

## 高斯分布，正态分布

x服从高斯分布，均值是 $\mu$ ，方差是 $\sigma^2$ ，标准差(standard deviation)是 $\sigma$

$$X \sim N(\mu, \sigma^2)$$

高斯分布概率密度

$$P(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

## 参数估计 parameter estimation

数据集： $\{x_1, x_2, \dots, x_m\}$

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$
$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)^2$$

## Anomaly detection algorithm 异常检测算法

1. 选择样本 $x_i$ ，你认为可能是异常的值
2. 计算 $\mu_1, \mu_2, \dots, \sigma_1, \sigma_2, \dots$
3. 对于例子 $x$ ，计算 $p(x)$

$$p(x) = \prod_{j=1}^n p(x_j; \mu_j, \sigma_j^2)$$

4. 如果 $p(x) < \epsilon$ ，就是异常的

## 多元高斯分布

多元高斯分布函数：

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

其中的  $\Sigma$  代表变量  $X$  的协方差矩阵， $i$ 行和 $j$ 列的元素值表示  $x_i$ 与 $x_j$ 的协方差

给定一组数据集， $\{x_1, x_2, \dots, x_m\}$

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