《最优化方法》2022 年秋季期末试卷(A 卷)

说明:
1. 本试卷考试时间 2 小时。
2. 本次考试不得提前交卷，所有人在正式开始考试满 2 小时之后，不得再动笔修改答卷。考试开始 2 小时后停笔然后将监考视频转到电脑摄像头，然后开始拍照答卷拼成 PDF 文件并上传至指定邮箱：zyh2023ruc@126.com。请尽快完成上传，确认无误之后，所有人退出考场。
3. 本试卷一共 10 道题，每题 10 分，卷面满分 100 分。
4. 本试卷中，如不加说明，则向量都是列向量。
5. 完成计算题和证明题的时候，应当有必要的步骤和简单的文字解释。
6. 考试结束时，提交一份 PDF 文档，内容是用纸笔完成的解答过程的照片。也就是说，每道题都是用纸笔完成。纸笔完成的试卷注意留存。
7. 涉及数值计算的时候，可以在列式之后用计算器或者计算机（用 Excel 或者代码实现均可）完成。但是求解最优化问题的时候不得通过调用软件包完成，而必须按照手工计算时代的原理和步骤完成。

一、(10 分) 设 $\mathbf{X}=(x_{ij})$ 是一个 $m\times n$ 实矩阵，$u=f(\mathbf{X})$ 为一实值函数，我们按如下规则记函数 u 对矩阵 $\mathbf{X}$ 的导数

$$\frac{\partial f}{\partial \mathbf{X}}=\begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{pmatrix}$$

证明如下两个结论:

(1) $\dfrac{\partial\, \mathrm{tr}(\mathbf{AX})}{\partial \mathbf{X}}=\mathbf{A}^{\mathrm{T}}$，其中 $\mathbf{A}$ 为 $n\times m$ 矩阵，$\mathbf{X}$ 为 $m\times n$ 矩阵;

(2) 设 $\mathbf{X}$ 为 $n\times n$ 可逆矩阵，则有 $\dfrac{\partial \det \mathbf{X}}{\partial \mathbf{X}}=(\det \mathbf{X})(\mathbf{X}^{-1})^{\mathrm{T}}$。

二、(10 分) 用拉格朗日乘子法求解最优化问题

$$\begin{aligned} \max_{x_1,x_2} \quad & x_1+x_2+x_1x_2 \\ \text{s.t.} \quad & x_1^2+x_2^2=1 \end{aligned}$$

三、(10 分) 在确定步长的方法中有一个所谓的 Wolf 条件，记最大步长为 α，必须满足如下不等式:

$$f(x^{(k)}+\alpha d^{(k)})\le f(x^{(k)})+\beta\alpha\nabla_{d^{(k)}} f(x^{(k)})$$

给定 $f(x)=x_1^2+x_2^2+3$，$\beta=10^{-4}$，$x^{(k)}=(-1,-1)^{\mathrm{T}},d^{(k)}=(1,0)^{\mathrm{T}}$，试求满足 Wolf 条件的最大步长 α。

四、(10 分) n 维欧氏空间中任意 k 个向量之间两两内积所组成的矩阵称为这 k 个向量的 Gram 矩阵，即

$$G(u_1,u_2,\dots,u_k)=\begin{pmatrix} u_1^{\mathrm{T}}u_1 & u_1^{\mathrm{T}}u_2 & \cdots & u_1^{\mathrm{T}}u_k \\ u_2^{\mathrm{T}}u_2 & u_2^{\mathrm{T}}u_2 & \cdots & u_2^{\mathrm{T}}u_k \\ \vdots & \vdots & \ddots & \vdots \\ u_k^{\mathrm{T}}u_1 & u_k^{\mathrm{T}}u_2 & \cdots & u_k^{\mathrm{T}}u_k \end{pmatrix}$$

按照教材上的通用记号，支持向量机的对偶问题具有如下形式

$$\begin{aligned} \max_{\alpha}\quad & \sum\nolimits_{i=1}^{m}\alpha_i-\frac{1}{2}\sum\nolimits_{i=1}^{m}\sum\nolimits_{j=1}^{m}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j \\ \text{s.t.}\quad & \sum\nolimits_{i=1}^{m}\alpha_i y_i=0 \\ & \alpha_i\ge 0, i=1,2,\dots,m. \end{aligned}$$

请利用 Gram 矩阵将上述最优化问题表示为标准的二次规划形式，并证明该问题存在最优解（当然，这个解在数值计算方面未必最优）。

五、(10 分) 我们知道，正态分布的密度函数形式具有某些最优性，比如在一定约束下可以取到最大熵。而这种形式的问题可以用经典的变分法求解。具体说来，μ,σ 为给定常数，定义在实数轴上的非负可导函数 $p(x)$ 满足如下条件

$$\int_{-\infty}^{+\infty}p(x)dx=1,\int_{-\infty}^{+\infty}xp(x)dx=\mu,\int_{-\infty}^{+\infty}(x-\mu)^2p(x)dx=\sigma^2$$

则在满足上述条件的函数中，正态密度函数可以取到最大熵。要求：说明变分原理的使用方式，然后可以用先猜后证的方式给出最终结果。

六、(10 分) 在使用共轭梯度下降的时候，目标函数为 $f(x,y)=x^2+xy+y^2+5$，迭代算法的初值为 $(x^{(0)},y^{(0)})=(1,1)^{\mathrm{T}}$。按照共轭梯度下降的原理用手工完成前两步递推计算并说明第二步之后得到的结果具有何种性质。

七、（10 分）利用 KKT 条件求解下面这个最优化问题，其中向量 $X=(x_1,x_2,x_3)^{\mathrm{T}}$

$$\begin{aligned} \min\quad & f(X)=x_1^2+x_2^2+x_3^2 \\ \text{s.t.}\quad & g_1(X)=2x_1+x_2-5\le 0 \\ & g_2(X)=x_1+x_3-2\le 0 \\ & g_3(X)=1-x_1\le 0 \\ & g_4(X)=2-x_2\le 0 \\ & g_5(X)=-x_3\le 0 \end{aligned}$$

八、(10 分) 本题有两问。

（1）考虑 Lasso 的单变量优化问题，即 β 为单变量，训练样本为 $\{y_i,z_i\}_{i=1}^{N}$，则最优化的目标函数可表示为

$$\min_{\beta}\left\{\frac{1}{2N}\sum_{i=1}^{N}(y_i-z_i\beta)^2+\lambda|\beta|\right\}$$

证明: 这个问题有最优解 $\hat{\beta}=S_{\lambda}\left(\frac{1}{N}\langle\mathbf{z},\mathbf{y}\rangle\right)$，其中软阈值算子为 $S_{\lambda}(x)=\mathrm{sgn}(x)(|x|-\lambda)_{+}$。

（2）关于 Lasso 和岭回归的对比，教材上有一个经典的对比图，请说明解读这张图的方法和理解的重点。

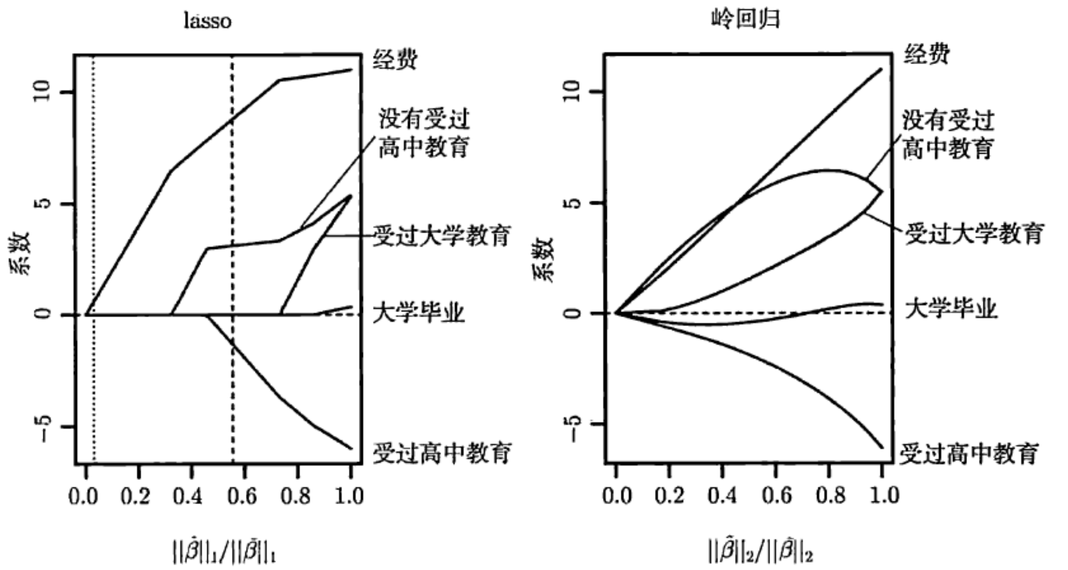
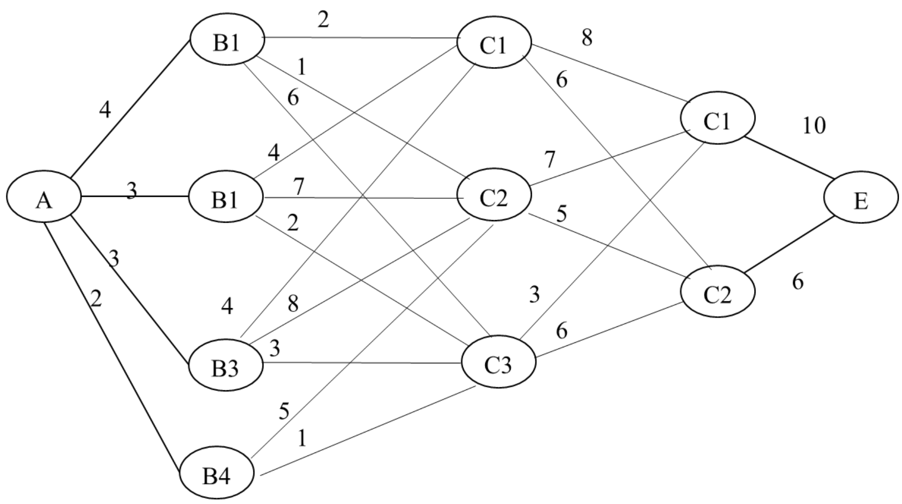


图 2-1　左图：横轴为系数向量 $\hat{\beta}$ 的 ℓ_1 范数与系数 $\tilde{\beta}$ 的 ℓ_1 范数之比。$\hat{\beta}$ 由 lasso 估计得出，$\tilde{\beta}$ 由无约束的最小二乘估计得出；右图：横轴为系数向量 $\hat{\beta}$ 的 ℓ_2 范数与系数 $\tilde{\beta}$ 的 ℓ_2 范数之比，$\hat{\beta}$ 由岭回归估计得到，$\tilde{\beta}$ 由无约束的最小二乘估计得到

九、(10 分) 用动态规划算法求从 A 到 E 的最短路径上的各个节点和最短路径长。为减少中间步骤的书写量，只需要满足两个对细节的要求：第一，简要展示从 B3 到 E 的最短路径的计算原理和过程；第二，直接写出各个节点到 E 的最短路径的长度。



十、(10 分) 约束下的最优化是个普遍适用的思考和数学建模框架，从 17 世纪的物理学到 20 世纪的人工智能，最优化问题的求解算法构成了现代科学研究和工程应用的基础。很多算法都可以类比于求导，于是根据不同的优化变量就发展出对求导、可行解空间、状态、邻域和搜索等等概念的封装和推广，结合本学期的课程举几个例子说明(越多越好，类型不要重复)。