



上海科技大学
ShanghaiTech University

SI140 Probability & Mathematical Statistics

Homework 1

陈昱聪

Chen Yucong ><E<>N

Student ID: 2019533079

Email: chenyc@shanghaitech.edu.cn

SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY

© Group#2 (TA: 曾理)

Problem 1.

Answer:

我在大一上学期学习了数学分析 I 以及线性代数 I, 在大一下学期学习了数学分析 II。我顺利地完成了这些课程, 并拿到了不算非常差的成绩。在此过程中, 我学习到了微积分、矩阵、向量空间等数学基础知识, 并掌握了基本的逻辑推理能力和一定的数学洞察力。经过长时间在数学海洋里的浸泡, 我领悟了许多数学思想, 拥有了基本的数学思维方式。除此之外, 我还学习了离散数学等课程; 在前不久, 我还参加了数学建模竞赛。

作为信息科学的重要数学基础, 这门课程对我而言无疑十分重要。首先我希望能够完全掌握这门课所要求的知识内容, 并进一步提升我的数学水平和逻辑思维能力。除此之外, 我还希望能够实践一些此门课在实际信息科学中的应用, 从而更好地将课堂知识转化为真实的学术能力。最后希望能拿一个好成绩!

Problem 2.

(a) **Answer:**

If T is a linear transformation from a vector space V over a field F into itself, and \mathbf{v} is a nonzero vector in V , then \mathbf{v} is the **Eigenvector** of T iff $T(\mathbf{v})$ is a scalar multiple of \mathbf{v} , that is

$$T(\mathbf{v}) = \lambda \mathbf{v}$$

And finally, λ is the **Eigenvalue** of T , which is the factor by which the eigenvector is scaled.

(b) **Proof:**

We couldn't find eigenvalues if a non-square matrix was given.

If there is a $m \times n$ matrix A , and $m \neq n$. Assume we could find an eigenvalue λ_a , then equivalently we could also find a corresponding eigenvector \mathbf{v}_a . And also \mathbf{v}_a is an n dimensional vector.

So that we have

$$A\mathbf{v}_a = \lambda_a \mathbf{v}_a$$

From the matrix multiplication rules, we know that $A\mathbf{v}_a$ is an m dimensional vector but $\lambda_a \mathbf{v}_a$ is an n dimensional vector. Obviously they can't be equal, so there is a contradiction!

So that we couldn't find eigenvalues if given a non-square matrix.

(c) **Answer:**

Geometrically, We consider eigenvalue as the factor by which the eigenvector is scaled. That is, when applying a linear transformation to a vector space, if there is a vector hasn't changed its direction over the transformation, then its norm may have changed. So we called the scale factor of the norm **Eigenvalue**.

(d) **Answer:**

Theoretically, for a linear transformation represented by a square matrix, it can be reduced to two transformations — Rotation and Stretching. The eigenvalues are used to describe degrees of stretching of the vectors which are not been rotated, it's a good mathematical concept to describe the characteristics of linear transformations.

Practically, we use eigenvalues for judging the similarity of two things that can be shown in matrices, like pictures. Therefore, eigenvalues are an important concept in many fields of information science.

Problem 3.

Solution:

(a)

$$\lim_{a \rightarrow 0^+} a^a = \lim_{a \rightarrow 0^+} e^{\ln a^a} = \lim_{a \rightarrow 0^+} e^{a \ln a} = \lim_{a \rightarrow 0^+} e^{\frac{\ln a}{\frac{1}{a}}} = e^{\lim_{a \rightarrow 0^+} \frac{\ln a}{\frac{1}{a}}}$$

By using L'Hôpital's rule, we have

$$\lim_{a \rightarrow 0^+} \frac{\ln a}{\frac{1}{a}} = \lim_{a \rightarrow 0^+} \frac{\frac{1/x}{-1/x^2}}{-1/x^2} = \lim_{a \rightarrow 0^+} -x = e^0 = 1$$

(b) We can figure out the square of the target integral first.

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-x^2} dx \right)^2 &= \int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy \\ &= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} e^{-x^2+y^2} dx \\ &= \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2} r dr \\ &= 2\pi \int_0^{\infty} e^{-r^2} r dr \\ &= \pi \int_{-\infty}^0 e^s ds \\ &= \pi(e^0 - e^{-\infty}) \\ &= \pi \end{aligned}$$

So we have

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

Problem 4.

Proof:

To prove $\phi(t) \leq \frac{1}{8}t^2$, we can first do some transformations to it:

$$\begin{aligned}\phi(t) &\leq \frac{1}{8}t^2 \\ -\theta t + \ln(1 - \theta + \theta e^t) &\leq \frac{1}{8}t^2 \\ \ln(1 - \theta + \theta e^t) &\leq \frac{1}{8}t^2 + \theta t\end{aligned}\tag{1}$$

Take exponential to Eq.(1), we have

$$1 - \theta + \theta e^t \leq e^{\frac{1}{8}t^2 + \theta t}\tag{2}$$

Take Taylor Expansion to the exponent terms in Eq.(2)

$$\begin{aligned}1 - \theta + \theta(1 + t + \frac{t^2}{2} + \dots) &\leq 1 + \theta t + (\frac{1}{4} + \theta^2)\frac{t^2}{2} + \dots \\ -(\theta - \frac{1}{2})^2 \cdot \frac{t^2}{2} - (\theta^2 - \frac{1}{4})\theta \cdot \frac{t^3}{6} + \dots &\leq 0\end{aligned}\tag{3}$$

By induction, all the coefs are non-positive for $\theta > 0, t > 0$. So Eq.(3) is True. So we have $\phi(t) \leq \frac{1}{8}t^2$.

Problem 5.

Proof:

For $a \in \{1, \dots, k\}$, the soft-max function can be shown as below:

$$\pi(x) = \frac{e^{H(x)}}{e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)}}$$

From the derivative rules,

(1) If $a = x$;

$$\begin{aligned} \frac{\partial \pi(x)}{\partial H(a)} &= \frac{\partial \pi(a)}{\partial H(a)} = \frac{e^{H(a)}(e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)}) - e^{H(a)}e^{H(a)}}{(e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)})^2} \\ &= \frac{e^{H(a)}(e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)}) - e^{H(a)}e^{H(a)}}{(e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)})(e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)})} \\ &= \frac{e^{H(a)}}{e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)}} - \left(\frac{e^{H(a)}}{e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)}} \right)^2 \\ &= \pi(a) - \pi^2(a) \\ &= \pi(x)(1 - \pi(a)) \end{aligned}$$

(2) If $a \neq x$;

$$\begin{aligned} \frac{\partial \pi(x)}{\partial H(a)} &= \frac{-e^{H(x)}e^{H(a)}}{(e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)})^2} \\ &= \frac{-e^{H(x)}}{e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)}} \cdot \frac{e^{H(a)}}{e^{H(1)} + \dots + e^{H(a)} + \dots + e^{H(k)}} \\ &= -\pi(x)\pi(a) \end{aligned}$$

From what has been discussed above, $\frac{\partial \pi(x)}{\partial H(a)} = \pi(x)(1_{\{x=a\}} - \pi(a))$ for any skill level $a \in \{1, \dots, k\}$.