

一个模型出现过拟合的情况，他的参数会趋向于无穷大

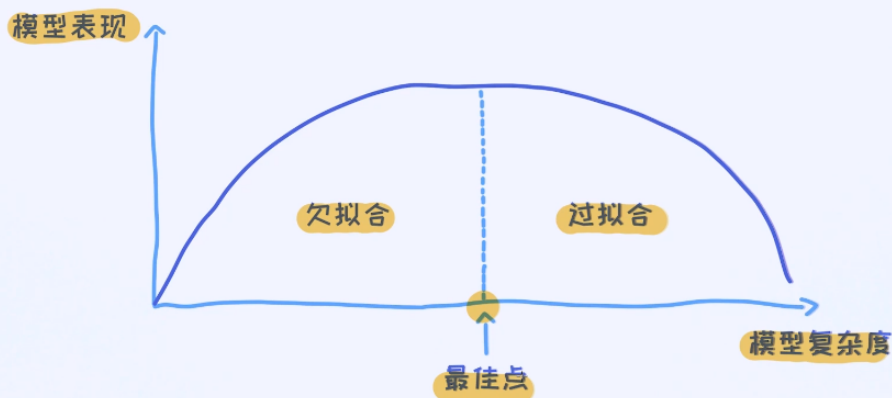
模型的过拟合

通俗来讲，当一个模型在训练数据上表现很不错，但在测试数据上表现比较差的现象叫做模型的过拟合，也就是在数据和测试数据上“判若两人”

构建泛化能力强的模型

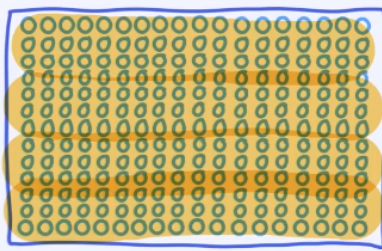
- 1 选择正确的数据
- 2 选择合适的模型
- 3 选择合适的优化算法
- 4 避免模型的过拟合

模型的复杂度

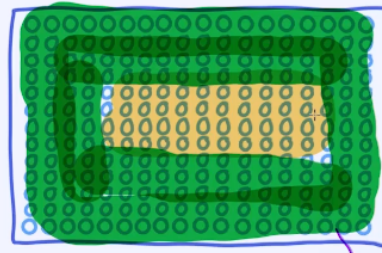


避免过拟合:1数据量增加 2加入正则项 3使用更简单的模型

正则的作用：对可行解空间（feasible region）的限制



加入正则之前



加入正则之后

复杂的模型

总体来讲，通过正则我们可以缩小可行解空间，而且在这里被我们丢弃掉的可行解是比较容易产生过拟合的。我们知道正则的强度是由超参数 λ 来控制，这个值越大，所选择的可行解空间也会相应变小。

正则是一个很大的领域，它既可以用在逻辑回归，也可以用在SVM、线性回归、神经网络、各类深度学习模型里。而且正则是一种能够把先验知识加入到模型里的最直接的方式。

l_1 范数具有稀疏性，比如用在逻辑回归可以用于特征选择

两种常见的正则项



正则

$$J(w) = f(w) + \lambda \|w\|_2^2$$

$$J(w) = f(w) + \lambda \|w\|_1$$

名称

L_2

L_1

数学表达 $\|w\|_2^2 = w_1^2 + w_2^2 + \dots + w_D^2$

$$\|w\|_1 = |w_1| + |w_2| + \dots + |w_D|$$

作用

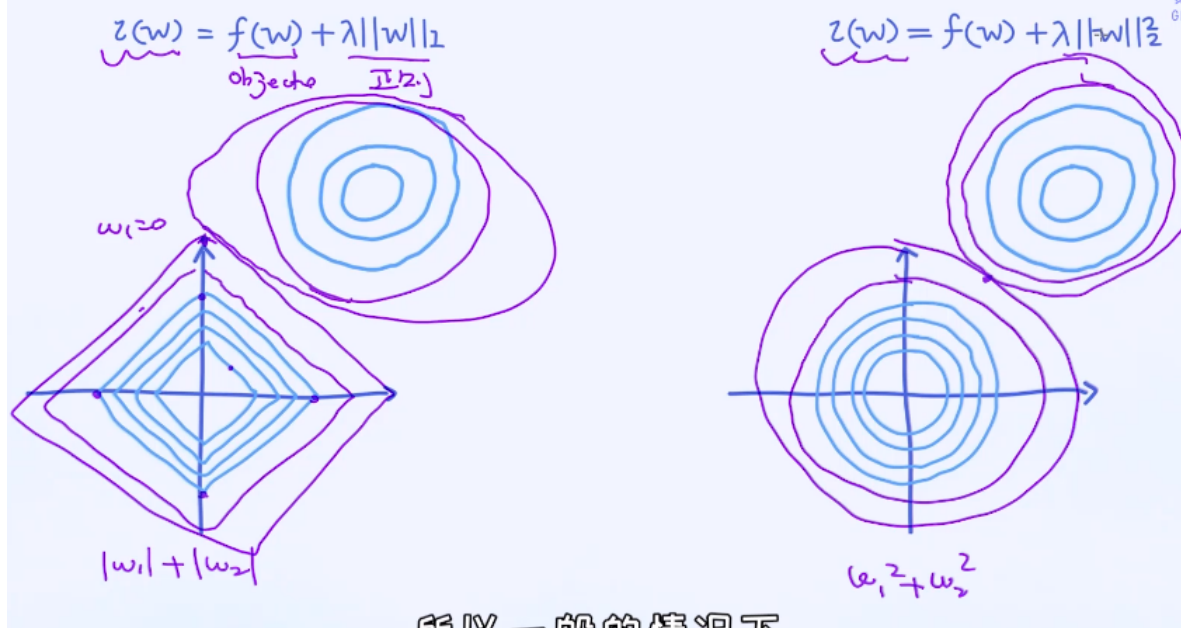
1. 让参数变小

1. 让参数变小

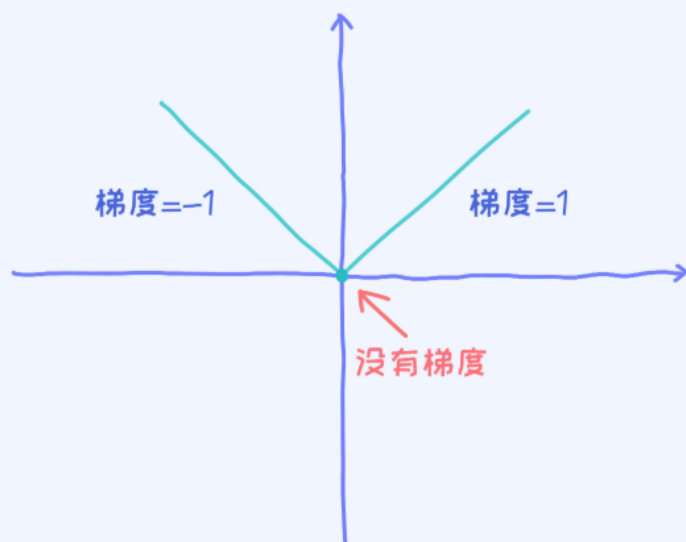
2. 稀疏性：很多参数等于0

l_1 范数更容易与 $f(w)$ 相交在顶点处，顶点处值为0

L1 和 L2 的几何意义



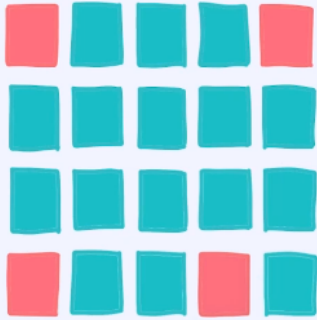
从计算的角度来讲，L1范数的挑战要大很多。一旦目标函数里包含了L1的正则，则优化起来会比较麻烦。主要的原因是L1范数在0点不具备梯度，所以需要做一些特殊处理，比如使用subgradient来代替梯度。



精确率和召回率 (precision, recall)



正常邮件



正常:

$$\text{精} = \frac{16}{18} = \frac{8}{9}$$

$$\text{召} = \frac{16}{20} = \frac{4}{5}$$

垃圾:

$$\text{精} = \frac{3}{7}$$

$$\text{召} = \frac{3}{5}$$

垃圾邮件



F1-Score

$$f1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{(\text{precision} + \text{recall})}$$