Part III-B: Medicine AI

Lecture by None
Note by THF

2025 年 3 月 8 日

目录

Learn 9

03.06

0.1 回归

简单来说，回归通过一系列**离散**的点，估计出一个**连续**的关系。回归的种类：

- 线性回归：使用线性函数 $f(x_i) = wx_i + b$
- 对数几率回归：本质是一种分类算法，通过引入 sigmoid 函数来确定哪些数据被激活
- 多项式回归：通过 $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta x^n + \varepsilon$ 来拟合数据
- 岭回归
- 套索回归
- ...

sigmoid 函数：

$$\sigma(z) = \frac{1}{1+e^{-z}}.$$

图像如下：

这个函数对应的回归模型为：

$$P(y=1 \mid \boldsymbol{x}) = \sigma\left(\boldsymbol{\theta}^\top \boldsymbol{x}\right) = \frac{1}{1+e^{-\boldsymbol{\theta}^\top \boldsymbol{x}}}$$
$$P(y=0 \mid \boldsymbol{x}) = 1 - P(y=1 \mid \boldsymbol{x}).$$

首先来看线性回归，有一个数据集合如下：

$$D = \{(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \ldots (\boldsymbol{x_n}, y_n)\}.$$

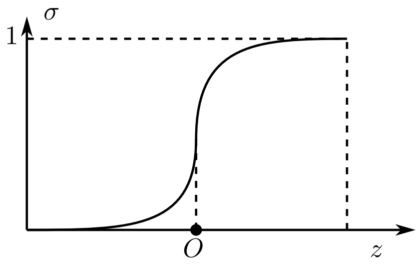

图 1: sigmoid 函数

其中集合中元素的第一个参数为一个 m 维向量，即：

$$\boldsymbol{x}_i = \begin{bmatrix} x_i^{(1)} & x_i^{(2)} & x_i^{(3)} \ldots x_i^{(m)} \end{bmatrix}^\top .$$

首先考虑一维向量，即常数，这个计算就是常见的线性回归计算，使用最小二乘法，求出 w, b 并拟合为以下表达式：

$$f(x_i) = wx_i + b.$$

如果输入为多维向量，如三维向量，那么由于向量需要通过点乘得到常数，因此 w 成为一个向量，即：

$$f(\boldsymbol{x}_i) = \boldsymbol{w}^\top \boldsymbol{x}_i + b + \varepsilon_i.$$

这里有一个误差项，表示回归所不能解释的部分（常常看作一个 $\mathbb{E}_{\varepsilon_i \sim p} = 0$ 的高斯噪声），一般不考虑，因此也写为：

$$f(\boldsymbol{x}_i) \approx \boldsymbol{w}^\top \boldsymbol{x}_i + b.$$

归类到多维后，称 $\boldsymbol{w} \in \mathbb{R}^m$ 为模型的**权重向量**。回归中常用 MSE 来评价性能，求解 $\boldsymbol{w}$ 和 b 一般使用**最小二乘法**，过程如下：

证明. 回顾残差平方和：

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 .$$

把 $\hat{y_i}$ 换为 $f(\boldsymbol{x}_i)$ 写出这个回归的均方误差表达：

$$E_{(\boldsymbol{w},b)} = \underset{(\boldsymbol{w},b)}{\arg\min} \sum_{i=1}^{m} (f(\boldsymbol{x}_i) - y_i)^2 = \underset{(\boldsymbol{w},b)}{\arg\min} \sum_{i=1}^{m} \left(\boldsymbol{w}^\top \boldsymbol{x}_i + b - y_i\right)^2 \tag{1}$$

这里 $\arg\min$ 的意思是通过取一组 $\boldsymbol{w}, b$ 的值使得表达式最小，严格定义上，这个式子所代表的过程就称为线性回归。求一个二元函数 $E_{(\boldsymbol{w},b)} = \sum_{i=1}^{m} \left(\boldsymbol{w}^\top \boldsymbol{x}_i + b - y_i\right)^2$ 的最小值，由于向量内积的性质 $\boldsymbol{w}^\top \boldsymbol{x}_i = \boldsymbol{x}_i^\top \boldsymbol{w}$ 也可以写为：

$$E = \sum_{i=1}^{m} \left(\boldsymbol{x}_i^\top \boldsymbol{w} + b - y_i\right)^2 .$$

这个式子可以直接求偏导来计算最小值，但是为了机器运算方便，尽量转化为矩阵运算。平方和很容易让人想到 L^2 范数，或者 Frobenius 范数。观察可得，这里的平方和可以拆成 y 和 $\boldsymbol{x}_i^\top \boldsymbol{w}+b$ 两部分。记住 $\boldsymbol{x}_i$ 和 $\boldsymbol{w}$ 都是列向量，这里为了统一输入，即输入的矩阵不含变量，给每一个 $\boldsymbol{x}_i$ 的末端加上一个 1，给 $\boldsymbol{w}$ 的末端加上一个 b ，这样做之后两个向量变成下列形式：

$$
\begin{aligned}
\widetilde{\boldsymbol{x}}_i &= \begin{bmatrix} \boldsymbol{x}_i^\top & 1 \end{bmatrix}^\top = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(n)} \\ 1 \end{bmatrix} \\
\widetilde{\boldsymbol{w}} &= \begin{bmatrix} \boldsymbol{w}^\top & b \end{bmatrix}^\top = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ b \end{bmatrix}.
\end{aligned}
$$

将所有的 $\widetilde{\boldsymbol{x}}_i$ 横向合并，记为 $\boldsymbol{X}$ ，形如：

$$
\boldsymbol{X} = \begin{bmatrix} \widetilde{\boldsymbol{x_1}}^\top \\ \widetilde{\boldsymbol{x_2}}^\top \\ \vdots \\ \widetilde{\boldsymbol{x_m}}^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1}^\top & 1 \\ \boldsymbol{x_2}^\top & 1 \\ \vdots & \vdots \\ \boldsymbol{x_m}^\top & 1 \end{bmatrix} \quad \boldsymbol{X} \in \mathbb{R}^{m\times(n+1)}.
$$

带回 $E_{(\boldsymbol{w},b)}$ 后可以把两个参数整合为一个，同时由于所有的 $\widetilde{\boldsymbol{x}}_i$ 在 $\boldsymbol{X}$ 中，计算发现：

$$
E_{\widetilde{\boldsymbol{w}}} = (\boldsymbol{X}\widetilde{\boldsymbol{w}} - \boldsymbol{y})^2 = \|\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{w}}\|_2^2.
$$

想要求一个函数的最小值还可以用梯度法，即 $\nabla f = 0$ 时 f 可能取得一个极值（在这里就是求导），对于 $E_{\widetilde{\boldsymbol{w}}}$，对 $\widetilde{\boldsymbol{w}}$ 求梯度：

$$
\nabla_{\widetilde{\boldsymbol{w}}} E_{(\widetilde{\boldsymbol{w}})} = \frac{\partial E_{\widetilde{\boldsymbol{w}}}}{\partial \widetilde{\boldsymbol{w}}}.
$$

这里有一些小细节，首先将 E 展开为内积形式：

$$
\begin{aligned}
E &= (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{w}})^\top (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{w}}) = \left(\boldsymbol{y}^\top - \widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top\right)(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{w}}) \\
&= \boldsymbol{y}^\top \boldsymbol{y} - 2\widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top \boldsymbol{y} + \widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top \boldsymbol{X}\widetilde{\boldsymbol{w}}.
\end{aligned}
$$

此处注意：$2\widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top \boldsymbol{y}$ 是一个**标量**，因此对其求转置仍为其本身，即：

$$
\begin{aligned}
E &= \boldsymbol{y}^\top \boldsymbol{y} - 2\widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top \boldsymbol{y} + \widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top \boldsymbol{X}\widetilde{\boldsymbol{w}} \\
&= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{y}^\top \widetilde{\boldsymbol{w}}\boldsymbol{X} + \widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top \boldsymbol{X}\widetilde{\boldsymbol{w}}.
\end{aligned}
$$

引入两个对矩阵求导的公式：

$$
\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x}^\top \boldsymbol{A} = \boldsymbol{A}^\top \tag{2}
$$

$$\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = 2\boldsymbol{A}\boldsymbol{x}(\text{对称矩阵}) \tag{3}$$

对于第一个公式，令 $\boldsymbol{A} = \boldsymbol{X}^\top \boldsymbol{y}$ ，可以解得

$$\frac{\partial}{\partial \widetilde{\boldsymbol{w}}} 2\widetilde{\boldsymbol{w}}^\top \boldsymbol{X}^\top \boldsymbol{y} = .$$

对于 □

Learn 10

03.08