人工智能数学基础复习

不喵半個世紀

日期：December 25, 2024

本文档是基于 2023-2024 秋季学期王子贺老师开设的人工智能数学基础课程所撰写的，旨在帮助后来的同学们复习这门知识范围广、难度高的课程。（不过不用担心，王子贺老师还是非常仁慈的 (^ v ^)，大家认真复习就好）此外，由于课程每年内容的变化和本人对课程记忆的淡忘，文档仅供参考，希望能帮助到大家。

1 微分

本节主要是掌握计算方法，基础性较强，在考题中往往结合其他题目出现。

1.1 一元函数

- 连续　　微分　　导数　　链式法则
- 泰勒级数　　麦克劳林级数

  Q: 解析的函数 f 的定义是什么？

1.2 多元函数

- 极限　　连续
- 偏导数　　全微分　　导数
  - 一元函数：可导 = 可微
  - 多元函数：可导 $\Rightarrow$ 可微
  - 连续必可微

  Q: 求 $f(x,y) = x\ln(x^3 + 4xy^2 + y^3)$ 的全微分
- 方向导数
- 链式法则

  Q: 设 $x = t^2, y = \cos t, f(x,y) = x\ln y + 4xy$，求 $\frac{df}{dt}$
- 梯度

  Q: 设 $x = t\sin s, y = t^2\cos s, f(x,y) = xy + \ln y$，求 $\frac{df}{d(s,t)^T}$

1.3 向量值函数的梯度

- 雅可比矩阵　　分子/分母布局（注意不同向量值函数梯度的维度大小）

- 向量值函数的梯度求解
  Q: 求解 MSE 的闭式解（即求解 $\arg\min_{\omega}||\mathbf{y}-\mathbf{X}\omega||_2^2$）

1.4 矩阵的梯度

- 矩阵关于向量的梯度
- 矩阵关于矩阵的梯度
- 几个重要的梯度
  Q: 设 $f(\mathbf{A})=\mathrm{tr}(\mathbf{A}^{\mathbf{T}}\mathbf{B}\mathbf{A})$，求 $\frac{df}{d\mathbf{A}}$

1.5 神经网络中的梯度

本节在 2023 年只介绍了反向传播与自动微分机制，感兴趣的同学可以了解前馈网络和反向传播的计算方法，但事实上这部分 2023 年没考。

1.6 高阶导数

- 高阶偏导
- 海森矩阵
- 多变量泰勒展开
  Q: 设 $f(x,y)=x^2y+2xy-y^3$，求 f 在 (2, 1) 的泰勒展开

1.7 微分方程

2023 年没讲，感觉不像重点

2 优化

期末重点之一，难度并不算高，考察计算为主，请务必掌握凸函数的定义、拉格朗日乘子法、牛顿迭代法

2.1 凸优化问题

- 凸集保凸运算
- 凸函数（重点）
  - 定义
  - 充要条件
  - 保凸运算

  这部分考了填空和大题各一道，笔者有些淡忘但下面的练习是大概类似的考点，难度也较为相似：
  Q1: $f_1(x,y)=x^2y-2xy+y^3$, $f_2(x,y)=xe^y$, $f_3(x,y)=x\ln(x-y)$, $f_4(x,y)=x^2-2xy+y^2$，上面函数中，凸函数是 _，凹函数是 _，非凸非凹函数是 _。

Q2: 假设 $f(x)$ 是凸函数，且 $g(x)$ 单调递增。证明 $h(x)=f(g(x))$ 仍是凸函数。

Q3: 假设 $f_1(x),f_2(x),\dots,f_m(x)$ 是凸函数，权重 $w_i>0$，且满足 $\sum_{i=1}^m w_i=1$。证明函数 $g(x)=\sum_{i=1}^m w_i f_i(x)$ 是凸函数。

- 下水平集上镜图

  虽然感觉不像是重点，但去年期末填空考了上镜图的定义，大家还是尽量记住定义，笔者虽然没记住但是考试绞尽脑汁写了个差不多的意思的，遂答对
- 优化问题凸优化问题
  - 定义
  - 局部最优 = 全局最优
  - 最优解充要条件

2.2 Lagrange 对偶函数

- Lagrange 函数
- Lagrange 对偶函数

  Q: 设问题为 $\min x_1^2-x_1x_2+x^2$, 有约束条件 $x_1^2+x_2^2\le 1$，$x_1+x_2=1$，求 Lagrange 函数和对偶函数。

2.3 对偶问题

- 原问题与对偶问题
- 弱对偶性与强对偶性

2.4 最优性条件

- 互补松弛性
- KTT 条件（重点）

  Q: 设问题为 $\min x_1^2+2x_2^2+x_3^2$, 有约束条件 $x_1^2+x_2^2+x_3^2\le 1$，$x_1+x_2+x_3=1$，求其最优解和最优值。

2.5 经典凸优化问题

试题里面一般都考凸优化, 特别是线性规划和二次规划, 要尽可能掌握凸优化的 lagrange 乘子法。但本节不重要

2.6 凸优化算法

- 无约束优化　梯度下降　回溯直线搜索
- 牛顿法（重点）

  去年考察了一道大题，下面是一道类似的题目

  Q: 设 $f(x_1,x_2)=x_1^2-2x_1x_2^2+3x_2^2$，请回答下面的问题：(1) 求 $f(x)$ 的海森矩阵，并说明

其极值的情况 (2) 求 $f(x)$ 在 $(1,2)$ 处的泰勒展开 (3) 以 $\mathbf{x}_0=(1,2)$ 为初始点，计算牛顿法迭代两次的结果。

- 其他方法

3 概率

期末考试的重难点，也是复习的大头，需着重复习

3.1 单变量

- 期望　方差
- 交叉熵　KL 散度　互信息

  去年似乎也考察了这部分内容，难度不高，需要注意的是与熵相关的题目

  Q: 给定联合概率分布 $p(x,y)$ 如下：

$X\backslash Y$	0	1
0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{3}{8}$

  其中 $p(x,y)$ 表示 X 和 Y 的联合分布。

  请计算以下四个量：

  (a) $H(X), H(Y)$

  (b) $H(X|Y), H(Y|X)$

  (c) $H(X,Y)$

  (d) $I(X;Y)$

3.2 多变量

基本与概统类似

- 联合分布　联合密度函数
- 边际分布　独立　条件期望
- 协方差　相关系数

3.3 高斯分布

- 多元正态分布

  填空考察了给定正态分布的概率密度函数，求均值和方差

  Q: 给定多元正态分布

$$f(\mathbf{x})=\frac{1}{2\pi\sqrt{5}}\exp\left(-\frac{1}{2}\left[\frac{3}{5}(x_1-3)^2+\frac{6}{5}(x_1-3)(x_2-5)+\frac{3}{5}(x_2-5)^2\right]\right).$$

  其均值和方差分别为?

- 变量组合

  若 $\mathbf{x}\sim\mathcal{N}(\mu_{\mathbf{x}},\mathbf{\Sigma}_{\mathbf{x}}), \mathbf{y}\sim\mathcal{N}(\mu_{\mathbf{y}},\mathbf{\Sigma}_{\mathbf{y}})$, 则 $a\mathbf{x}+b\mathbf{y}\sim\mathcal{N}(a\mu_{\mathbf{x}}+b\mu_{\mathbf{y}},a^2\mathbf{\Sigma}_{\mathbf{x}}+b^2\mathbf{\Sigma}_{\mathbf{y}})$

若 $\mathbf{x} \sim \mathcal{N}(\mu_{\mathbf{x}}, \mathbf{\Sigma_x}), \mathbf{y} \sim \mathcal{N}(\mu_{\mathbf{y}}, \mathbf{\Sigma_y})$, 则 $\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mu_{\mathbf{x}|\mathbf{y}}, \mathbf{\Sigma_{x|y}})$

若 $\mathbf{x} \sim \mathcal{N}(\mu_{\mathbf{x}}, \mathbf{\Sigma_x}), \mathbf{y} = \mathbf{A}\mathbf{x}$, 则 $\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mu_{\mathbf{x}}, \mathbf{A\Sigma_x A^T})$

但是需注意 $\alpha p(x) + (1-\alpha)p(y)$ 不服从正态分布

- 协方差　相关系数

3.4 统计推断

考试中相关部分似乎涉及不多，好像考了简单的 MLE；但是是下学期机器学习的基本知识

- 最大似然

  参考概率论与数理统计

- 贝叶斯估计

  Q: 设 $y_1, y_2, \ldots, y_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其中 σ^2 已知，μ 未知。假设 μ 的先验分布为均匀分布 Uniform(a, b)，即 μ 在区间 $[a, b]$ 内均匀分布，且 a 和 b 已知。试求 μ 的贝叶斯估计。

3.5 共轭分布

去年没考，下学期的机器学习会部分涉及到

- 共轭分布

  若后验与先验分布相同，则先验与似然共轭分布

- 指数族分布充分统计量

  指数族分布都有共轭分布

3.6 概率不等式

- 马尔可夫不等式　切比雪夫不等式

  概统的重点，但去年只涉及了简单的马尔可夫不等式

- 切尔诺夫界　霍夫丁界

  没考

3.7 高维数据

3.7.1 几何特性

- 单位球　表面积　体积
- 体积分布

  定理：对于 $c \geq 1, d \leq 3$, 至少 $1 - \frac{2}{c}e^{-\frac{c^2}{2}}$ 比例的体积有 $x_1 \leq \frac{c}{\sqrt{d-1}}$

- 随机采样

  近似正交性：从 d 维单位球中均匀采样 n 个点 $x_1, \ldots, x_n$, 以至少 $1O(\frac{1}{n})$ 的概率满足

$$|x_i \cdot x_j| \leq \frac{\sqrt{6\ln n}}{d-1}, \text{for } i, j \in [1, n]$$

3.7.2 概率特性

- 均匀分布

  23 年拔尖班必做题：如何从一元标准正态分布得到 d 维单位球面的标准正态分布

  Q： 当时考试的部分同学从极坐标的角度进行了回答，请问这样做有什么问题？

- 高斯分布

  高斯环定理：在 d 维高斯空间中，每一维的方差都是 1，对于 $\beta \leq \sqrt{d}$, 至多有 $3e^{-c\beta^2}$ 的概率分布于环 $\sqrt{d} - \beta \leq |x| \leq \sqrt{d} + \beta$ 之外。

  Q： $E(|x|) =?$

3.8 数据降维

- 随机投影定理
- JL 算法

  23 年第一道大题：简述 JL 算法

3.9 采样

- 基本采样算法
- 拒绝采样　自适应拒绝采样
- 重要性采样
- MCMC

  似乎是第二道大题，第一问简述 MCMC 算法的流程，第二问计算马尔科夫链的稳态分布，与下面的题目类似：

  Q： 假设我们有两个离散随机变量 X_1 和 X_2，它们的状态空间分别为 $X_1 \in \{1, 2, 3\}$ 和 $X_2 \in \{1, 2\}$。其中，状态转移概率分别为 $\begin{pmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{pmatrix}$ 和 $\begin{pmatrix} 0.1 & 0.7 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}$

  (1) 假设我们使用 Metropolis-Hastings 算法进行采样，请简述该算法的基本流程。

  (2) 计算该马尔科夫链的稳态分布 π。

- 高斯分布采样

  逆变换法　Box-Muller　Marsaglia polar

  没考但感觉也是重点

4 算法

该部分 2023 年未教授。