统计机器学习 课后作业1

陈劭涵 17300180049

September 26, 2020

1 问题1

解:

似然函数:

$$Likelihood = \prod_{i=1}^{N} P(Y|X)$$

极大似然估计:

$$\max_{f\in\mathcal{F}} \prod_{i=1}^{N} P(Y|X) \iff \max_{f\in\mathcal{F}} log \prod_{i=1}^{N} P(Y|X) \iff \max_{f\in\mathcal{F}} \sum_{i=1}^{N} log(P(Y|X))$$

经验风险最小化:

$$\min_{f\in\mathcal{F}} R_{emp} = \min_{f\in\mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i))$$

对数风险函数:

$$L(Y, P(Y|X)) = -log(P(Y|X))$$

等价推导:

$$\begin{aligned}
\because \min_{f\in\mathcal{F}} R_{emp} &= \min_{f\in\mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i)) \\
&= \min_{f\in\mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} -log(P(Y|X)) \\
&= \min_{f\in\mathcal{F}} -\frac{1}{N} log \prod_{i=1}^{N} P(Y|X)
\end{aligned}$$

$$\because \min_{f\in\mathcal{F}} -\frac{1}{N} log \prod_{i=1}^{N} P(Y|X) \iff \max_{f\in\mathcal{F}} log \prod_{i=1}^{N} P(Y|X)$$

$$\therefore \min_{f\in\mathcal{F}} R_{emp} \iff \max_{f\in\mathcal{F}} \prod_{i=1}^{N} P(Y|X)$$

2 问题2

解:

由马尔科夫不等式和引理:

$$\begin{aligned} P(S_n - E(S_n) \geq t) &= P(e^{s(S_n - E(S_n))} \geq e^{st}) \\ &\leq \frac{1}{e^{st}} E(e^{s(S_n - E(S_n))}) \\ &= \frac{1}{e^{st}} \prod_{i=1}^{n} E(e^{s(X_i - E(X_i))}) \\ &\leq \frac{1}{e^{st}} \prod e^{\frac{s^2(b_i - a_i)^2}{8}} \\ &= exp\{-st + \frac{s^2 \sum\limits_{i=1}^{n} (b_i - a_i)^2}{8}\} \end{aligned}$$

观察等式右侧部分，视为关于s的函数:

$$exp\{-st + \frac{s^2 \sum\limits_{i=1}^{n} (b_i - a_i)^2}{8}\}$$

对该函数求导，得到其最大值:

$$g'(s) = 0 \iff s = \frac{4t}{\sum\limits_{i=1}^{n} (b_i - a_i)^2}$$

带入不等式即可得到结论:

$$P(S_n - E(S_n) \geq t) \leq exp(-\frac{2t^2}{\sum\limits_{i=1}^{n} (b_i - a_i)^2})$$

如果取反:

$$S_n = -S_n$$

则得到另一半结论：

$$P(E(S_n) - S_n \geq t) \leq exp(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2})$$

3 问题3

解：

(1) 问题背景：评估地块商品房的房价，房价预测
(2) 因变量和自变量：因变量是该地块商品房的房价；自变量是涉及地块房价的一系列因素如：城市发展水平，与市中心距离，周围配套设施状况，交通条件，政府政策等；
(3) 机器学习建模：通过从一些已被开发地块的商品房房价、地块信息和政府政策等作为训练数据，对这些数据进行一系列的数据处理之后通过回归方法建模得到房价与地块信息之间的关系，从而实现对未开发地块房价的评估和房价预测等；

4 问题4

解：

assume that:

$$x^T = \begin{pmatrix} x_1, & x_2, & ... & x_n \end{pmatrix}$$
$$y^T = \begin{pmatrix} y_1, & x_2, & ... & y_n \end{pmatrix}$$

（a）：

$$
\begin{aligned}
&\because y = Ax \\
&\ y_i = \sum_{j=1}^{n} a_{ij} x_j \\
&\therefore \frac{\partial y_i}{\partial x_j} = \frac{\partial \sum\limits_{j=1}^{n} a_{ij} x_j}{\partial x_j} = a_{ij} \\
&\because \frac{\partial y}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} \\
&\therefore \frac{\partial y}{\partial x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \\
&\iff \frac{\partial y}{\partial x} = A
\end{aligned}
$$

（b）：

$$
\begin{aligned}
&\because \alpha = y^T A x \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} y_i a_{ij} x_j \\
&\therefore \frac{\partial \alpha}{\partial x} = \frac{\partial (\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} y_i a_{ij} x_j)}{\partial x} \\
&= \left(\frac{\partial \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} y_i a_{ij} x_j}{\partial x_1}, \quad \frac{\partial \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} y_i a_{ij} x_j}{\partial x_2}, \quad \cdots \quad , \frac{\partial \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} y_i a_{ij} x_j}{\partial x_n} \right) \\
&= \left(\sum_{i=1}^{m} y_i a_{i1}, \quad \sum_{i=1}^{m} y_i a_{i2}, \quad \cdots \quad , \sum_{i=1}^{m} y_i a_{in} \right) \\
&= y^T A \\
&\therefore \frac{\partial \alpha}{\partial x} = y^T A
\end{aligned}
$$

(c)：

$$
\begin{aligned}
&\because \alpha = x^T A x \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j \\
&\therefore \frac{\partial \alpha}{\partial x} = \frac{(\partial \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i a_{ij} x_j)}{\partial x} \\
&= \left(\frac{\partial \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i a_{ij} x_j}{\partial x_1}, \quad \frac{\partial \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i a_{ij} x_j}{\partial x_2}, \quad \cdots \quad , \frac{\partial \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i a_{ij} x_j}{\partial x_n} \right) \\
&= \left(\sum_{i=1}^{n} x_i a_{i1} + \sum_{j=1}^{n} a_{1j} x_j, \quad \sum_{i=1}^{n} x_i a_{i2} + \sum_{j=1}^{n} a_{2j} x_j, \quad \cdots \quad , \sum_{i=1}^{n} x_i a_{in} + \sum_{j=1}^{n} a_{nj} x_j \right) \\
&= x^T \left(A[:,1] + A[1,:]^T, \quad A[:,2] + A[2,:]^T, \quad \cdots \quad , A[:,n] + A[n,:]^T \right) \\
&= x^T (A + A^T)
\end{aligned}
$$

(d)：

$$
\begin{aligned}
&\because \alpha = y^T A x \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} y_i a_{ij} x_j \\
&\therefore \frac{\partial \alpha}{\partial z} = \frac{\partial \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} y_i a_{ij} x_j}{\partial z} \\
&= \left(\frac{\partial \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} y_i a_{ij} x_j}{\partial z_1}, \quad \frac{\partial \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} y_i a_{ij} x_j}{\partial z_2}, \quad \cdots , \quad \frac{\partial \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} y_i a_{ij} x_j}{\partial z_l} \right) \\
&= \left(\sum_{i=1}^{m}\sum_{j=1}^{n} \tfrac{\partial y_i}{\partial z_1} a_{ij} x_j, \quad \sum_{i=1}^{m}\sum_{j=1}^{n} \tfrac{\partial y_i}{\partial z_2} a_{ij} x_j, \quad \cdots , \quad \sum_{i=1}^{m}\sum_{j=1}^{n} \tfrac{\partial y_i}{\partial z_l} a_{ij} x_j \right) \\
&+ \left(\sum_{i=1}^{m}\sum_{j=1}^{n} y_i a_{ij} \tfrac{\partial x_j}{\partial z_1}, \quad \sum_{i=1}^{m}\sum_{j=1}^{n} y_i a_{ij} \tfrac{\partial x_j}{\partial z_2}, \quad \cdots , \quad \sum_{i=1}^{m}\sum_{j=1}^{n} y_i a_{ij} \tfrac{\partial x_j}{\partial z_l} \right) \\
&= \left((\tfrac{\partial y}{\partial z_1})^T A x, \quad (\tfrac{\partial y}{\partial z_2})^T A x, \quad \cdots , \quad (\tfrac{\partial y}{\partial z_l})^T A x \right) + \left(y^T A \tfrac{\partial x}{\partial z_1}, \quad y^T A \tfrac{\partial x}{\partial z_2}, \quad \cdots , \quad y^T A \tfrac{\partial x}{\partial z_l} \right) \\
&= \left(x^T A^T \tfrac{\partial y}{\partial z_1}, \quad x^T A^T \tfrac{\partial y}{\partial z_2}, \quad \cdots , \quad x^T A^T \tfrac{\partial y}{\partial z_l} \right) + \left(y^T A \tfrac{\partial x}{\partial z_1}, \quad y^T A \tfrac{\partial x}{\partial z_2}, \quad \cdots , \quad y^T A \tfrac{\partial x}{\partial z_l} \right) \\
&= x^T A^T \left(\tfrac{\partial y}{\partial z_1}, \quad \tfrac{\partial y}{\partial z_2}, \quad \cdots , \quad \tfrac{\partial y}{\partial z_l} \right) + y^T A \left(\tfrac{\partial x}{\partial z_1}, \quad \tfrac{\partial x}{\partial z_2}, \quad \cdots , \quad \tfrac{\partial x}{\partial z_l} \right) \\
&= x^T A^T \frac{\partial y}{\partial z} + y^T A \frac{\partial x}{\partial z}
\end{aligned}
$$

(e)：

$$\begin{aligned}
&\because (\frac{\partial AB}{\partial \alpha})_{ij}\\
&= \frac{\partial \sum\limits_{k=1}^{n} a_{ik}b_{kj}}{\partial \alpha}\\
&= \sum_{k=1}^{n}\frac{\partial a_{ik}}{\partial \alpha}b_{kj} + \sum_{k=1}^{n}a_{ik}\frac{\partial b_{kj}}{\partial \alpha}\\
&= (A\frac{\partial B}{\partial \alpha})_{ij} + (\frac{\partial A}{\partial \alpha}B)_{ij}\\
&\therefore \frac{\partial AB}{\partial \alpha} = A\frac{\partial B}{\partial \alpha} + \frac{\partial A}{\partial \alpha}B\\
&\because AA^{-1} = I\\
&\therefore \frac{\partial AA^{-1}}{\partial \alpha}\\
&= A\frac{\partial A^{-1}}{\partial \alpha} + \frac{\partial A}{\partial \alpha}A^{-1}\\
&= \frac{\partial I}{\partial \alpha}\\
&= \mathbf{0}\\
&\iff A\frac{\partial A^{-1}}{\partial \alpha} + \frac{\partial A}{\partial \alpha}A^{-1} = 0\\
&\iff A\frac{\partial A^{-1}}{\partial \alpha} = -\frac{\partial A}{\partial \alpha}A^{-1}\\
&\iff \frac{\partial A^{-1}}{\partial \alpha} = -A^{-1}\frac{\partial A}{\partial \alpha}A^{-1}
\end{aligned}$$

5 问题5

解:

为了方便求解，由范数本身的定义，不难得知:

$$\underset{a}{min}\,\|Xa-y\| \iff \underset{a}{min}\,\|Xa-y\|^2$$

根据定义还可得知：

$$
\begin{aligned}
&\min_a \|Xa - y\|^2 \iff \frac{\partial \|Xa - y\|^2}{\partial a} = 0 \\
&\|Xa - y\|^2 = (Xa - y)^T (Xa - y) = a^T X^T X a - 2y^T X a + y^T y
\end{aligned}
$$

因此求其导数即可得出结论：

$$
\begin{aligned}
&\because \frac{\partial \|Xa - y\|^2}{\partial a} = 2X^T X a - 2X^T y = 0 \\
&and\ X^T X\ is\ nonsingular,\ so\ (X^T X)^{-1}\ exists \\
&\therefore \hat{a} = (X^T X)^{-1} X^T y
\end{aligned}
$$