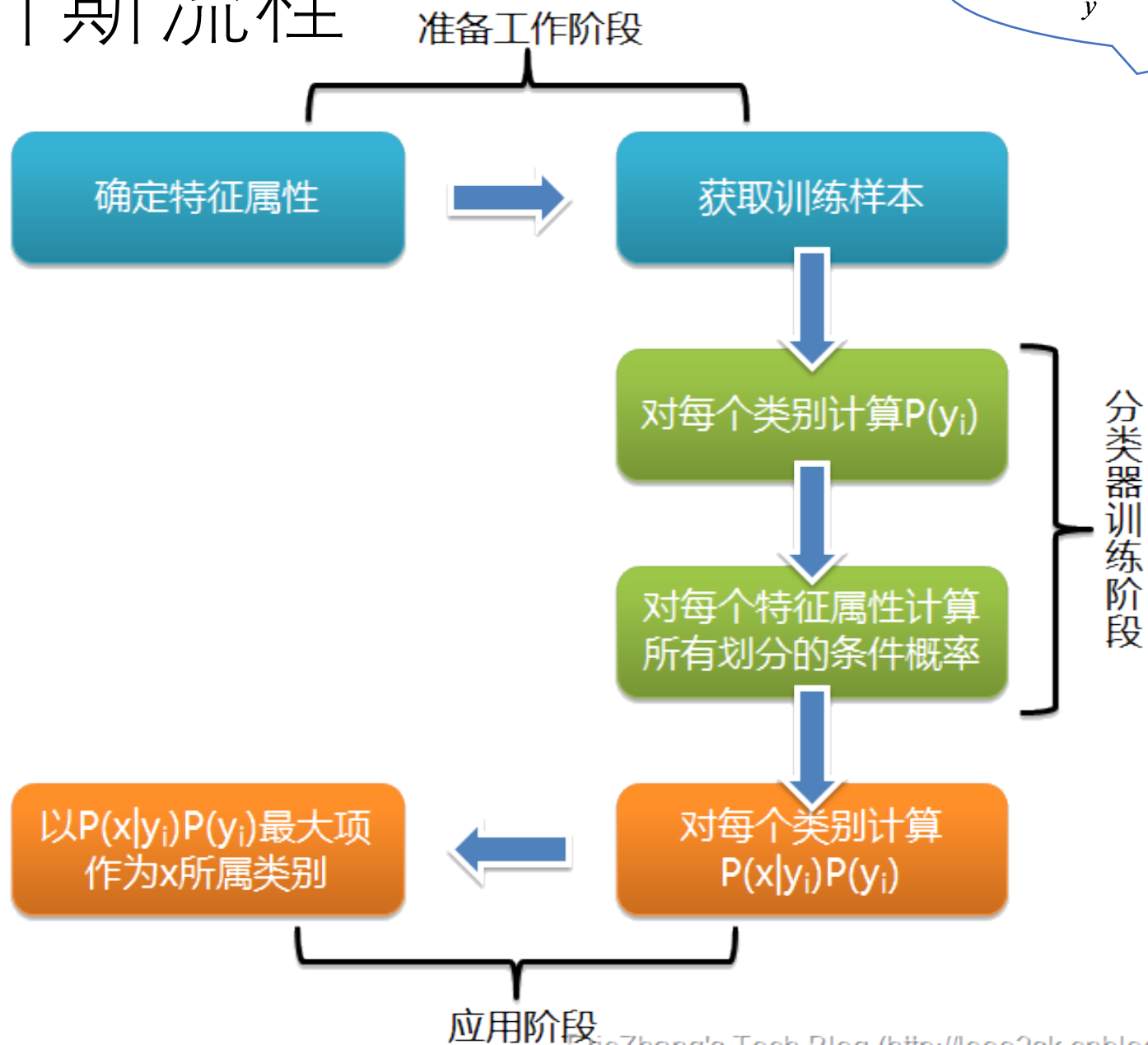


高斯和伯努利朴素贝叶斯 --连续特征

L先生AI课堂

朴素贝叶斯流程

$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^m P(x_i | y)$$



高斯朴素贝叶斯

- Gaussian Naive Bayes是指当特征属性为**连续值**时，而且分布服从**高斯分布**，那么在计算 **$P(\mathbf{x}|\mathbf{y})$** 的时候可以直接使用高斯分布的概率公式：

$$g(x, u, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2}{2\sigma^2}}$$

$$P(x_i | y_k) = g(x_i, u_{i,y_k}, \sigma_{i,y_k})$$

- 因此只需要计算出各个类别中此特征项划分的各个**均值和标准差**

	1	2	3	4	5	6	7	8	9	10
x1	1.1	1.2	1.3	2.1	2.2	2.3	2.4	3.1	3.1	4.1
x2	1	1	1	2	2	2	2	3	3	4
y	-1	1	1	-1	-1	-1	1	1	1	1



$$g(x, u, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2}{2\sigma^2}}$$

$$P(x_i | y_k) = g(x_i, u_{i,y_k}, \sigma_{i,y_k})$$



$$U(x1|y=1) = (1.2 + 1.3 + 2.4 + 3.1 + 3.1 + 4.1) / 6 = \mathbf{u11}$$

$$\sigma(x1|y=1) = ((1.2 - u11)^2 + (1.3 - u11)^2 + (2.4 - u11)^2 + (3.1 - u11)^2 + (3.1 - u11)^2 + (4.1 - u11)^2) / 6$$

伯努利朴素贝叶斯

- Bernoulli Naive Bayes是指当**特征属性为连续值时**，而且分布服从伯努利分布，那么在计算 $P(x|y)$ 的时候可以直接使用伯努利分布的概率公式：

$$P(x_k | y) = P(1 | y)x_k + (1 - P(1 | y))(1 - x_k)$$

- 伯努利分布是一种离散分布，只有两种可能的结果。1表示成功，出现的概率为 p ；0表示失败，出现的概率为 $q=1-p$.

	1	2	3	4	5	6	7	8	9	10
x1	1.1	1.2	1.3	0	0	0	0	0	0	0
x2	1	1	1	0	0	0	0	0	0	0
y	-1	1	1	-1	-1	-1	1	1	1	1



$$P(x_k | y) = P(1 | y)x_k + (1 - P(1 | y))(1 - x_k) \longrightarrow P(x=1.1|y=1)=2/6=1/3$$