

3. 计算机-学习优化：

正常短信
9条



$P(\text{正常}) = 0.75$

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

此时，我们手机又收到一条短信：
“验证码 链接 链接 链接”



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

“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。

垃圾短信
3条



$P(\text{垃圾}) = 0.25$

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

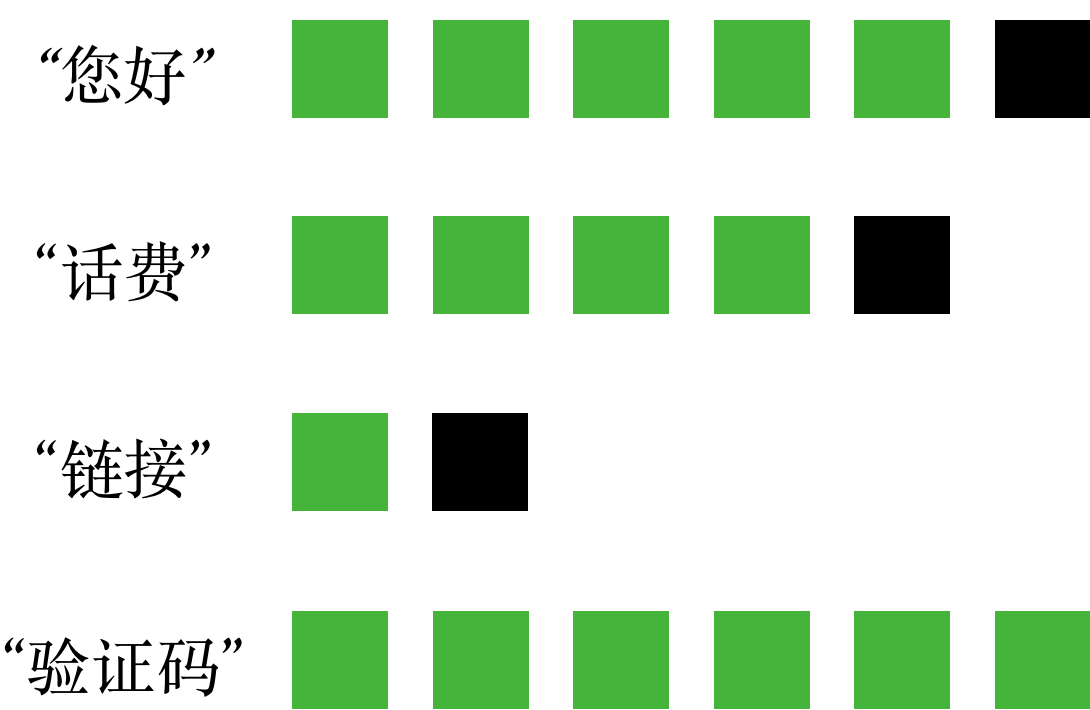
$$P(\text{垃圾}) \times P(\text{“验证码”}|\text{垃圾}) \times P(\text{“链接”}|\text{垃圾})^3 = 0.25 \times \frac{0}{8} \times (\frac{5}{8})^3 = 0$$

“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：



$P(\text{正常}) = 0.75$



$P(\text{“您好”}|\text{正常}) = \frac{5 + 1}{16 + 4} = \frac{6}{20}$

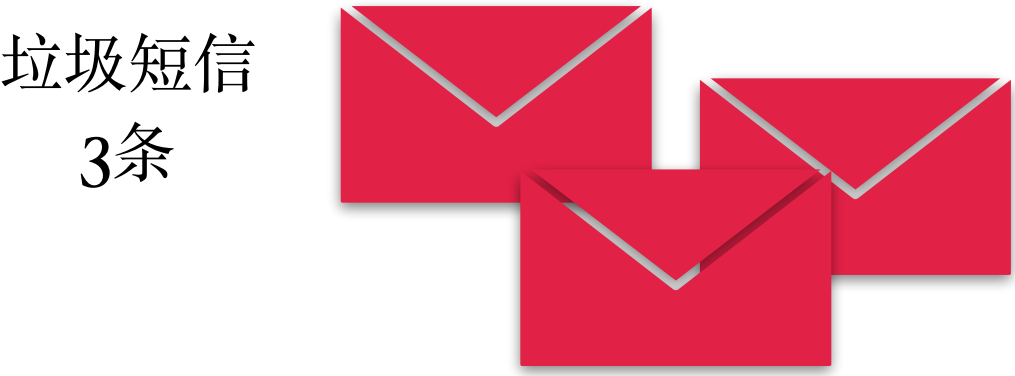
$P(\text{“话费”}|\text{正常}) = \frac{4 + 1}{16 + 4} = \frac{5}{20}$

$P(\text{“链接”}|\text{正常}) = \frac{1 + 1}{16 + 4} = \frac{2}{20}$

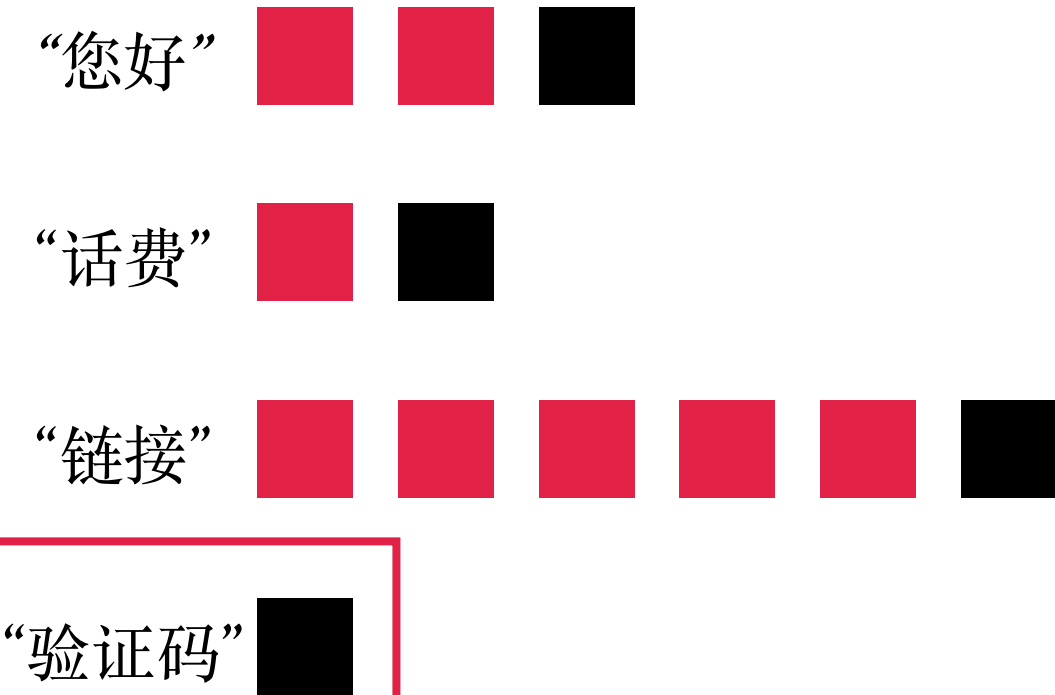
$P(\text{“验证码”}|\text{正常}) = \frac{6 + 1}{16 + 4} = \frac{7}{20}$

解决方法为，让计算机默认每个词都**多出现一次**，

保证不会有0出现，并重新计算学习结果



$P(\text{垃圾}) = 0.25$



$P(\text{“您好”}|\text{垃圾}) = \frac{2 + 1}{8 + 4} = \frac{3}{12}$

$P(\text{“话费”}|\text{垃圾}) = \frac{1 + 1}{8 + 4} = \frac{2}{12}$

$P(\text{“链接”}|\text{垃圾}) = \frac{5 + 1}{8 + 4} = \frac{6}{12}$

$P(\text{“验证码”}|\text{垃圾}) = \frac{0 + 1}{8 + 4} = \frac{1}{12}$