线性回归

机器学习的四个关键部分

• 数据(经验): 你有什么样的数据;

• 模型(假设): 对给定的问题有什么样的假设;

损失函数(目标):如何评估一个模型;优化函数(改善):如何获取最优的模型。

线性回归

模型

$$y = f_{ heta}(x) = heta_0 + \sum_{j=1}^d heta_j x_j = heta^ op x, x = (1, x_1, x_2, \ldots, x_d)$$

这个 x_i 可以是非线性的,对于 $y=a\cos(x)+b\sin(x)+c$,可以令 $x_1=\cos(x),x_2=\sin(x)$ 得到线性方程。

目标/损失函数

目标:

$$\min_{ heta} rac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(y_{i}, f_{ heta}\left(x_{i}
ight)
ight)$$

最常使用均方误差 (MSE):

$$J_{ heta} = rac{1}{2N} \sum_{i=1}^{N} \left(y_{i} - f_{ heta}\left(x_{i}
ight)
ight)^{2} \quad \min_{ heta} J_{ heta}$$

优化

使用最常用的梯度下降: $\theta_{\text{new}} \leftarrow \theta_{\text{old}} - \eta \frac{\partial \mathcal{L}(\theta)}{\partial \theta}$

梯度下降可以细分为以下三种:

方法	损失函数	参数更新方程	优缺点
BGD	$J(heta) = rac{1}{2N} \sum_{i=1}^{N} \left(y_i - f_{ heta}\left(x_i ight) ight)^2 \min_{ heta} J(heta)$	$ heta_{ ext{new}} = heta_{ ext{old}} + \eta rac{1}{N} \sum_{i=1}^{N} \left(y_i - f_{ heta} \left(x_i ight) ight) \! x_i$	更稳定,但是收 敛速度慢,容易 获得局部最优 解。
SGD	$J^{(i)}(heta) = rac{1}{2}(y_i - f_ heta\left(x_i ight))^2 \min_ heta rac{1}{N} \sum_i J^{(i)}(heta)$	$ heta_{ ext{new}} = heta_{ ext{old}} + \eta \left(y_i - f_{ heta} \left(x_i ight) ight) \! x_i$	收敛速度快,但 是可能会导致波 动和不确定性。
MBGD	$J^{(k)}(heta) = rac{1}{2N_k} \sum_{i=1}^{N_k} \left(y_i - f_{ heta}\left(x_i ight) ight)^2$	$ heta_{ m new} = heta_{ m old} - \eta rac{\partial J^{(k)}(heta)}{\partial heta}$	结合了SGD和 BGD的优点,并 且适合并行处 理。

学习率

学习率是一个需要用户选择的超参数。如果学习率太小,收敛速度太慢;如果太大,可能会波动甚至发散。

矩阵形式

目标: $J(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^{\top}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) \quad \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$

梯度下降: $\frac{\partial J(oldsymbol{ heta})}{\partial oldsymbol{ heta}} = -oldsymbol{X}^ op(oldsymbol{y} - oldsymbol{X}oldsymbol{ heta})$

最优解:

$$egin{aligned} rac{\partial J(oldsymbol{ heta})}{\partial oldsymbol{ heta}} &= oldsymbol{0} \Rightarrow oldsymbol{X}^ op (oldsymbol{y} - oldsymbol{X} oldsymbol{ heta}) = oldsymbol{0} \ &\Rightarrow oldsymbol{X}^ op oldsymbol{y} &= oldsymbol{X}^ op oldsymbol{X}^ op$$

正则化

有可能 $m{X}^ op m{X}$ 是不可逆的,此时可以引入正则化: $J(m{ heta}) = rac{1}{2}(m{y} - m{X}m{ heta})^ op (m{y} - m{X}m{ heta}) + rac{\lambda}{2}\|m{ heta}\|_2^2$

梯度下降: $\frac{\partial J(m{ heta})}{\partial m{ heta}} = -m{X}^{ op}(m{y} - m{X}m{ heta}) + \lambdam{ heta}$

最优解: