deepshare.net
堔享网

统计学习方法：

那些你应该知道的基础知识

讲师：K

基础知识

快速回顾

- 线性代数
- 微积分

→ B站3blue1brown微积分和现代视频

- 最优化理论 → Stephen Boyd的凸优化理论（有中文译本）
- 概率论 → 概率论与数理统计

注意：

我们这门课的重点并不是学习上述的这些内容，但是我们需要一些基础的知识。

使用方法建议：

需要了解的知识请及时翻看上面的内容~

线性代数

向量和矩阵

- 向量和矩阵的表达形式：

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ a_{11} & a_{12} & \dots & a_{1d} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nd} \end{bmatrix}$$

- 点乘的理解：

$$\boldsymbol{w} \bullet \boldsymbol{x} = \sum_{i=1}^{n} w_i x_i$$

- 对于几何的理解：

向量在N维空间的夹角↓

$$\boldsymbol{w} \bullet \boldsymbol{x} = |\boldsymbol{w}| \, |\boldsymbol{x}| \cos(\theta)$$

微积分

- 对于y = f(x)：

  偏导$\frac{dy}{dx}$是梯度或者斜率

  积分是$\int_{x=a}^{x=b} f(x)\,\mathrm{dx}$是曲线下的面积

- 对于多变量的功能 y = f(x_1, x_2 , …,x_p)

  偏导$\frac{\partial f}{\partial xi}$，对x求导假设其他变量都保持不变

  梯度向量表示为：

$$\nabla \boldsymbol{f} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{pmatrix}$$

最优化理论

deepshare.net
探享网

- 不受约束的最优化：min f(x)
- 受约束的最优化：

$$\min f(\boldsymbol{x})$$
$$\text{subject to } g_i(\boldsymbol{x}) \le b_i,\ \ i = 1, 2, .., m$$

- 梯度和海森矩阵：

$$\boldsymbol{\nabla f} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{pmatrix} \qquad \boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_p} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_p \partial x_1} & \frac{\partial^2 f}{\partial x_p \partial x_2} & \cdots & \frac{\partial^2 f}{\partial^2 x_p} \end{bmatrix}$$

- 梯度下降理论：

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \eta \boldsymbol{\nabla f}$$

- 牛顿法：

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - H^{-1} \boldsymbol{\nabla f}$$

- 拉格朗日乘数：

$$\min f(\boldsymbol{x})$$
$$\text{subject to } g_i(\boldsymbol{x}) \le b_i,\ \ i = 1, 2, .., m$$

$$F(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \left[b_i - g_i(\boldsymbol{x})\right]$$

概率论

单变量VS多变量

- 离散概率：P[X]
- 连续概率密度：p(x)
- 联合概率：P[X,Y]
- 条件概率：P[X|Y]
- 单变量高斯分布：
- 多变量高斯分布：

$$
\begin{aligned}
P[Y \mid X] &= \frac{P[X|Y]\; P[Y]}{P[X]} \\
P[X] &= \sum_{Y} P[X|Y]\; P[Y] \\
P[X, Y] &= P[X|Y]\; P[Y] \\
P[X] &= \sum_{y} P[X, Y] \\
&= \sum_{Y} P[X|Y]\; P[Y]
\end{aligned}
$$

$$
p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right\}
$$

$$
p(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}(\det \boldsymbol{C})^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^t\, \boldsymbol{C}^{-1}\, (\boldsymbol{x}-\boldsymbol{m})\right\}
$$

M表示平均值，是向量；
C代表协方差矩阵，它是对称半正定矩阵

deepshare.net
堔享网

联系我们：
电话：18001992849
邮箱：service@deepshare.net
Q Q： 2677693114

公众号

客服微信