    - 重复的特征
  - 参数求解公式
    - $\theta = (X^T X)^{-1} X^T y$
  - 与梯度下降区别
    - 当n>10K时，梯度下降也能够工作的很好
    - O(kn^2)与O(n^3)
    - 需要许多迭代步骤
    - 需要步长alpha
- 多变量的梯度下降



- 多种特征

$$h_\theta(x) = [\theta_0 \quad \theta_1 \quad \ldots \quad \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \theta^T x$$

- 特征标准化
  - 方法
    - mean normalization: (x − mu) / sigma or range(x), 除数可能是标准差也可能是range差值
    - feature scaling: x / (max(x) - min(x))
  - 加快theta梯度下降的速度
- 学习步长
  - 确保损失函数一直变小, alpha太小收敛较慢；太大可能不收敛
  - 注意特征标准化
- 多项式回归
  - 不是直线，而是曲线
  - 特征转换, x1, x2=x1^2, x3 = sqrt(x1), x4 = x1^3, x5 = x1 * x2