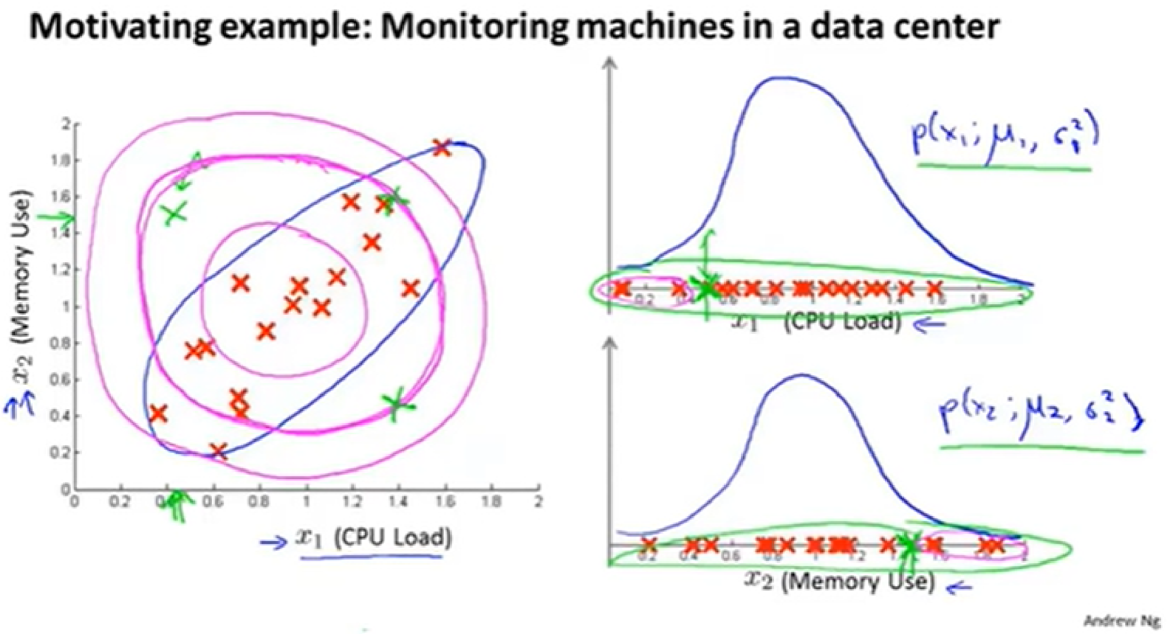
**单元高斯分布问题：**




在检测计算机的问题中，当CPU负载高，内存使用也应该高，这两个变量是线性增长的关系。所以绿色的点应该是异常点，但在两个变量的高斯分布中，显示他们是异常点的概率并不高，所以这种高斯分布不能识别出异常点，相反，他们会将最里面的粉圈视为概率最高。

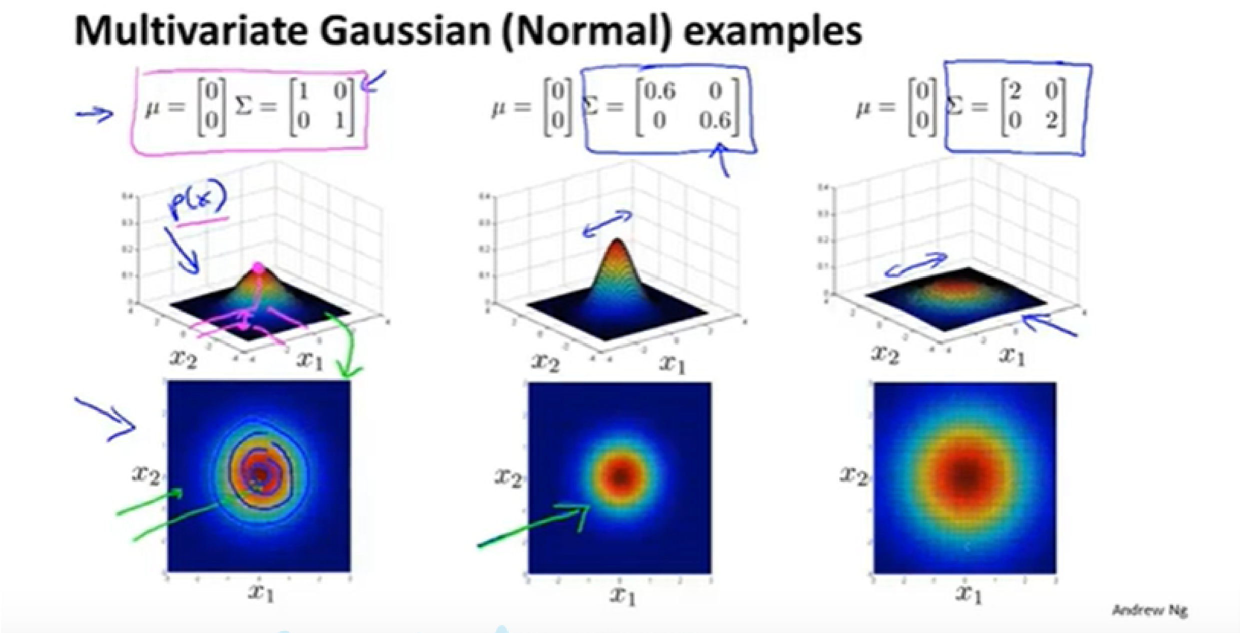
**Multivariate Gaussian (Normal) distribution**

$x \in \mathbb{R}^n$. Don't model $p(x_1), p(x_2), \ldots$, etc. separately.
Model $p(x)$ all in one go.
Parameters: $\mu \in \mathbb{R}^n$, $\boxed{\Sigma \in \mathbb{R}^{n \times n}}$ (covariance matrix)

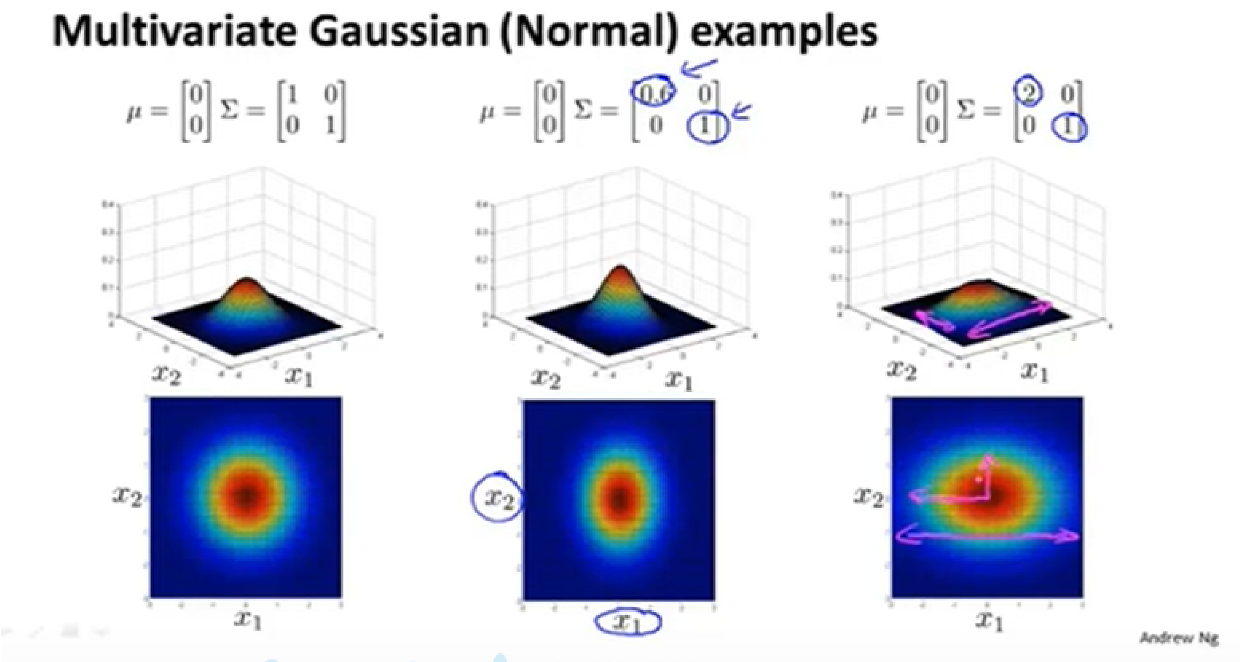
$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$|\Sigma|$ = determinant of $\Sigma$ | det (Sigma)

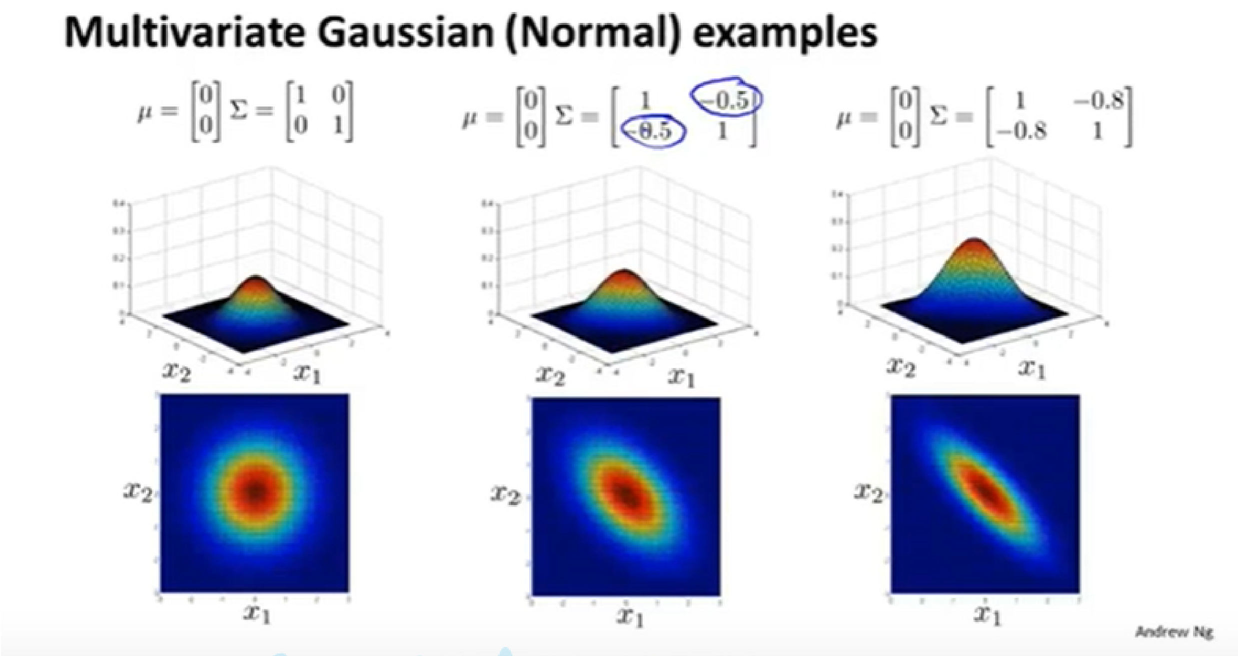
钟形的体积不变，当协方差矩阵中参数的值更小，钟形的高度更高，宽度更小。



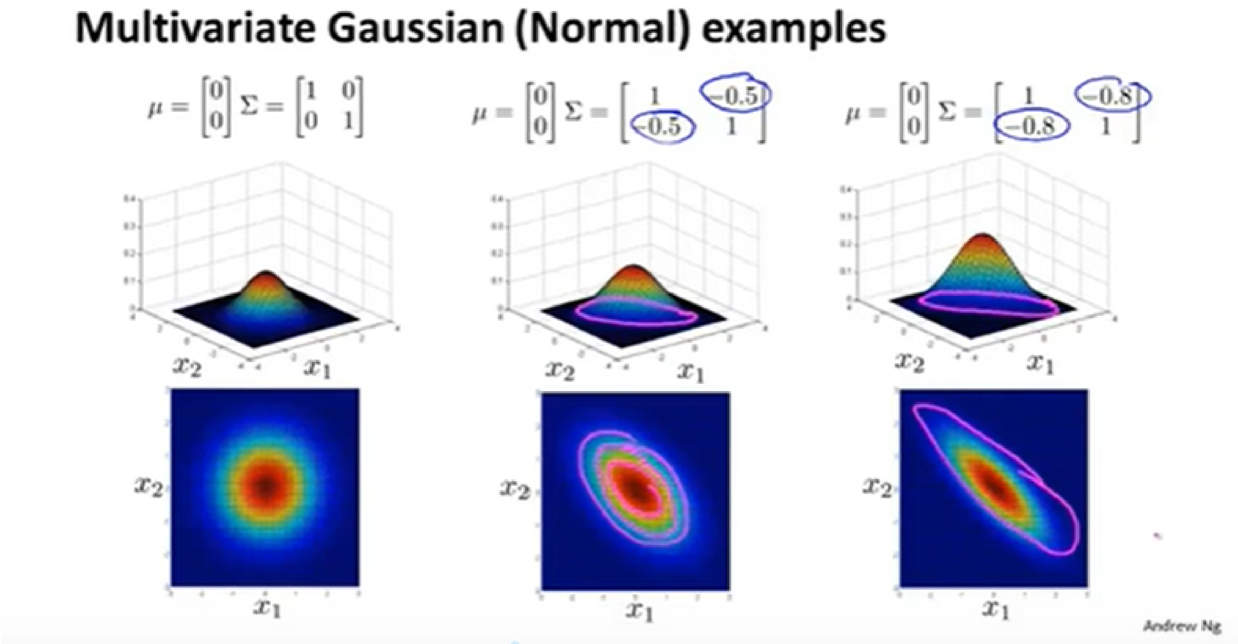

当改变矩阵中的一个值，一个变量的变化范围产生变化。



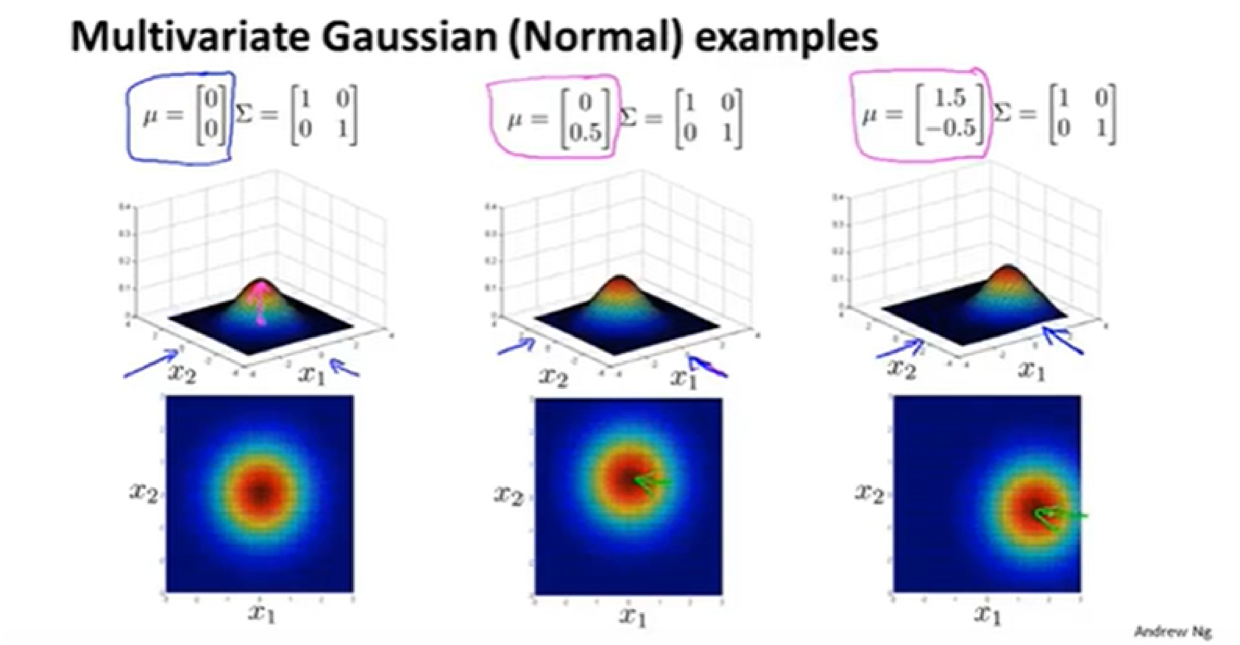

当改变非对角线上的元素，x和y正相关，随着非对角线上的值增大，区域更加狭窄，有更高更细的分布



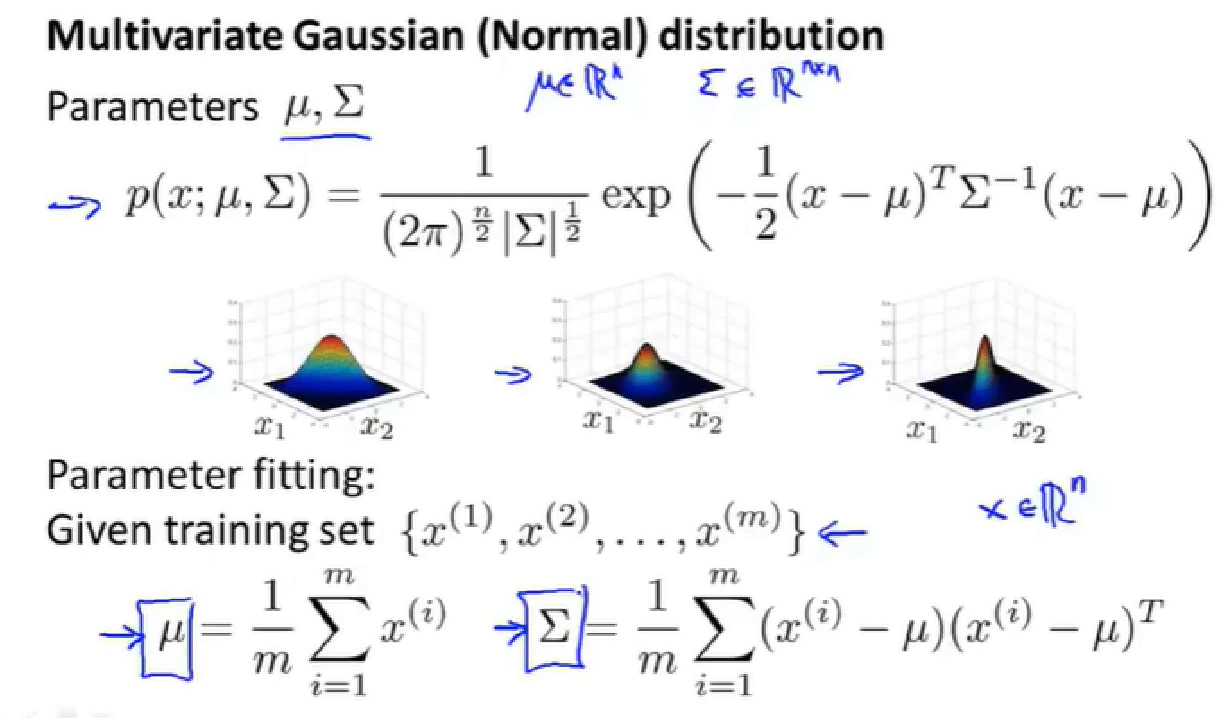

当非对角线上的元素为负，x和y负相关




当改变μ，会改变峰值产生的位置




**多元高斯分优势在于能描述两个特征变量之间的正相关，负相关的情况**



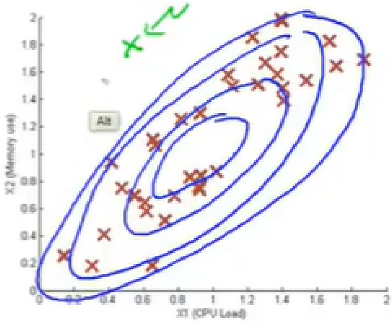

**多元高斯分布应用于异常检测**

## Anomaly detection with the multivariate Gaussian

1. Fit model $p(x)$ by setting

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu)(x^{(i)}-\mu)^T$$




2. Given a new example $x$, compute

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

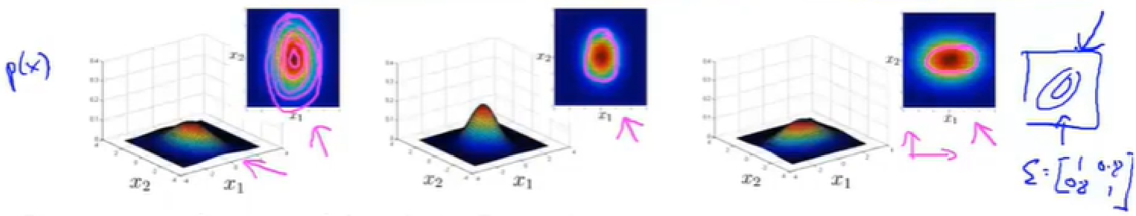
Flag an anomaly if $p(x) < \varepsilon$

Andrew Ng

异常点计算出的概率小于给定值，所以可以检测出这个异常点。

## Relationship to original model

Original model: $p(x) = p(x_1;\mu_1,\sigma_1^2)\times p(x_2;\mu_2,\sigma_2^2)\times\cdots\times p(x_n;\mu_n,\sigma_n^2)$




Corresponds to multivariate Gaussian

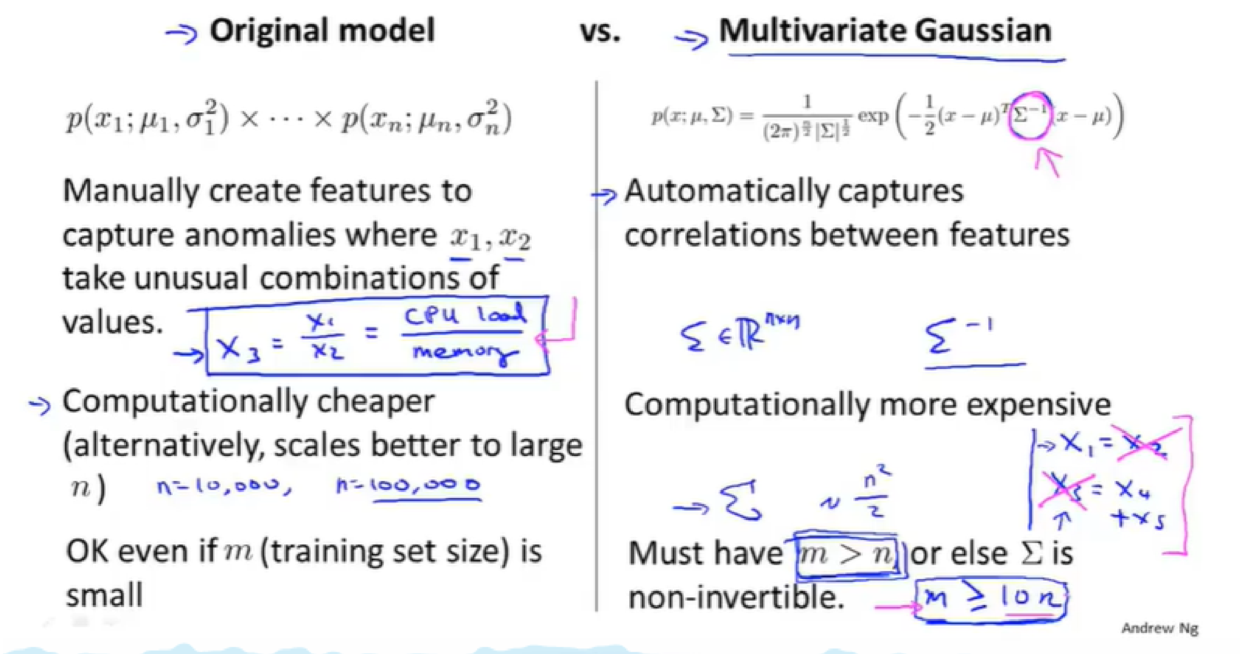
$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

where $\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ & & \ddots \\ 0 & & & \sigma_n^2 \end{bmatrix}$

Andrew Ng

当非对角线上的元素都是0时，多元高斯分布的模型相当于各个单元高斯分布模型得出的概率相乘。

## 高斯分布模型 vs 多元高斯分布模型




1.原始模型需要手动添加一些特征(表明特征之间关系的新特征)来捕捉一些特殊的异常信息，比如新添加CPU负载和内存比例的特征值。而多元高斯分布模型可以自动捕捉特征之间的关系。

2.原始模型计算负担更小，对于n很大的情况比较友好。而多元高斯分布模型计算成本高，sigma矩阵是n*n的。

3.原始模型即使训练集很小，也可以运行。但是多元高斯分布中，必须有m(训练集大小) > n(特征维度)，否则sigma矩阵就是奇异矩阵，不可逆。

一般当m >= 10n的时候，采用多元高斯分布模型。当多元高斯分布模型出现sigma矩阵不可逆时，一般是**m < n**，或者是出现了**冗余特征**。