

一、线性神经网络

1.1 线性回归:

模型:

input: $X = [x_1, x_2, x_3, \dots, x_n]^T$

权重, 标量偏差: $w = [w_1, w_2, w_3, \dots, w_n]^T, b$

output: $y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$

向量: $y = \langle w, x \rangle + b$

平方损失 $l(\hat{y}, y) = \frac{1}{2} (y - \hat{y})^2$

参数学习:

训练损失: $L = \frac{1}{2n} \|y - Xw - b\|^2$

最小化损失学习参数: $w^*, b^* = \arg \min (L(X, y, w, b))$

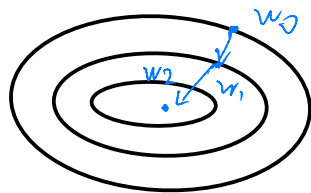
1.2 基础优化方法:

梯度下降:

① 初始化 w_0

② 重复迭代, $t=1, 2, 3, \dots, n$

$$w_t = w_{t-1} - \eta \frac{\partial L}{\partial w_{t-1}}$$



学习率: 一次走的步长

太大: 在最优解处震荡

太小: 迭代次数最多

小批量随机梯度下降:

随机采样 b 个样本 $i_1, i_2, i_3, \dots, i_b$ 近似损失

$$\frac{1}{b} \sum_{i \in I_b} \ell(x_i, y_i, w)$$

b : 批量大小