【一看就懂】机器学习之L1和L2正则化

尘缘墨语
百家号 03-23 15:18

摘要：本文主要分为三部分，先讲述什么是正则化，再讲L1和L2正则化数学原理，最后小结对比。

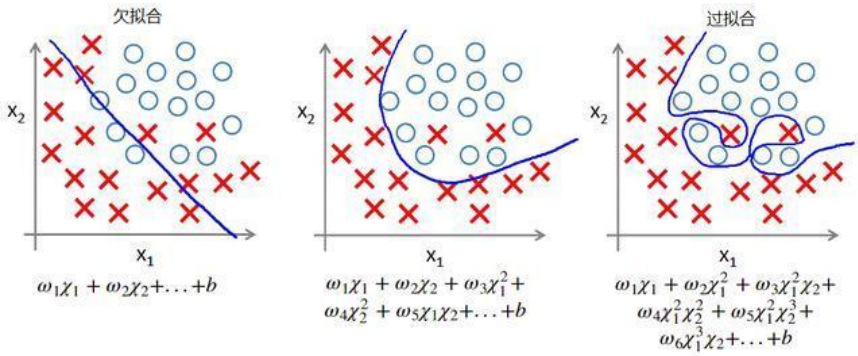
一、什么是正则化

上篇机器学习之线性方程与学习率中引入损失函数，以便寻找最优权重W。

然而正如大名鼎鼎的奥卡姆剃刀定律，

> *奥卡姆剃刀定律：“如无必要，勿增实体”，即“简单有效原理”*
> *“切勿浪费较多东西去做，用较少的东西，同样可以做好的事情。”*

模型越复杂，越容易过拟合。



因此，原先以最小化损失（经验风险最小化）为目标：

$$\text{minimize}(\text{Loss}(\text{Data}|\text{Model}))$$

现在以最小化损失和模型复杂度（结构风险最小化）为目标：

$$\text{minimize}(\text{Loss}(\text{Data}|\text{Model}) + \text{complexity}(\text{Model}))$$

通过降低复杂模型的复杂度来防止过拟合的规则称为正则化。

二、L1 和 L2 正则化的数学原理

机器学习中最常见的即L1和L2正则化。

1. L1正则化，即原损失函数 + 所有权重的平均绝对值 * λ ，其中λ >0

根据损失更新ω，需要先对ω求导：

那么权重ω的更新规则为：

尘缘墨语
百家号 最近更新 03-23 15:18

简介:用笔尖的墨滴，写下尘缘的轨迹

作者最新文章

干货|教你如何判断一台3D打印机的精度

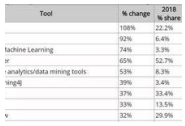
手机拍照技巧经验分享二

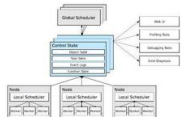
大庆人要当心！微信号随意买卖既违规又危险

相关文章


从苹果店员到机器学习工程师：学习AI，我…
机器之心 10-08


非算法类人工智能从业者须知的十件事
嘟嘟味 10-09


最受欢迎机器学习框架王座之争：PyTorch存…
DeepTech深科技 10-08


实时机器学习框架的设计原则
大众影迷 10-11


医疗保健如何为人工智能，机器学习做准备
只留下了印记 10-10

η即学习率

比原始的更新规则多出了η * λ * sgn(ω)/n。

- 当ω为正时，更新后的ω变小
- 当ω为负时，更新后的ω变大

可见每次更新，ω都是往0靠，即使网络中的权重尽可能为0。

2. L2正则化，即原损失函数 + 所有权重平方和的平均值 * λ / 2，λ>0

$$C = C_0 + \frac{\lambda}{2n}\sum_w w^2$$

为什么是λ / 2？纯粹是为了后面的数学计算方便。

因为λ是需要设置的正数，因此 λ 和 λ / 2并无区别。

同样，需要先对ω求导：

$$\frac{\partial C}{\partial w} = \frac{\partial C_0}{\partial w} + \frac{\lambda}{n} w$$

那么权重ω的更新规则为：

$$\begin{aligned} w &\rightarrow w - \eta\frac{\partial C_0}{\partial w} - \frac{\eta\lambda}{n} w \\ &= \left(1 - \frac{\eta\lambda}{n}\right) w - \eta\frac{\partial C_0}{\partial w}. \end{aligned}$$

原始的更新规则ω系数为1，现在为 1ηλ/n 。

因为η、λ、n都>0，所以 1ηλ/n

三、L1和L2正则化小结

L2 和 L1 采用不同的方式降低权重：
- L1 会降低 |权重|。
- L2 会降低权重2。

因此，L2 和 L1 具有不同的导数：
- L1 的导数为 常数，其值与权重无关）。
- L2 的导数为 2 * 权重。

L1 正则化可以理解为每次从权重中减去一个常数。

L2 正则化可以理解为每次移除权重的 x%。

本质都是为了降低模型的复杂度，防止过拟合。