FDU 统计机器学习 2. 线性回归

本文参考以下教材:

- An Introduction To Statistical Learning (2nd Edition, G. James, D. Witten, T. Hastie) Chapter 3
- 统计学习导论: 基于 R 应用 (第二版 G. James, D. Witten, T. Hastie) 第 3 章

欢迎批评指正!

2.1 多元线性回归

考虑输入向量 $x=[x_1,\dots,x_d]^{\mathrm{T}}$ (其中 x_i 是 x 在第 i 个特征上的取值)
线性模型试图学习一个通过特征的线性组合来进行预测的函数:

$$\begin{aligned} f(x) &= w_1x_1+\cdots+w_dx_d+b \\ &= w^{\mathrm{T}}x+b \qquad\qquad (\text{where } \tilde{w}=[b,w_1,\dots,w_d]^{\mathrm{T}}, \tilde{x}=[1,x_1,\dots,x_d]^{\mathrm{T}}) \\ &= \tilde{w}^{\mathrm{T}}\tilde{x} \end{aligned}$$

由于权重向量 w 直观地表达了各特征在预测中的重要程度，故线性模型有很好的可解释性 (comprehensibility).

给定数据集 $D:=\{(x_i,y_i):i=1,\dots,n\}$，其中 $x_i=[x_{i1},\dots,x_{id}]^{\mathrm{T}}$ 而 $y_i\in\mathbb{R}$
均方误差是回归任务中最常用的性能度量:

$$\begin{aligned} (w^\star,b^\star) &= \arg\min_{(w,b)}\sum_{i=1}^m(f(x_i)-y_i)^2 \\ &= \arg\min_{(w,b)}\sum_{i=1}^m(y_i-w^{\mathrm{T}}x_i-b)^2 \ (\text{where } \begin{cases} y=[y_1,\dots,y_n]^{\mathrm{T}} \\ \tilde{x}_i=[1,x_{i1},\dots,x_{id}]^{\mathrm{T}} \\ X=[\tilde{x}_1,\dots,\tilde{x}_n]^{\mathrm{T}}\in\mathbb{R}^{n\times(d+1)} \\ \tilde{w}=[w_1,\dots,w_d,b]^{\mathrm{T}} \end{cases}) \\ &= \arg\min_{\tilde{w}}\|y-X\tilde{w}\|_2^2 \end{aligned}$$

通过令 $\nabla_{\tilde{w}}\|y-X\tilde{w}\|_2^2=-2X^{\mathrm{T}}(y-X\tilde{w})=0_{d+1}$ 可解得 $\tilde{w}^\star=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$

然而在实际应用中 $X^{\mathrm{T}}X$ 往往不是非奇异矩阵.
此时可以解出多个 $\tilde{w}$，它们都能使均方误差最小化.
选择哪个解作为输出将由学习算法决定，常见的做法是引入正则化项控制 $\tilde{w}$ 的大小.
(更多内容请参考 FDU 回归分析 2. 多元线性回归 & 3. 岭回归)

广义线性模型

- ① 其刻画的因变量 Y 来自指数分布族:

  $$f(y|\theta,\phi)=\exp\left\{\frac{yb(\theta)-c(\theta)}{a(\phi)}+d(y,\phi)\right\}$$

  其中 θ 称为中心参数，与 Y 的均值有关，ϕ 称为尺度参数，与 Y 的方差有关.

- ② 系统成分 $y=x^{\mathrm{T}}\beta$

- ③ 链接函数 $g(\cdot)$ (双射，连续可导)
  线性回归 $g(\mu)=\mu$
  逻辑回归 $g(\mu)=\log\left(\frac{\mu}{1-\mu}\right)$
  Possion 回归 $g(\mu)=\log(\mu)$

The End