


### 3. 计算机-学习优化:

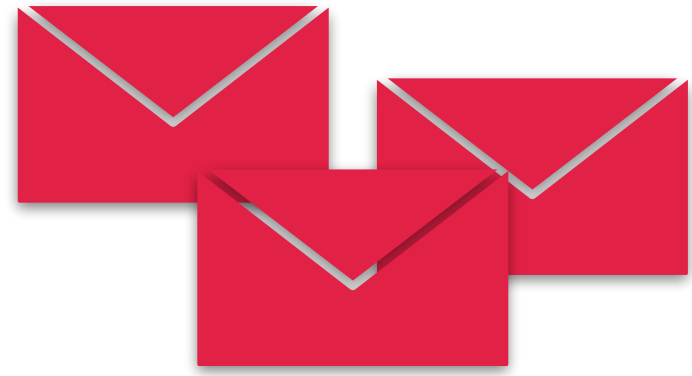
正常短信  
9条



$P(\text{正常}) = 0.75$

$$P(\text{“您好”}|\text{正常}) = \frac{5}{16}$$
$$P(\text{“话费”}|\text{正常}) = \frac{4}{16}$$
$$P(\text{“链接”}|\text{正常}) = \frac{1}{16}$$
$$P(\text{“验证码”}|\text{正常}) = \frac{6}{16}$$

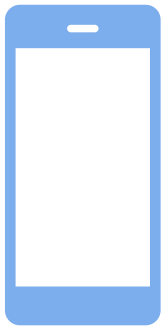
垃圾短信  
3条



$P(\text{垃圾}) = 0.25$

$$P(\text{“您好”}|\text{垃圾}) = \frac{2}{8}$$
$$P(\text{“话费”}|\text{垃圾}) = \frac{1}{8}$$
$$P(\text{“链接”}|\text{垃圾}) = \frac{5}{8}$$
$$P(\text{“验证码”}|\text{垃圾}) = \frac{0}{8}$$

此时，我们手机又收到一条短信：  
“验证码 链接 链接 链接”



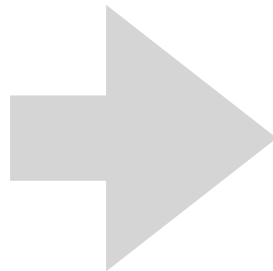
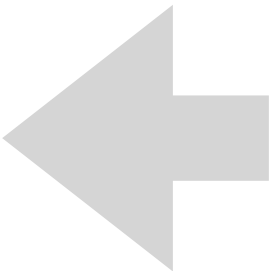
$$P(\text{正常}) \times P(\text{“验证码”}|\text{正常}) \times P(\text{“链接”}|\text{正常})^3 = 0.75 \times \frac{6}{16} \times \left(\frac{1}{16}\right)^3 = 0.00006$$

“验证码 链接 链接 链接”为正常短信的概率

计算机重复之前类似的步骤，得出两类的概率，  
计算机得出这条短信为**正常短信**，因为为垃圾短信的概率是0

$$P(\text{垃圾}) \times P(\text{“验证码”}|\text{垃圾}) \times P(\text{“链接”}|\text{垃圾})^3 = 0.25 \times \frac{0}{8} \times \left(\frac{5}{8}\right)^3 = 0$$

“验证码 链接 链接 链接”为垃圾短信的概率



### 3. 计算机-学习优化:



$P(\text{正常}) = 0.75$

$$\begin{aligned} P(\text{“您好”}|\text{正常}) &= \frac{5}{16} \\ P(\text{“话费”}|\text{正常}) &= \frac{4}{16} \\ P(\text{“链接”}|\text{正常}) &= \frac{1}{16} \\ P(\text{“验证码”}|\text{正常}) &= \frac{6}{16} \end{aligned}$$

此时，我们手机又收到一条短信：  
“验证码 链接 链接 链接”



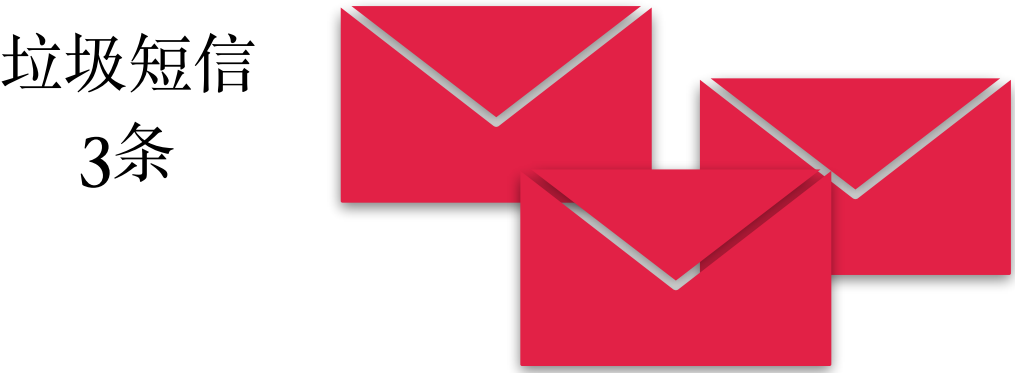
$$P(\text{正常}) \times P(\text{“验证码”}|\text{正常}) \times P(\text{“链接”}|\text{正常})^3 = 0.75 \times \frac{6}{16} \times \left(\frac{1}{16}\right)^3 = 0.00006$$

“验证码 链接 链接 链接”为正常短信的概率

但是，这明显是错误的，

因为在计算机学习时，“验证码”一词没有出现在垃圾短信中，

导致  $P(\text{验证码}|\text{垃圾})$  为0，使最后结果为0。



$P(\text{垃圾}) = 0.25$

$$\begin{aligned} P(\text{“您好”}|\text{垃圾}) &= \frac{2}{8} \\ P(\text{“话费”}|\text{垃圾}) &= \frac{1}{8} \\ P(\text{“链接”}|\text{垃圾}) &= \frac{5}{8} \\ P(\text{“验证码”}|\text{垃圾}) &= \frac{0}{8} \end{aligned}$$

$$P(\text{垃圾}) \times P(\text{“验证码”}|\text{垃圾}) \times P(\text{“链接”}|\text{垃圾})^3 = 0.25 \times \frac{0}{8} \times \left(\frac{5}{8}\right)^3 = 0$$

“验证码 链接 链接 链接”为垃圾短信的概率

### 3. 计算机-学习优化：

正常短信  
9条



$P(\text{正常}) = 0.75$

$$P(\text{“您好”}|\text{正常}) = \frac{5}{16}$$
$$P(\text{“话费”}|\text{正常}) = \frac{4}{16}$$
$$P(\text{“链接”}|\text{正常}) = \frac{1}{16}$$
$$P(\text{“验证码”}|\text{正常}) = \frac{6}{16}$$

垃圾短信  
3条



$P(\text{垃圾}) = 0.25$

$$P(\text{“您好”}|\text{垃圾}) = \frac{2}{8}$$
$$P(\text{“话费”}|\text{垃圾}) = \frac{1}{8}$$
$$P(\text{“链接”}|\text{垃圾}) = \frac{5}{8}$$
$$P(\text{“验证码”}|\text{垃圾}) = \frac{0}{8}$$

此时，我们手机又收到一条短信：  
“验证码 链接 链接 链接”



根据我们的直觉，这应该是一条**垃圾短信**，因为出现了很多“链接”词语，  
让我们看看计算机是如何判断的。

