Linear Regression

实战目标：

1. 掌握线性回归模型的基本原理
2. 实现模型训练、预测、评价、存储和加载
3. 实现利用向量运算的模型训练
4. 可视化

一元线性回归

设数据集有 m 个样本 $X=\{x^{[1]},x^{[2]},...,x^{[m]}\},Y=\{y^{[1]},y^{[2]},...,y^{[m]}\},x\in\mathbb{R},y\in\mathbb{R}$

模型函数：

$$\hat{y}=a+bx$$

均方差损失函数：

$$\begin{aligned}J(a,b)&=\frac{1}{2m}\Sigma_{i=1}^{m}(\hat{y}^{[i]}-y^{[i]})^2\\&=\frac{1}{2m}\Sigma_{i=1}^{m}(a+bx^{[i]}-y^{[i]})^2\end{aligned}$$

损失函数在a，b方向上的偏导数：

$$\begin{aligned}a:\frac{dJ(a,b)}{da}&=\frac{1}{m}\Sigma_{i=1}^{m}(a+bx^{[i]}-y^{[i]})\\b:\frac{dJ(a,b)}{db}&=\frac{1}{m}\Sigma_{i=1}^{m}(a+bx^{[i]}-y^{[i]})*x^{[i]}\end{aligned}$$

最小二乘法求解：

J(a,b) 是一个凸函数，当 J(a,b) 的偏导数为0时达到极值

$$\begin{cases} \frac{dJ(a,b)}{da} = 0 \\ \frac{dJ(a,b)}{db} = 0 \end{cases}$$

得出:

$$\begin{cases} a = \frac{1}{m}\Sigma_{i=1}^{m} y^{[i]} - \frac{b}{m}\Sigma_{i=1}^{m} x^{[i]} \\ b = \frac{\Sigma_{i=1}^{m} x^{[i]} y^{[i]} - \frac{1}{m}\Sigma_{i=1}^{m} x^{[i]} \Sigma_{i=1}^{m} y^{[i]}}{\Sigma_{i=1}^{m} {x^{[i]}}^2 - \frac{1}{m}\Sigma_{i=1}^{m} x^{[i]} \Sigma_{i=1}^{m} x^{[i]}} \end{cases}$$

梯度下降法求解:

参数延着逆梯度方向下降，逐步逼近损失函数的极小值

设学习率为 η:

$$a = a - \eta * \frac{dJ(a,b)}{da}$$
$$b = b - \eta * \frac{dJ(a,b)}{db}$$

多元线性回归

设数据集有 m 个样本 $X = \{X^{[1]}, X^{[2]}, ..., X^{[m]}\}, Y = \{y^{[1]}, y^{[2]}, ..., y^{[m]}\}, X^{[i]} \in \mathbb{R}^n, y^{[i]} \in \mathbb{R}$

模型函数:

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_n x_n$$

为了方便计算，对 X 再添加一维，即令 $x_0^{[i]} = 1, X^{[i]} \in \mathbb{R}^{n+1}$, 令$\Theta = [\theta_0, \theta_1, ..., \theta_n]^T, \Theta \in \mathbb{R}^{n+1}$, 则:

$$\begin{aligned} \hat{y}^{[i]} &= \theta_0 x_0^{[i]} + \theta_1 x_1^{[i]} + \theta_2 x_2^{[i]} + ... + \theta_n x_n^{[i]} \\ &= \Theta^T X^{[i]} \end{aligned}$$

均方差损失函数:

$$\begin{aligned} J(\Theta) &= \frac{1}{2m}\Sigma_{i=1}^{m}(\hat{y}^{[i]} - y^{[i]})^2 \\ &= \frac{1}{2m}\Sigma_{i=1}^{m}(\Theta^T X^{[i]} - y^{[i]})^2 \end{aligned}$$

损失函数在 θ_k 上的偏导数:

$$\frac{dJ(\Theta)}{d\theta_k} = \frac{1}{m}\Sigma_{i=1}^{m}(\Theta^T X^{[i]} - y^{[i]})X_k^{[i]}$$

梯度下降法求解:

参数延着逆梯度方向下降，逐步逼近损失函数的极小值

设学习率为 η:

$$\theta_k = \theta_k - \eta * \frac{dJ(\Theta)}{d\theta_k}$$

用向量表示

损失函数在 Θ 上的偏导数:

$$\frac{dJ(\Theta)}{d\Theta} = X^{[i]}\frac{1}{m}\Sigma_{i=1}^{m}(\Theta^T X^{[i]} - y^{[i]})$$

梯度下降法求解:

依旧设学习率为 η，此时的梯度下降变化为

$$\Theta = \Theta - \eta * \frac{dJ(\Theta)}{d\Theta}$$