# 第一章　回归

回归 (regression) 是一种研究变量之间关系的机器学习方法. 回归问题建立了可观测的自变量 (observation) 到目标变量 (target variable) 之间的一种映射，从而可以根据新的观测值去预测目标变量. 回归既可以是有监督的，也可以是无监督的. 这一部分将主要讨论有监督的线性回归 (linear regression).

## 1.1　线性回归

设样本矩阵 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{m\times n}$，其中 $\mathbf{x} \in \mathbb{R}^m$，样本对应的标签集为 $\mathcal{T} = \mathbb{R}$，即连续实数集，标签向量为 $\mathbf{y} = [y_1, y_2, \cdots, y_n]^T \in \mathbb{R}^n$. 现在考虑建立观测变量 $\mathbf{x}_i$ 和目标变量 $t$ 之间的关系，我们需要构造一个函数 $f(\cdot) : \mathbb{R}^m \mapsto \mathbb{R}$，使得对于每个样本，其输出总是接近于它的标签，即 $f(\mathbf{x}_i) \approx y_i$. 这里我们考虑下面的线性模型：

$$\bar{f}(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b, \tag{1.1.1}$$

其中，$\mathbf{w} \in \mathbb{R}^m$ 是权重向量 (weight vector)， $b \in \mathbb{R}$ 是一个偏移项 (bias). 实际中，由于我们可以把 $\bar{f}(\mathbf{x})$ 写成：

$$\bar{f}(\mathbf{x}) = \hat{\mathbf{w}}^T\hat{\mathbf{x}}, \quad \hat{\mathbf{w}} = \begin{bmatrix}\mathbf{w}\\ b\end{bmatrix}, \quad \hat{\mathbf{x}} = \begin{bmatrix}\mathbf{x}\\ 1\end{bmatrix}. \tag{1.1.2}$$

于是偏移项很自然地归入了权重向量中. 所以，我们在这里更多讨论的线性模型是不带偏移的，即

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}. \tag{1.1.3}$$

我们现在建立 $f(\mathbf{x})$ 和 $y$ 之间的关系：

$$y = f(\mathbf{x}) + \epsilon. \tag{1.1.4}$$

其中 $\epsilon$ 是用 $f(\mathbf{x})$ 近似 $y$ 时产生的误差. 于是，为了使得每一个样本的**平均平方误差 (mean square error, MAE)** 最小，我们有如下的优化问题：

$$\min_f \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2 \tag{1.1.5}$$

将线性模型定义 (1.1.3) 代入，可以得到等价的优化问题：

$$\min_{\mathbf{w}} \sum_{i=1}^n (y_i - \mathbf{w}^T\mathbf{x}_i)^2 \tag{1.1.6}$$

$$= \min_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}^T\mathbf{w}\|_2^2 = \min_{\mathbf{w}} \mathbf{y}^T\mathbf{y} - 2\mathbf{w}^T\mathbf{X}\mathbf{y} + \mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w}. \tag{1.1.7}$$

这是一个无约束的凸优化问题，它的最优点在目标函数对 $\mathbf{w}$ 的导数为0时取到，于是有

$$-2\mathbf{X}\mathbf{y} + 2\mathbf{X}\mathbf{X}^T\mathbf{w} = 0 \Rightarrow \mathbf{w} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{y}. \tag{1.1.8}$$

这个求解 $\mathbf{w}$ 的方法也被称为**最小二乘法 (least square method)** 在获得 $\mathbf{w}$ 之后，我们就获得了一个最优的 $f(\cdot)$，根据它可以对未知点 $\mathbf{x}^*$ 的标签进行预测.

线性回归可以扩展到更一般的情况. 假设标签 $\mathbf{y} \in \mathbb{R}^k$ 是一个 $k$ 维的向量. 此时，我们的线性模型则变成了

$$\mathbf{f}(\mathbf{x}) = \mathbf{W}^T\mathbf{x}, \tag{1.1.9}$$

其中，$\mathbf{W} \in \mathbb{R}^{m\times k}$. 因此，优化问题扩展为

$$\min_{\mathbf{W}} \sum_{i=1}^{n} \|\mathbf{y}_i - \mathbf{W}^T\mathbf{x}_i\|_2^2 \tag{1.1.10}$$

$$= \min_{\mathbf{W}} \|\mathbf{Y} - \mathbf{X}^T\mathbf{W}\|_F^2, \tag{1.1.11}$$

其中，$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]^T \in \mathbb{R}^{k\times n}$. 同样的，通过最小二乘法可以得到一个 $\mathbf{W}$ 的闭式解

$$\mathbf{W} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Y}. \tag{1.1.12}$$

高维情况的回归，即 $k > 1$ 时的情况也是十分常见的. 例如，对于一个多分类问题，它有 $k$ 个类，那么我们就可以利用**独热编码**进行标签表示，我们就能通过回归中的方法解决分类问题. 下一节中所介绍的交叉熵回归就是一种十分常用于解决分类问题的回归方法.

## 1.2 交叉熵回归

### 1.2.1 Logistic 回归

### 1.2.2 Softmax 回归

## 1.3 正则化回归模型

### 1.3.1 岭回归

### 1.3.2 LASSO 回归

## 1.4 概率视角的回归模型