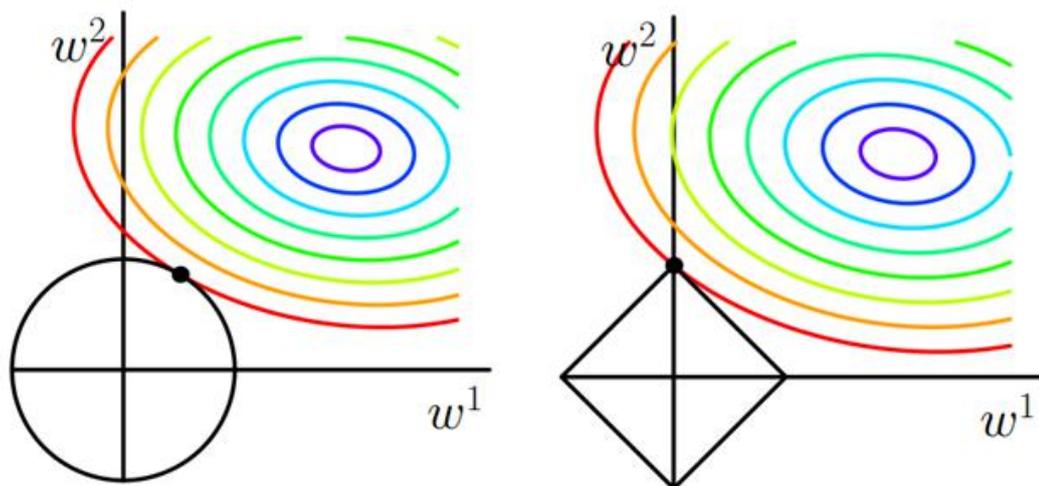
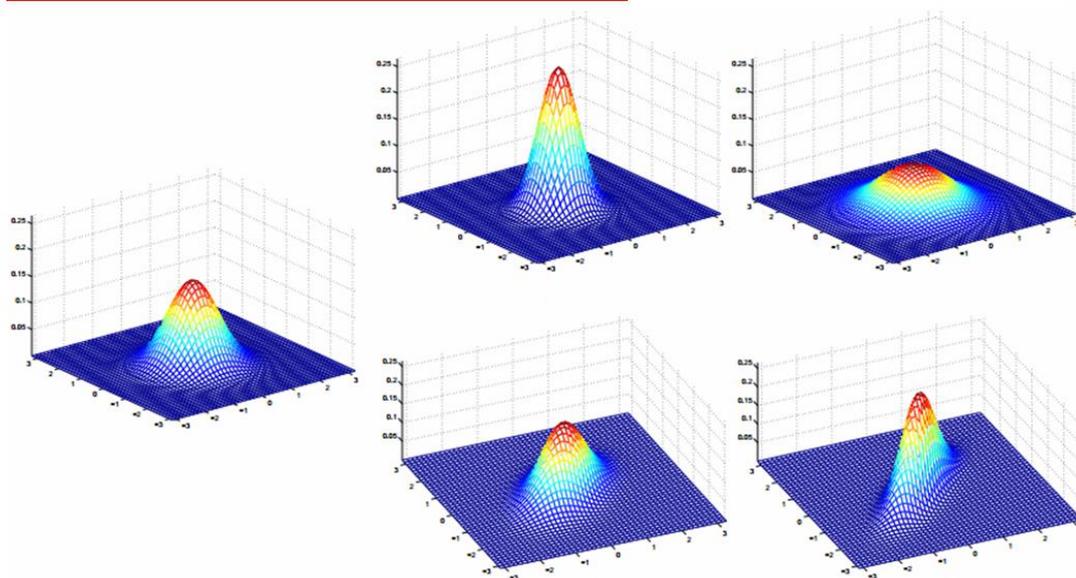


正则化

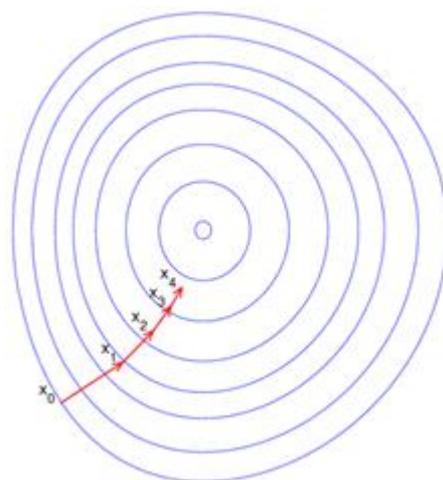
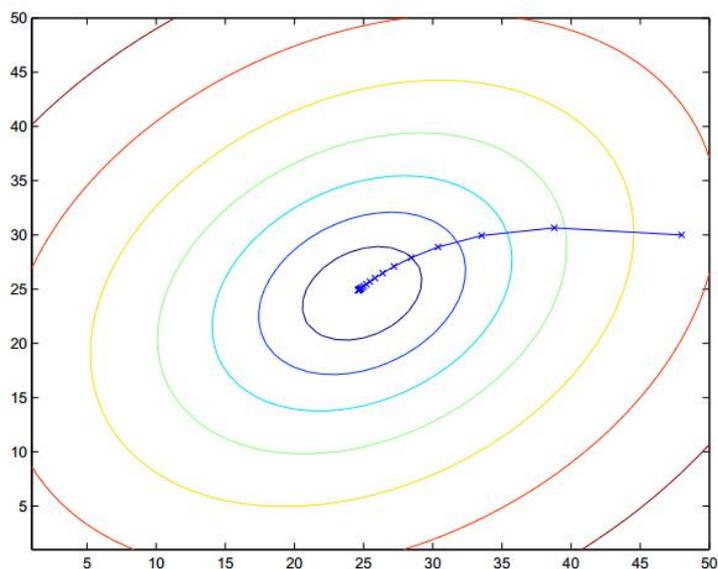


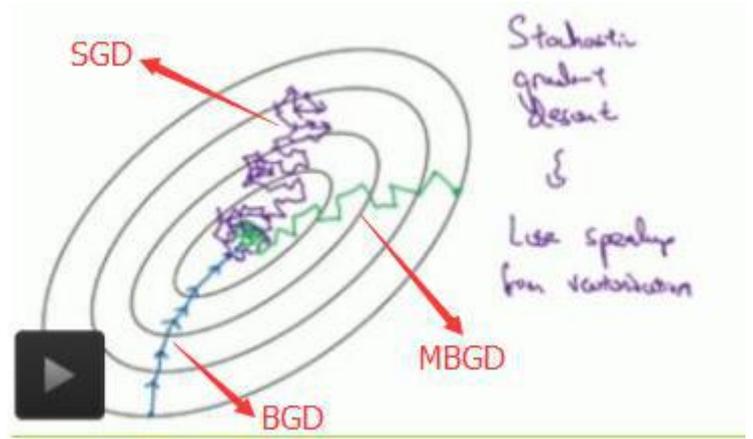
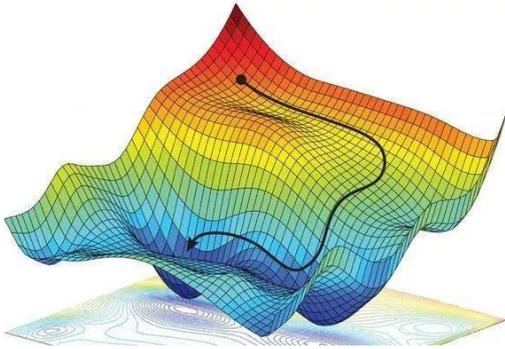
数据分布

## 二元正态分布



损失的 梯度下降





特征缩放

例如:  $x_1 = \text{size (0-2000 feet}^2\text{)}$   
 $x_2 = \text{number of bedrooms (1-5)}$

方法一: ↵

$$x_1 = \frac{x_1}{2000} \qquad x_2 = \frac{x_2}{5}$$

方法二: ↵

$$x_j^{(i)} = \frac{x_j^{(i)} - \text{avg}(x_j)}{\text{max}(x_j) - \text{min}(x_j)}$$

**Feature Scaling**

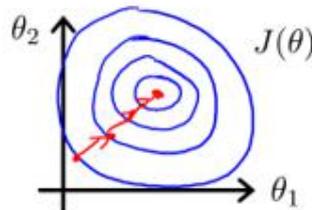
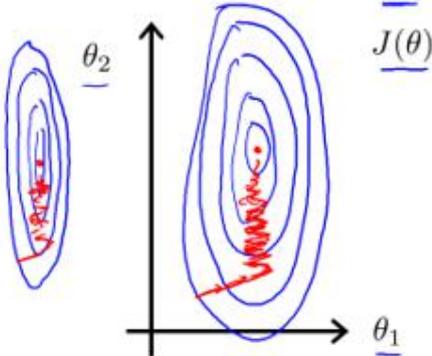
Idea: Make sure features are on a similar scale.

E.g.  $x_1 = \text{size (0-2000 feet}^2\text{)}$  ↵  
 $x_2 = \text{number of bedrooms (1-5)}$  ↵

$$\rightarrow x_1 = \frac{\text{size (feet}^2\text{)}}{2000} \quad \leftarrow$$

$$\rightarrow x_2 = \frac{\text{number of bedrooms}}{5} \quad \leftarrow$$

$$0 \leq x_1 \leq 1 \quad 0 \leq x_2 \leq 1$$



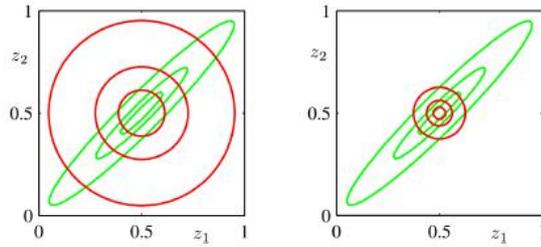
Andrew Ng ↵

## 两个KL散度的区别

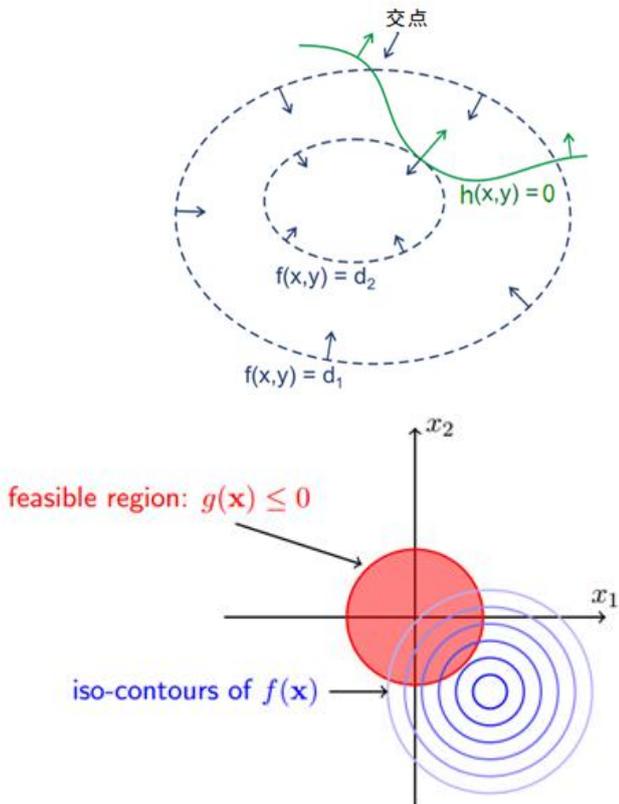
□ 绿色曲线是真实分布 $p$ 的等高线；红色曲线是使用近似 $p(z_1, z_2) = p(z_1)p(z_2)$ 得到的等高线

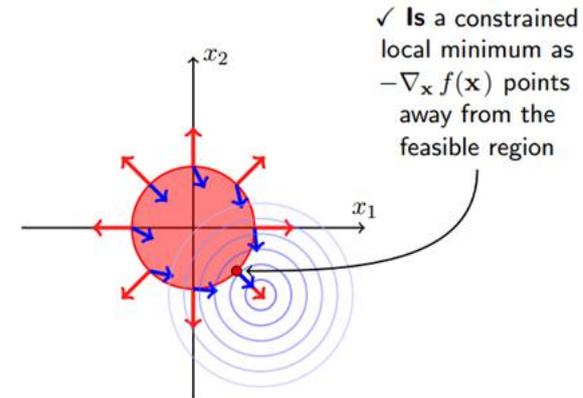
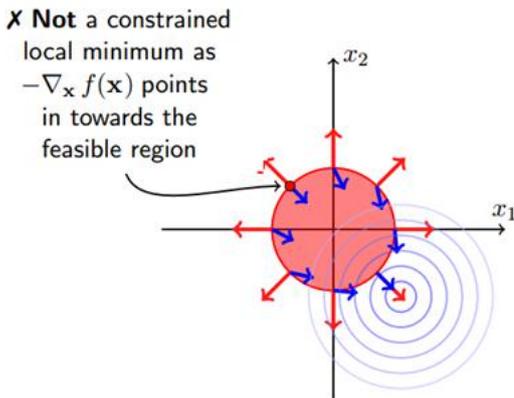
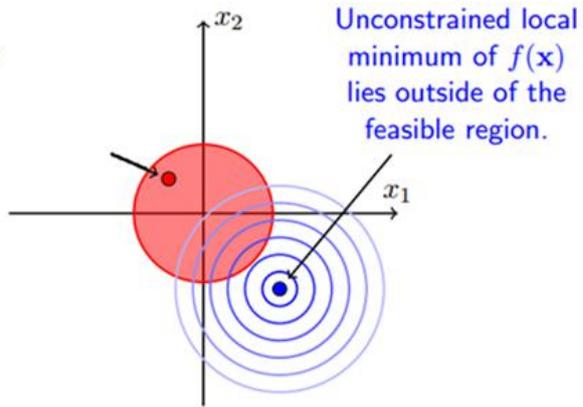
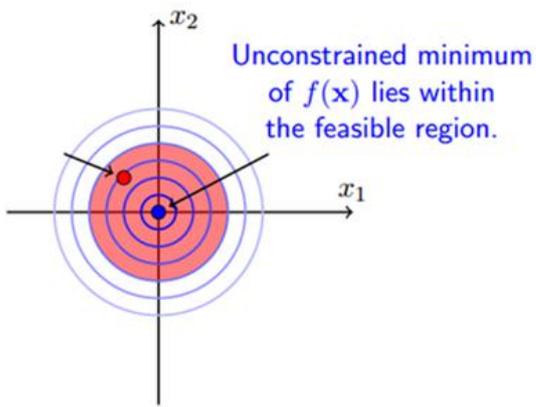
■ 左:  $KL(p||q)$ : zero avoiding

■ 右:  $KL(q||p)$ : zero forcing



拉格朗日乘子法





### 等高线的绘制

