HW1

1. 经验风险：$\min_{f\to\mathcal{F}} \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i)) \Leftrightarrow \min_{f\to\mathcal{F}} \frac{1}{N}\sum_{i=1}^{N}(-\log P(y_i|x_i))$.

$$= \max_{f\to\mathcal{F}} \frac{1}{N}\log\underbrace{\left(\prod_{i=1}^{N} P(y_i|x_i)\right)}_{\text{求最大值}}.$$

极大似然估计. $\max_{f\to\mathcal{F}} \prod_{i=1}^{N} P(y_i|x_i)$ 可见两者等价.

2. $\exp(sx) \le \frac{b-x}{b-a}\exp(sa) + \frac{x-a}{b-a}\exp(sb)$.

$$E\{\exp(sX)\} \le \frac{b-E(X)}{b-a}\exp(sa) + \frac{E(X)-a}{b-a}\exp(sb)$$

$$= \frac{b}{b-a}\exp(sa) - \frac{a}{b-a}\exp(sb).$$

$$= \exp\left\{sa + \ln\left(1+\frac{a-\exp(s(b-a))a}{b-a}\right)\right\}.$$

令 $L(s(b-a)) = sa + \ln\left(1+\frac{a-\exp(s(b-a))a}{b-a}\right)$.

$L(x) = \frac{a}{b-a}x + \ln\left(1+\frac{a-\exp(x)a}{b-a}\right)$. $L'(x) = \frac{a}{b-a} + \frac{b-a}{b-\exp(x)a}\cdot\frac{-a}{b-a}\exp(x) = \frac{a}{b-a} + \frac{-a}{b-\exp(x)a}\exp(x)$

$L''(x) = \left[\frac{-a}{b-\exp(x)a} + \frac{a}{[b-\exp(x)a]^2}\cdot(-a\exp(x))\right]\cdot\exp(x)$. $L''(0) = \frac{-a}{b-a} + \frac{-a^2}{(b-a)^2} = -\left(\frac{a}{b-a}+\frac{1}{2}\right)^2 + \frac{1}{4} \le \frac{1}{4}$. 当且仅当 $a+b=0$ 时取等.

$L(0) = 0$. $L'(0) = \frac{a}{b-a} + \frac{-a}{b-a} = 0$.

Taylor 展开. $L(x) = L(0) + L'(0)x + \frac{L''(0)}{2}x^2 + o(x^2)$

由 Taylor 中值定理，$\exists\, 0\le\theta\le 1$ 使 $L(x) = L(0) + L'(0)x + \frac{1}{2}L''(\theta x)x^2 \le \frac{1}{8}x^2$.

$\therefore E(\exp(sX)) \le \exp\left\{\frac{1}{8}s^2(b-a)^2\right\}$. 得证.

3. (1) 问题背景：关于信贷风险，根据用户行为预测该用户是否会在将来还款逾期.

(2) 因变量：该用户是否曾经发生过逾期行为.

自变量：该用户的行为（如月收入、月消费额、单笔最高支出等）.

(3) 机器学习建模：这是一个二分类问题，可以使用决策树、随机森林、逻辑回归、SVM 等分类器进行建模，在数据集上训练后即可得到模型.

4. (a) $y=(y_1,y_2,\dots,y_m)'$, $x=(x_1,x_2,\dots,x_n)'$.

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}_{m\times n}$$

$$y=Ax \quad \Rightarrow \quad y_i=\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_j.$$

$\therefore \dfrac{\partial y_i}{\partial x_j}=a_{ij}$ (因为A与x无关).

$$\therefore \frac{\partial y}{\partial x^T}=\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n}\end{pmatrix}_{m\times n}=(a_{ij})_{m\times n}=A. \quad 得证.$$

(b) $\alpha=y^TAx=(y^TA)_{1\times n}x.=x^TA^Ty=\alpha^T.$ (α是常数,故$\alpha=\alpha^T$).

本来根据(a)的结论,$\dfrac{\partial\alpha}{\partial x^T}=y^TA$,但是在本题中$y=\varphi(x)$,还需要对$y$进行求导.

$\dfrac{\partial\alpha}{\partial x^T}=y^TA+\dfrac{\partial\alpha}{\partial y}\cdot\dfrac{\partial y}{\partial x^T}=y^TA+x^TA^T\left(\dfrac{\partial y}{\partial x^T}\right).$ 得证.

(c) 本题对应(b)中$y=x$的情况.

$\therefore \dfrac{\partial\alpha}{\partial x^T}=x^TA+x^TA^T\cdot I=x^T(A+A^T).$ 得证.

(d) $\dfrac{\partial\alpha}{\partial z^T}=\dfrac{\partial\alpha}{\partial x}\cdot\dfrac{\partial x}{\partial z^T}+\dfrac{\partial\alpha}{\partial y}\cdot\dfrac{\partial y}{\partial z^T}$

$=y^TA\left(\dfrac{\partial x}{\partial z^T}\right)+x^TA^T\left(\dfrac{\partial y}{\partial z^T}\right)$ 得证.

(e) $AA^{-1}=I.$

$A^{-1}\dfrac{\partial A}{\partial\alpha}+\dfrac{\partial A^{-1}}{\partial\alpha}A=0$

同乘A^{-1}, $A^{-1}\cdot\dfrac{\partial A}{\partial\alpha}\cdot A^{-1}+\dfrac{\partial A^{-1}}{\partial\alpha}A\cdot A^{-1}=0. \Rightarrow \dfrac{\partial A^{-1}}{\partial\alpha}=-A^{-1}\cdot\dfrac{\partial A}{\partial\alpha}\cdot A^{-1}.$ 得证.

5. $\min\limits_{a}\sqrt{(Xa-y)^T(Xa-y)}.$

令$S=(Xa-y)^T(Xa-y)=a^TX^TXa-2a^TX^Ty+y^Ty.$

对目标函数S关于a求导.

$\dfrac{\partial S}{\partial a^T}=2X^TXa-2X^Ty.$ 令$\dfrac{\partial S}{\partial a^T}=0.$ 则 $X^TXa=X^Ty.$ 两边同左乘 $(X^TX)^{-1}$. (因为X^TX满秩).

$\therefore \hat{a}=(X^TX)^{-1}X^Ty.$