2024阿里巴巴全球数学竞赛决赛试题

代数与数论

问题1

记$\mathrm{Mat}_2(\mathbb{Z})$为$2\times 2$整系数矩阵环，$R$为子环

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}_2(\mathbb{Z}) \middle| a \equiv d, c \equiv 0 \mod 2024 \right\}.$$

我们通过$\mathrm{Mat}_2(\mathbb{Z})$在$V := \mathbb{Q}^2$上自然的左作用将$V$看成一个左$R$-模。$V$中的一个$R$-格是指一个左$R$-子模$L \subset V$使得$L$作为$\mathbb{Z}$-模是有限生成的，并且满足$L \otimes_{\mathbb{Z}} \mathbb{Q} = V$。两个$R$-格是等价的，如果它们作为左$R$-模是同构的。找出$V$中$R$-格等价类的个数（有限或者无限）；并证明你的答案。

问题2

我们说多项式环$\mathbb{C}[x,y]$的一个理想I是缩放不变的，如果对于任意一对元素$\lambda, \mu \in \mathbb{C} \setminus \{0\}$，由所有形如$f(\lambda x, \mu y)$的元素（$f(x,y) \in I$）生成的理想就是$I$本身。找出$\mathbb{C}[x,y]$中满足$\dim_{\mathbb{C}} (\mathbb{C}[x,y]/I) = 6$的缩放不变的理想$I$的个数（有限或者无限）；并证明你的答案。

问题3

设n和d为正整数。设$F_1, \ldots, F_m$为$\mathbb{C}[X_0, \ldots, X_n]$中次数最多为$d$的齐次多项式使得

$$V(F_1, \ldots, F_m) := \{[x_0 : \cdots : x_n] \in \mathbb{CP}^n | F_1(x_0, \ldots, x_n) = \cdots = F_m(x_0, \ldots, x_n) = 0\}$$

是个有限集；这里$\mathbb{CP}^n$是指n维复射影空间。证明：$V(F_1, \ldots, F_m)$的元素个数至多是d^n。

问题4

设$p > 5$是一个素数。证明：方程

$$\prod_{k=1}^{(p-1)/2} (X - 2\cos(2\pi k/p)Y) = p^2$$

没有整数解。

问题5

设H是$\mathrm{GL}_4(\mathbb{R})$的由矩阵

$$\begin{pmatrix}1 & a & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix},\ \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & b & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix},\ \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & c\\ 0 & 0 & 0 & 1\end{pmatrix}$$

（$a,b,c\geq 0$）生成的幺半子群。选取元素$y_i,z_i\in H$（$i=1,2,3,\ldots$）使得序列$(y_iz_i)_{i\geq 1}$收敛。证明：存在$(1,2,3,\ldots)$的一个无穷子序列$(i_n)_{n\geq 1}$，使得序列$(y_{i_n})_{n\geq 1}$与$(z_{i_n})_{n\geq 1}$均收敛。

问题6

设p是一个素数。考虑半直积群$G=L\rtimes H$，其中L是一个循环p-群，H是一个有限循环群，以及一个有限生成的$\mathbb{F}_p[H]$-模M，满足$\mathrm{Hom}_H(L,\mathrm{End}_{\mathbb{F}_p}(M))=0$。这里，元素$h\in H$通过公式$(h\varphi)(m)=h\varphi(h^{-1}m)$（$m\in M$）作用在元素$\varphi\in\mathrm{End}_{\mathbb{F}_p}(M)$上。假设我们有一个$\mathbb{F}_p[H]$-模分解

$$M=M_1\oplus\cdots\oplus M_n,$$

使得对于任意$i\neq j$，$\mathrm{Hom}_H(M_i,M_j)=0$成立。证明：对任意正整数$d$，以及任意$(\mathbb{Z}/p^d\mathbb{Z})[G]$-模$N$满足$N\otimes_{\mathbb{Z}/p^d\mathbb{Z}}\mathbb{F}_p$作为$\mathbb{F}_p[G]$-模同构于$M$（这里我们将$H$自然地看作$G$的商群），都存在唯一一个$(\mathbb{Z}/p^d\mathbb{Z})[G]$-模分解

$$N=N_1\oplus\cdots\oplus N_n,$$

使得对于任意$1\leq i\leq n$，$N_i\otimes_{\mathbb{Z}/p^d\mathbb{Z}}\mathbb{F}_p\simeq M_i$成立。

几何与拓扑

问题1

设(M,g)是Ricci曲率非负的紧致黎曼流形. 假设任给正数δ, 均存在有限覆盖$\pi:\hat{M}\to M$使得$(\hat{M},\pi^*g)$的单射半径大于δ. 证明: (M,g)是平坦流形.

问题2

T^n为n维环面, $f: T^n \to T^n$是连续映射, 记$f_*: H_1(T^n;\mathbb{R})\to H_1(T^n;\mathbb{R})$为诱导映射. 假设$H_1(T^n;\mathbb{R})$上存在范数$||\cdot||$使得对$H_1(T^n;\mathbb{Z})$中的每个非零元$a$, 都存在一个正整数$k$使得$||f_*^k(a)||<||a||$, 其中$f^k$是$f$的$k$次迭代. 证明: f有不动点.
（向量空间V上的范数是一个映射$||\cdot||: V\to\mathbb{R}$满足如下条件:
- $||v||\geq 0$ 对所有$v\in V$成立, 且等号成立当且仅当$v=0$;
- $||\lambda v||=|\lambda|\cdot||v||$对所有$\lambda\in\mathbb{R}$ 和$v\in V$成立;
- $||u+v||\leq||u||+||v||$ 对所有$u,v\in V$成立.）

问题3

考虑3维复射影空间$\mathbb{C}P^3$中由
$$z_0^n+z_1^n+z_2^n+z_3^n=0$$
定义的复曲面, 其中n为正整数. 假设该复曲面上存在一个恰有素数个不动点的光滑$\mathbb{S}^1$作用, 证明: $n=1$, 并构造出相应的作用.

问题4

记N_g为连通不可定向的亏格为g的闭曲面. 设M为连通可定向的闭3维流形使得每个光滑嵌入的2维球面都是一个3维球体的边界, 并且N_1能光滑嵌入M中. 证明: N_g能光滑嵌入M中当且仅当g是奇数. （N_g表示g个$\mathbb{R}P^2$的连通和，g称为其亏格.）

问题5

构造一个紧致4维流形M, 其边界为3维环面, 且具有如下性质：存在一个内点P以及$M\setminus\{P\}$上4个处处线性无关的向量场X_1,X_2,X_3,X_4, 使得X_1,X_2,X_3在边界上的限制构成3维环面上的左不变标架场. 对M 上的光滑度量g, 设其在边界附近的限制是$I\times S^1\times S^1\times S^1$上的乘积度量, 计算$\int_M(-\frac{\mathrm{Tr}(\Omega^2)}{8\pi^2})$, 其中$\Omega$是$g$的Levi-Civita联络的曲率形式.

问题6

设Σ是$\mathbb{R}^{n+1}$ $(n\geq 4)$中的嵌入超曲面, 其上诱导度量记为g. 假设对Σ上任意一点p, 存在局部坐标$(x_1,\ldots,x_n)$以及光滑函数u使得$g=e^u(\sum_i dx_i\otimes dx_i)$. 证明：对$\Sigma$上任意一点$p$，某个主曲

率的重数至少为$n-1$. 进一步假设非脐点集合U是非空的, 证明: U上由重数$n-1$的主方向构成的分布是可积的.

分析与方程

问题1

给定常数$\omega > 0$，考虑$\mathbb{R}$上满足方程

$$x\frac{d^2u}{dx^2} + (1-\omega^2 x)u = 0$$

不恒等于0的缓增分布$u \in \mathscr{S}'(\mathbb{R})$。

(1) 证明$u \in C(\mathbb{R}) \cap L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ (这里空间L^p是由$\mathbb{R}$上勒贝格测度给出)。

(2) 计算

$$A := \frac{\left|\int_{\mathbb{R}} u(x)dx\right|^2}{\int_{\mathbb{R}} |u(x)|^2 dx}.$$

(3) 对$\omega = \frac{1}{2}$，求u的显式表达式。

问题2

记B是所有定义在实轴$\mathbb{R}$上正的以2π为周期并且满足如下条件

$$f > 0, \quad \int_0^{2\pi} (f''(x))^2 dx \le 1, \quad \forall f \in B$$

的光滑周期函数构成的集合。对$k > 0$，记$S(k)$是所有满足条件

$$\sup_{f\in B} \int_0^{2\pi} \frac{|f'(x)|^k}{(f(x))^\alpha} dx < \infty$$

的实数α组成的集合。

(1) 证明$S(4)$是闭区间，并且找出$S(4)$的最大值。

(2) 证明存在常数C使得

$$|f'(x)| \le C f^{\frac{1}{3}}(x), \quad \forall f \in B.$$

(3) 证明$S(2024)$是闭区间，并求$S(2024)$的最大值。

问题3

假设M和Q是给定的正的常数。定义函数

$$f(r) = 1 - \frac{M}{r^2} + \frac{Q}{r^4} - r^2, \quad r > 0.$$

如果f有三个不同的正根$r_c > r_+ > r_- > 0$，证明$f'(r_+) + f'(r_-) < 0$。

问题4

定义序列

$$a_{n+1} = a_n + \frac{a_n^2}{n^2}, \quad 0 \le a_1 < 1.$$

证明极限 $\lim_{n\to\infty} a_n$存在并且有限。

问题5

假设Ω是$\mathbb{R}^d$中连通的开集使得其补集$\mathbb{R}^d \backslash \Omega$ 包含一个开的锥$\mathcal{C}$。假设$u : \bar{\Omega} \to \mathbb{R}$是有界的连续函数，在$\Omega$中是$C^2$的，并且满足

$$\begin{cases} \Delta u(x) \ge 0 & , \quad x \in \Omega, \\ u(x) \le 0 & , \quad x \in \partial\Omega. \end{cases}$$

证明

$$u(x) \le 0, \quad \forall x \in \Omega.$$

这里开的锥$\mathcal{C}$指的是存在顶点x_0，非零方向$v \in \mathbb{R}^d$以及$\theta \in (0, \frac{\pi}{2})$使得

$$\mathcal{C} = \{x : |x - x_0||v| \cos\theta < v \cdot (x - x_0)\}.$$

问题6

假设$\mathcal{F}$是所有单调增1-Lipschitz函数$f : [0,1] \to [0,1]$构成的集合，即是f满足

$$0 \le f(x) - f(y) \le x - y, \quad \forall 0 \le y \le x \le 1, \quad \forall f \in \mathcal{F}.$$

(1) 证明对任意$\epsilon > 0$，存在只依赖ϵ的常数$\tau > 0$使得对任意$f \in \mathcal{F}$，存在区间$[a,b] \subset [0,1]$使得$b - a > \tau$，$\sigma = \sigma(a,b) := \frac{f(b)-f(a)}{b-a}$，

$$f(a) + \sigma(x-a) + \epsilon(b-a) \ge f(x) \ge f(a) + \sigma(x-a) - \epsilon(b-a), \quad \forall x \in [a,b].$$

(2) 证明对任意$\epsilon > 0$，存在只依赖于ϵ的常数$\tau > 0$使得对任意$f \in \mathcal{F}$，存在$[0,1]$区间的分割$\{[a_j, a_{j+1}]\}_{j=1}^{J}$以及$0 \le \sigma_j < \sigma_{j+1} \le 1$使得$a_{j+1} - a_j \ge \tau$，

$$f(x) \ge f(a_j) + \sigma_j(x - a_j) - \epsilon(a_{j+1} - a_j), \forall x \in [a_j, a_{j+1}],$$

并且

$$\sum_j \sigma_j(a_{j+1} - a_j) \ge f(1) - \epsilon.$$

应用与计算数学

问题1

设$A, B \in \mathbb{C}^{n\times n}$，$\lambda(A), \lambda(B)$分别为矩阵$A, B$的特征值集合，设矩阵$A$的Jordan分解为$A = P^{-1}JP$，其中$J$为$A$的Jordan标准形，对于任意的$\mu \in \lambda(B)$，证明：

(1) 若$\mu \notin \lambda(A)$，则必有$\|(\mu I - J)^{-1}\|_2^{-1} \leq \theta$;

(2) $\min_{\lambda\in\lambda(A)} |\lambda - \mu| \leq 2(1+\theta)\ln(1+\theta^{1/m})$;

其中$\theta = m\kappa_2(P)\|A - B\|_2$，$\kappa_2(P) = \|P\|_2\|P^{-1}\|_2$为矩阵$P$在2-范数下的条件数，$m$ 为J中最大Jordan块的阶数.

问题2

假设$F(x;w)$是一个输出标量的深度神经网络,其中x是输入，w表示权重。假设F关于w连续可微，并且对于训练数据$\{x_j, y_j\}_{j=1}^m$过参数化：即存在$w^\star$使得对所有j满足$F(x_j, w^\star) = y_j$。为了研究训练神经网络时在$w^\star$的局部优化动力学，我们考虑线性化神经网络$\widetilde{F}(x;w) = F(x;w^\star) + (w - w^\star)^\top \nabla F(x;w^\star)$，其损失函数为

$$\mathrm{Loss}(w) := \frac{1}{2m}\sum_{j=1}^{m}(y_j - \widetilde{F}(x_j;w))^2.$$

令s表示学习率，梯度下降法为$w_{i+1} = w_i - s\nabla \mathrm{Loss}(w_i)$，而随机梯度下降法为$w_{i+1} = w_i - s(\nabla \mathrm{Loss}(w_i) + \epsilon_i)$，其中噪声项$\epsilon_i$满足$\mathbb{E}\epsilon_i = 0$ 和$\mathbb{E}\epsilon_i\epsilon_i^\top = M(w_i)/b$，$b$是mini-batch的大小。假设协方差矩阵$M$与

$$\Sigma = \frac{1}{m}\sum_{j=1}^{m}\nabla F(x_j, w^\star)\nabla F(x_j, w^\star)^\top$$

在以下意义上对齐：

$$\frac{\mathrm{Tr}(M(w)\Sigma)}{2\mathrm{Loss}(w)\|\Sigma\|_F^2} \geq \delta$$

对于$\delta > 0$和所有w成立。这里$\|\cdot\|_F$表示Frobenius 范数。

(1) 对于梯度下降，证明如果Σ的谱范数满足

$$\|\Sigma\|_2 \leq \frac{2}{s},$$

则梯度下降是局部稳定的（即对所有i，$\mathrm{Loss}(w_i)$是有界的）。(注意，这蕴含了一个依赖维度的界：$\|\Sigma\|_F \leq \frac{2\sqrt{d}}{s}$，其中$d$是$w$的维度。)

(2) 对于随机梯度下降，如果$\mathbb{E}\mathrm{Loss}(w_i)$对所有$i$都有界，则以下独立于维度的不等式必须成立：

$$\|\Sigma\|_F \leq \frac{\sqrt{b/\delta}}{s}.$$

问题3

考虑以下系统：

$$\partial_t f + v \cdot \nabla_x f + E \cdot \nabla_v f = 0, \quad \partial_t E = -\int_{\mathbb{R}^d} v f \mathrm{d}v \,. \tag{1}$$

其中变量有$t \in \mathbb{R}^+$, $x \in \mathbb{R}^d$, $v \in \mathbb{R}^d$, f 是一个关于t, x 和v 的未知函数，E 只关于t 和x.

a) 证明质量和能量

$$m(t) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f \mathrm{d}v \mathrm{d}x \,, \qquad e(t) := \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |v|^2 f \mathrm{d}x \mathrm{d}v + \frac{1}{2} \int_{\mathbb{R}^d} |E|^2 \mathrm{d}x$$

随时间保持不变。

b) 现假设$E(t, x)$ 已知, 则(1) 简化为

$$\partial_t f + v \cdot \nabla_x f + E \cdot \nabla_v f = 0. \tag{2}$$

我们将使用粒子方法来求解(2)，初始条件为：

$$f(t = 0, x, v) = f_{\mathrm{in}}(x, v).$$

考虑由$\{x_p(t), v_p(t)\}_{1 \le p \le P}$ 表示P 个粒子, 其中$x_p(t)$ 表示第P 个粒子的位置，$v_p(t)$ 表示其速度, 每个粒子的权重为$\frac{1}{P}$。$x_p(t)$ 和$v_p(t)$ 的演化由以下方程决定：

$$\frac{dx_p}{dt} = v_p, \quad \frac{dv_p}{dt} = E(x_p) \,.$$

证明

$$f^P(t, x, v) := \frac{1}{P} \sum_{p=1}^{P} \delta(x - x_p(t)) \delta(v - v_p(t))$$

在分布意义下是(2) 的弱解，其中(2) 的初始条件是$f_0^P = \frac{1}{P} \sum_{p=1}^{P} \delta(x - x_p(0)) \delta(v - v_p(0))$。这里$\delta$ 是Dirac-Delta 函数。

c) 回到原始系统(1)。考虑以下粒子方法

$$\begin{aligned} x_p^{n+1} &= x_p^n + \Delta t v_p^n \\ v_p^{n+1} &= v_p^n - E(x_p^n) \Delta t \\ E_i^{n+1} &= E_i^n - J_i^n \Delta t \,. \end{aligned}$$

这里上标n表示时间步长，$E(x_p^n)$和J_i^n分别定义如下

$$E(x_p^n) := \sum_i S(x_i - x_p^n) E_i^n \Delta x^d \,,$$

$$J_i^n := \frac{1}{P} \sum_{p=1}^{P} S(x_i - x_p^n) v_p^n \,.$$

在这些表达式中，x_i表示具有网格大小Δx的均匀网格，S是一个满足$\int_{\mathbb{R}^d} S(x)\mathrm{d}x = 1$. 这种方法是否保留能量？也就是说，我们是否有

$$\frac{1}{2P}\sum_p (v_p^n)^2 + \frac{1}{2}\sum_i (E_i^n)^2 \Delta x^d = \frac{1}{2P}\sum_p (v_p^{n+1})^2 + \frac{1}{2}\sum_i (E_i^{n+1})^2 \Delta x^d\,?$$

如果是，证明它。如果不是，构造一个在离散级别上保持能量的离散格式。

问题4

问题1:
考虑如下两个优化问题:

$$(\mathrm{A}):\ \begin{array}{ll} \min\limits_{\mathbf{x}} & f(\mathbf{x}), \\ \text{s. t.} & \mathbf{g}(\mathbf{x}) = 0, \\ & \mathbf{x}_i \in \mathcal{F}_i,\ i = 1, \ldots, n \end{array} \qquad \text{和} \qquad (\mathrm{B}):\ \begin{array}{ll} \min\limits_{\mathbf{x}} & f(\mathbf{x}) + \beta \mathbf{g}(\mathbf{x})^\top \mathbf{1}_p, \\ \text{s. t.} & \mathbf{x}_i \in \mathcal{F}_i,\ i = 1, \ldots, n, \end{array}$$

其中$\mathbf{x} := [\mathbf{x}_1^\top, \ldots, \mathbf{x}_n^\top]^\top \in \mathbb{R}^m$ ($m,\ n \in \mathbb{N}$), 且对于$i = 1, \ldots, n$, $\mathbf{x}_i \in \mathbb{R}^{m_i}$ ($m_i \in \mathbb{N}$), 满足$\sum_{i=1}^n m_i = m$. 函数$f : \mathbb{R}^m \to \mathbb{R}$和$\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^p$ ($p \in \mathbb{N}$)是多仿射的, 即对任意$i \in \{1, \ldots, n\}$, 在固定下标在$\{1, \ldots, N\} \setminus \{i\}$中的块后, 它们对$\mathbf{x}_i$ 是仿射函数. 这里, 我们称一个函数$\mathbf{h} : \mathbb{R}^q \to \mathbb{R}^r$ ($q,\ r \in \mathbb{N}$) 是仿射的, 若对任意$a \in \mathbb{R}$ 以及$\mathbf{y}^{(1)}$, $\mathbf{y}^{(2)} \in \mathbb{R}^q$, 有

$$\mathbf{h}(a\mathbf{y}^{(1)} + (1-a)\mathbf{y}^{(2)}) = a\mathbf{h}(\mathbf{y}^{(1)}) + (1-a)\mathbf{h}(\mathbf{y}^{(2)}).$$

对任意$i \in \{1, \ldots, n\}$, 集合$\mathcal{F}_i \subseteq \mathbb{R}^{m_i}$是一个有界多面体. 函数$\mathbf{g}$ 在$\bigtimes_{i=1}^n \mathcal{F}_i$上非负, 其中"$\times$"表示集合的笛卡尔积. 此外, 在问题(B)中, 标量β 是一个实数, $\mathbf{1}_p \in \mathbb{R}^p$代表$p$维全一向量.

请证明如下三个结论.

1. 对任意$\beta \in \mathbb{R}$, 问题(B)至少有一个最优解是可行域的极点(即极点最优解).
2. 存在$\bar{\beta} \in \mathbb{R}$, 使得对任意$\beta \geq \bar{\beta}$, 问题(B)的极点最优解都是问题(A)的最优解.
3. 存在$\tilde{\beta} \in \mathbb{R}$, 使得对任意$\beta \geq \tilde{\beta}$, 问题(A)和(B)的最优解集相同.

问题5

考虑下面的由随机微分方程组(SDEs)给出的粒子系统

$$\mathrm{d}x_i^t = -x_i^t \mathrm{d}t - \frac{1}{N}\sum_{j=1}^N \nabla W(x_i^t - x_j^t)\mathrm{d}t + \sqrt{\frac{2}{\beta}}\mathrm{d}B_t^i, \quad i = 1, 2, \cdots, N.$$

这里每个粒子x_i^t属于欧氏空间$\mathbb{R}^d$, 参数$\beta > 0$表示温度的倒数. 注意到$(B_\cdot^i)$表示$\mathbb{R}^d$上的N个独立的布朗运动. 我们另外假设$W \in C_c^\infty(\mathbb{R}^d)$, $W(0) = 0$, 并且对于任意$x \in \mathbb{R}^d$, $W(x) = W(-x)$.

1. 用$\rho_N(t, \cdot)$表示这N个粒子的联合分布. 请清晰地写出ρ_N满足的Fokker-Planck方程.

2. 定义的Gibbs分布M_N为

$$M_N(x_1,x_2,\cdots,x_N)=\frac{1}{Z_N}\exp\Big(-\beta\Big(\sum_{i=1}^{N}\frac{1}{2}|x_i|^2+\frac{1}{N}\sum_{1\le i<j\le N}W(x_i-x_j)\Big)\Big),$$

这儿我们选取常数Z_N使得M_N是一个概率密度函数. 证明M_N是ρ_N 满足的Fokker-Planck方程的唯一的平稳解.

3. 假设一个概率密度函数ρ满足下面的非线性方程

$$\rho=\frac{1}{Z}\exp\Big(-\beta\big(\frac{1}{2}|x|^2+W*\rho(x)\big)\Big),$$

这儿归一化常数Z定义为

$$Z=Z(\rho)=\int_{\mathbb{R}^d}\exp\Big(-\beta\big(\frac{1}{2}|x|^2+W*\rho(x)\big)\Big)\mathrm{d}x.$$

证明存在一个临界的常数$\beta_c>0$, 使得当$\beta<\beta_c$时, 我们有结论

$$\sup_{N\ge2}\int_{(\mathbb{R}^d)^N}M_N\log\frac{M_N}{\rho^{\otimes N}}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_N<\infty,$$

这儿$\rho^{\otimes N}(x_1,x_2,\cdots,x_N)=\rho(x_1)\rho(x_2)\cdots\rho(x_N)$.

提示: 对于第三部分, 可以直接使用下面的结论: 存在一个常数$c_0>0$, 使得当$\|\phi\|_{L^\infty}\le c_0$ 时, 我们有

$$\int_{(\mathbb{R}^d)^N}\rho^{\otimes N}\exp\Big(N\int_{\mathbb{R}^{2d}}\phi(x,y)\mathrm{d}(\mu_N(x)-\rho(x))\mathrm{d}(\mu_N(y)-\rho(y))\Big)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_N,$$

是关于N一致有界的, 这里ρ是任一概率测度, $\mu_N=\frac{1}{N}\sum_{i=1}^N\delta_{x_i}$ 表示点$(x_1,x_2,\cdots,x_N)\in(\mathbb{R}^d)^N$ 对应的经验测度.

问题6

研究大模型的缩放定律对减少其训练开销至关重要. 即，最终的测试损失是如何随着训练步数和模型大小的变化而变化的？本题中，我们研究训练线性模型时的缩放定律.

1. 在本小问中，我们考虑使用梯度下降学习一个一维线性模型的情况.

   - 定义数据分布$\mathcal{D}$ 为一个$\mathbb{R}^2$ 上的分布. 每个数据是一个数对(x,y)，分别代表输入和输出，并服从分布$x\sim\mathcal{N}(0,1)$，$y\sim\mathcal{N}(3x,1)$.
   - 用梯度下降算法学习线性模型$f_w(x)=w\cdot x$，其中$w,x\in\mathbb{R}$. 初始化$w_0=0$并进行多步迭代. 每次迭代时，从$\mathcal{D}$中采样(x_t,y_t)，然后更新w_t为$w_{t+1}\leftarrow w_t-\eta\nabla\ell_t(w_t)$，其中$\ell_t(w)=\frac{1}{2}(f_w(x_t)-y_t)^2$是平方损失函数，$\eta>0$ 是学习率.

   设学习率$\eta\in(0,\frac{1}{3}]$，那么$T\ge0$ 步迭代之后的测试损失的期望

   $$\overline{\mathcal{L}}_{\eta,T}=\mathbb{E}_{w_T}\mathbb{E}_{(x,y)\sim D}[\frac{1}{2}(f_{w_T}(x)-y)^2]$$

   是多少？

2. 现在我们在第一小问的设定下，考虑学习率η 被调到最优的情况. 求函数$g(T)$，使得当$T \to +\infty$时，以下条件成立：

$$\left| \inf_{\eta \in (0, \frac{1}{3}]} \overline{\mathcal{L}}_{\eta,T} - g(T) \right| = O\left(\frac{(\log T)^2}{T^2} \right)$$

3. 一个常常被观测到的实验现象是大语言模型的预训练过程大致遵循Chinchilla 缩放定律：

$$\overline{\mathcal{L}}_{N,T} \approx \frac{A}{N^\alpha} + \frac{B}{T^\beta} + C,$$

其中$\overline{\mathcal{L}}_{N,T}$ 是在经过T 步训练后具有N 个参数的模型的测试损失的期望，A, B, α, β, C 是常数. 现在，我们来举一个训练多维线性模型的例子，使其也遵循类似的缩放定律.

- 固定$a > 0, b \geq 1$. 每个数据$(x_\bullet, y)$ 由一个输入和输出组成，其中输入$x_\bullet$ 是一个无限维向量（可看作一个序列），输出y 满足$y \in \mathbb{R}$. 定义数据分布$\mathcal{D}$ 如下. 首先，从Zipf 分布中采样k，$\Pr[k = i] \propto i^{-(a+1)}$ $(i \geq 1)$. 令$j := \lceil k^b \rceil$. 然后，从$\mathcal{N}(0, 1)$中采样得到$x_\bullet$ 的第j个坐标x_j，并令其余坐标为0. 最后，$y \sim \mathcal{N}(3x_j, 1)$. 这样得到的$(x_\bullet, y)$ 的分布即数据分布$\mathcal{D}$.
- 我们研究一个仅关注前N个输入坐标的线性模型. 定义函数$\phi_N(x_\bullet) = (x_1, \ldots, x_N)$. 我们研究的线性模型具有参数$\boldsymbol{w} \in \mathbb{R}^N$，输出为$f_{\boldsymbol{w}}(x_\bullet) = \langle \boldsymbol{w}, \phi_N(x_\bullet) \rangle$.
- 我们使用梯度下降算法来学习该线性模型. 初始化$\boldsymbol{w}_0 = 0$ 并进行多步迭代. 每次迭代时，从$\mathcal{D}$中采样$(x_{t,\bullet}, y_t)$，然后更新$\boldsymbol{w}_t$为$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - \eta \nabla \ell_t(\boldsymbol{w}_t)$，其中$\ell_t(\boldsymbol{w}) = \frac{1}{2}(f_{\boldsymbol{w}}(x_{t,\bullet}) - y_t)^2$.

令$\overline{\mathcal{L}}_{\eta,N,T} = \mathbb{E}_{\boldsymbol{w}_T} \mathbb{E}_{(\boldsymbol{x},y)\sim D}[\frac{1}{2}(f_{\boldsymbol{w}_T}(\boldsymbol{x}) - y)^2]$为以学习率$\eta \in (0, \frac{1}{3}]$ 对具有N 个参数的线性模型进行$T \geq 0$步训练后的测试损失的期望.

请求出α, β, C，使得$\exists \gamma > 0, \forall c > 0$，当$T = N^{c+o(1)}$且$N$足够大时，以下条件成立：

$$\epsilon(N,T) := \frac{\inf_{\eta \in (0,\frac{1}{3}]} \overline{\mathcal{L}}_{\eta,N,T} - C}{\frac{1}{N^\alpha} + \frac{1}{T^\beta}}, \qquad (\log N + \log T)^{-\gamma} \leq \epsilon(N,T) \leq (\log N + \log T)^{\gamma}.$$

即$\inf_{\eta \in (0,\frac{1}{3}]} \overline{\mathcal{L}}_{\eta,N,T} = \tilde{\Theta}(N^{-\alpha} + T^{-\beta}) + C$，其中$\tilde{\Theta}$表示忽略任何关于$\log N$和$\log T$的多项式.

组合与概率

问题1

令 m 为一正整数.考虑一个整数集合 $\mathbb{Z}$ 上的马氏链 $X=(X_n)_{n\geq 0}$，其转移概率 $p_{i,j}:=\mathbb{P}[X_{n+1}=j|X_n=i]$ 满足如下条件：(1). $p_{i,j}\neq 0$ 当且仅当 $|j-i|=1$; (2). 当 $j-i=m$ 时 $p_{i,i+1}=p_{j,j+1}$. 令 $Y_n=X_n \mod m$. 则 $Y=(Y_n)_{n\geq 0}$ 可看成一个状态空间为 $\{0,1,\ldots,m-1\}$ 的马氏链. 令 $(\mu_i)_{0\leq i<m}$ 为 Y 的平稳概率分布. 令 $A=\sum_{i=0}^{m-1}\mu_i p_{i,i+1}$. 令 $T=\inf\{n\geq 0: X_n=m\}$. 证明：如果 $A>\frac{1}{2}$，则 $(2A-1)\mathbb{E}[T\mid X_0=0]=m$.

问题2

一个有向图 G 被称作简单图, 如果它没有自环且任意两点间最多连一条边. 设 u,v 为 $V(G)$ 中两个不同顶点. 我们将从 u 射向 v 的边记作 $u\to v$, 并且记 u 是 v 的入邻居, v 是 u 的出邻居. 从 u 到 v 的距离 $d(u,v)$ 为 G 中从 u 到 v 的最短有向路的长度. 对于整数 $j\geq 1$, 定义 $N_j^+(u)$ 为满足 $d(u,v)=j$ 的顶点 $v\in V(G)$ 构成的集合.

设 G 是一个简单有向图使得对于任意 $u\in V(G)$, u 的入邻居与 u 的出邻居数目相等. 假设 G 中不存在三个顶点 u,v,w 满足 $u\to v, v\to w$ 且 $u\to w$. 证明

$$\sum_{v\in V(G)}|N_2^+(v)|\geq\sum_{v\in V(G)}|N_1^+(v)|.$$

问题3

假设 Y 是取值在 $(-1,1)$上的随机变量. 对于一个实数 $y\in(-1,1)$, 令其二进制展开为

$$y=\sum_{k=1}^{\infty}a_k 2^{-k},\ \text{其中}a_k\in\{-1,1\}\ \text{对每个}k\in\mathbb{N}.$$

这里, 二进制展开是唯一的, 因为我们规定不存在 k_0 使得对所有的 $n\geq k_0$ 都有 $a_n=1$. 对每个 $s\in\mathbb{N}$, 令 $y_s=\sum_{k=1}^{s}a_k 2^{-k}$, 即 y 的最初 s 位展开. 如上定义随机过程 $(Y_s)_{s\in\mathbb{N}}$, 即逐渐读取 Y 的二进制展开. 注意到 $Y_s\to Y$ 当$s\to\infty$ (点点收敛). 对一个严格增函数 $g:[-1,1]\to\mathbb{R}$, 定义随机过程 $(X_s)_{s\in\mathbb{N}}:=(g(Y_s))_{s\in\mathbb{N}}$. 假设 $(X_s)_{s\in\mathbb{N}}$ 是一个关于其自然 σ-代数族的鞅. 证明存在常数 $r<0.9$ 和 $C>0$ 使得

$$\mathbb{E}[(X_s-g(Y))^2]\leq Cr^s\ \text{对所有的}s\in\mathbb{N}.$$

(数值 0.9 并非最优.)

提示: 你可以尝试证明并使用以下结果: 假设 $k\geq 3$ 并且正实数 $a_1,\ldots,a_k,b_1,\ldots,b_k$ 满足

$$b_{s-2}\geq\min\{a_s,a_{s-1}\}\ \text{对于}s=3,\ldots,k;$$
$$b_{s-1}\geq\min\{a_s,b_s\}\ \text{对于}s=2,\ldots,k.$$

则存在 $r<0.9$ 使得

$$\prod_{s=1}^{k}\frac{a_s}{a_s+b_s}\leq\sqrt{\frac{b_1}{b_k}}\,r^{k-2}.$$

问题4

假设 C 是 $\mathbb{R}^2$ 中的一个凸图形, 其面积为 1, 并假设 $\mathcal{S}$ 是 $\mathbb{R}^2$ 中的一组（可能是无限个）凸图形. 对于 $\mathcal{S}$ 中的每个凸图形 D, 存在一个常数 $k \in \mathbb{R}$ 使得 $D = kC := \{k\vec{x}\colon \vec{x} \in C\}$. 如果存在一个映射 $t\colon \mathcal{S} \to \mathbb{R}^2$, 使得对于 $\mathcal{S}$ 中的每个 D, D 通过向量 $t(D)$ 平移后的内部包含在 C 内, 并且任何 $\mathcal{S}$ 中两个不同的 D 和 D' 在通过 $t(D)$ 和 $t(D')$ 平移后, D 的内部不与 D' 的内部重叠, 那么我们说 $\mathcal{S}$ 中的凸形状可以仅通过平移的方式被填充到 C 内. 证明如果 $\mathcal{S}$ 中凸形状的总面积最多为 1/8, 则它们可以仅通过平移的方式被填充到 C 内.

问题5

在第 0 天，某债券价值 1 元. 在第 n 天，其价值为 $S_n := \exp(X_1 + \cdots + X_n)$ 元，其中 X_i 是独立同分布随机变量，满足 $P(X_i = 1) = P(X_i = -1) = 1/2$. 爱丽丝有一些直到第 N 天都可投资该债券的闲钱.作为一位有独特投资理念的投资人，她只愿意在“信号时刻”购入债券.所谓信号时刻，指的是某个日子 K，其中 $K \in [1, \ldots, N-1]$，并且该债券在第 K 天的价格是第 0 天到第 K 天中最高的，却是第 K 天到第 N 天中最低的.

当然即使这样的日子 K 存在，她也无法在当时确认这是否就是“信号时间”.但事后来看，我们很自然地想知道，这样的信号时间是否真的存在.试证明:存在不依赖于 N 的常数 $c, C > 0$ 使得这样的日子 K 存在的概率 $\in (c/\log N, C/\log N)$.

提示: 定义 $p_n = P[S_i \geq 1,\ \forall i = 1, \ldots, n]$ 并注意到

$$p_n^2 \leq P[1 \leq S_i \leq S_n,\ \forall 1 \leq i \leq n] \leq p_{\lfloor n/2 \rfloor}^2.$$

[如果你能证明该提示，你也能得到部分分数！]

问题6

给定图 G, 我们称 σ 是它的一个合法 q-染色, 如果 σ 赋给每个点 q 个颜色中的一个, 并且没有任何一条边的两个端点是同色的. 给定一个染色 σ 和一个顶点 $v \in V$, 令 $L_\sigma(v)$ 为点 v 处可用的颜色的集合, 即在染色 σ 下没有出现在 v 的邻居的颜色的集合.

证明存在一个正整数 $d_0 \geq 1$ 使得对任意整数 $d \geq d_0$, 以下命题成立: 对任意的没有三角形且最大度数为 d 的图 $G = (V, E)$, 以及它的任意一个顶点 $v \in V$,

$$\mathbb{E}_\sigma[|L_\sigma(v)|] \geq d/3,$$

这里 σ 是一个符合均匀分布的随机合法 d-染色.