# 模式识别与计算机视觉 HW1

**姓名**：石睿

**学号**：211300024

**所属院系**：人工智能学院

习题 1.

(a) ① 类型约束：a 是 int() 或 float()

( type(a) == int ) or ( type(a) == float )

② 数值约束：根号有效

$a \geq \frac{1}{8}$

(b) 1

(c) 令 $b = \frac{a+1}{3}\sqrt{\frac{8a-1}{3}}$

原式 $= \sqrt[3]{a+b} + \sqrt[3]{a-b}$

尝试令 $a=b$，也即 $a = \frac{a+1}{3}\sqrt{\frac{8a-1}{3}}$

化简得 $8a^3 - 12a^2 + 6a - 1 = a^2 \cdot (8a-4) - a(8a-4) + (2a-1)$

$= (2a-1)\cdot(4a^2-4a+1)$

$= (2a-1)\cdot(2a-1)^2$

$= (2a-1)^3 = 0$

$\therefore$ 令 $a=\frac{1}{2}$，原式 $=1$.

(d) $1.2182 + 0.1260i$

(e) MATLAB计算方根时总是返回辐角最小的那个根.

代码需修改成 f= (a+ (a+1)/3 * sqrt((8*a-1)/3))^(1/3) + ⋯

− (abs(a−(a+1)/3 * sqrt((8*a-1)/3)))^(1/3)

$\Rightarrow$ 尝试后（对 $\forall a \geq \frac{1}{8}$）函数值为 1

(f)

结论（断言）

问题描述 $\sqrt[3]{a + \frac{a+1}{3}\cdot\sqrt{\frac{8a-1}{3}}} + \sqrt[3]{a - \frac{a+1}{3}\cdot\sqrt{\frac{8a-1}{3}}} = 1, \forall a \geq \frac{1}{8}$

证明：

符号描述 令 $A = \sqrt[3]{a + \frac{a+1}{3}\cdot\sqrt{\frac{8a-1}{3}}} \in \mathbb{R}$.

$B = \sqrt[3]{a - \frac{a+1}{3}\cdot\sqrt{\frac{8a-1}{8}}} \in \mathbb{R}$.

↓

[观察] $A^3 + B^3 = 2a^3$

[观察] $A^3 \cdot B^3 = a^2 - \frac{(a+1)^2\cdot(8a-1)}{27} = \left(\frac{1-2a}{3}\right)^3$

$\because A, B \in \mathbb{R}$

$\therefore A \cdot B = \frac{1-2a}{3}$.

$A^3 + B^3 = (A+B)\cdot(A^2 - AB + B^2)$

$= (A+B)\cdot((A+B)^2 - 3AB)$

$= (A+B)\cdot((A+B)^2 + (2a-1)) = 2a$.

问题

原问题即证 $A+B=1$

不妨令 $A+B = C \in \mathbb{R}$，原问题即证 $C=1$ 问题

有 $C\cdot(C^2 + (2a-1)) = 2a$.

即 $C^3 + (2a-1)\cdot C - 2a$

$= (C-1)\cdot(C^2 - C + 2a) = 0$

$\therefore C=1$ 或 $C^2 - C + 2a = 0$

由于 $\Delta = 1 - 8a < 0$

$\therefore C^2 - C + 2a > 0$ 恒成立

$\therefore C = 1$

也即原问题得证

(g) 令 $a=2$，（满足 $a\ge\frac{1}{8}$）

$$\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}}=\sqrt[3]{a+\frac{a+1}{3}\cdot\sqrt{\frac{8a-1}{3}}}+\sqrt[3]{a-\frac{a+1}{3}\cdot\sqrt{\frac{8a-1}{3}}}\equiv 1$$

(h)

由 Cardano 公式可知，式(1.1)是方程 $x^3+(2a-1)x-2a=0$ 的第一个复根（a 变化时始终为实数）

也即在(f)证明中给出的 $c^3+(2a-1)\cdot c-2a=0$ 有唯一闭式解 $c=1$.

故式(1.1)的值在定义域内恒为1

---

习题 2

(a) 换元，积分下限变为0

$$P(X\ge\varepsilon)=\int_{\varepsilon}^{+\infty}p(x)\,dx=\int_{\varepsilon}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}dx\overset{t=x-\varepsilon}{=\!=}\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{(t+\varepsilon)^2}{2}}d(t+\varepsilon)$$

$$=\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{\left(-\frac{t^2+2t\varepsilon+\varepsilon^2}{2}\right)}d(t+\varepsilon)=\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{\varepsilon^2}{2}}\cdot e^{-t\varepsilon}\cdot e^{-\frac{t^2}{2}}dt$$

$$\le e^{-\frac{\varepsilon^2}{2}}\cdot\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}dt=\frac{1}{2}\cdot e^{-\frac{\varepsilon^2}{2}}$$

(b)

$$P(|X|\ge\varepsilon)=2\cdot P(X\ge\varepsilon)\neq 2\cdot\int_{\varepsilon}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}dx.$$

法一 ✗

$$\overset{t=x-\varepsilon}{=\!=}2\cdot\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{(t+\varepsilon)^2}{2}}d(t+\varepsilon)$$

$$=2\cdot\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}\cdot e^{-t\varepsilon}\cdot e^{-\frac{\varepsilon^2}{2}}dt.$$

证明选择的失误导致的回溯

同 ATP 证明

法二 用 hint ✓

$\boxed{1}$ $$P(|X|\ge\varepsilon)=2\cdot P(X\ge\varepsilon)\neq 2\cdot\int_{\varepsilon}^{+\infty}f(x)\,dx=2\int_{\varepsilon}^{+\infty}-\frac{f'(x)}{x}dx.\le 2\cdot\int_{\varepsilon}^{+\infty}-\frac{1}{\varepsilon}\cdot f'(x)\,dx.$$

$$=-\frac{2}{\varepsilon}\int_{\varepsilon}^{+\infty}f'(x)\,dx.$$

$$=-\frac{2}{\varepsilon}f(x)\Big|_{\varepsilon}^{+\infty}=\frac{2}{\varepsilon}\cdot f(\varepsilon)=\frac{2}{\varepsilon}\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{\varepsilon^2}{2}}=\sqrt{\frac{2}{\pi}}\cdot\frac{e^{-\frac{\varepsilon^2}{2}}}{\varepsilon}$$

$\boxed{2}$ $$P(|X|\ge\varepsilon)\le P(|X|\ge 0)=\int_{-\infty}^{+\infty}f(x)\,dx=1$$

综上：$$P(|X|\ge\varepsilon)\le\min\left\{1,\sqrt{\frac{2}{\pi}}\cdot\frac{e^{-\frac{\varepsilon^2}{2}}}{\varepsilon}\right\}$$

习题 3

本题中的问题描述如下

输入 $x \in \mathbb{R}^d$，两个FC层的权重矩阵和偏置向量分别为 $W_1 \in \mathbb{R}^{m\times d}$，$b_1 \in \mathbb{R}^m$
$W_2 \in \mathbb{R}^{n\times m}$，$b_2 \in \mathbb{R}^n$.

$\Gamma$ 和 $\beta$ 是批正则化层(BN)中的缩放因子和偏移量

$\varepsilon > 0$ 是个很小的常数.

(a) 此时假设FC中没有BN，也没有激活函数层.（有不同的定义方式）

FC层：$y = WX + b$. FC层通过权重矩阵 $W$ 和偏置向量 $b$ 的线性变换（仿射变换）学习输入、输出间映射关系。

BN层：
$$\begin{cases} \mu = \frac{1}{n}\sum_{j=1}^{n} x_j & \text{均值} \\ \sigma = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2} & \text{标准差} \end{cases}$$

$$y = \frac{x-\mu}{\sqrt{\sigma}+\varepsilon}\cdot\Gamma + \beta$$

BN层对一个批次数据 $X = \{x_1, x_2, \dots, x_n\}$ 计算均值和方差，对X进行归一化处理，使 $X \sim N(0,1)$。把X的分布拉回激活函数的敏感区域，使网络中间层输入稳定，减轻梯度消失或爆炸的问题

(b) 证明：

$\because y_1 = W_1 X + b_1$

$\therefore y_2 = W_2 y_1 + b_2 = W_2(W_1 X + b_1) + b_2 = (W_2 W_1)\cdot X + (W_2 b_1 + b_2)$

故两层相连可视作一个以 $W_2W_1$ 为权重矩阵，$W_2 b_1 + b_2$ 为偏置向量的新FC层.

(c) 在较简单任务或计算资源有限时.

① 两FC层直接相连，在保证参数量不减小时减化了网络结构（中间不加入BN/激活层）可提高训练速度

② 无中间的BN，激活层，在无需复杂的特征变换时，避免了冗余计算

③ 无BN.激活层，反向传播更易计算，且避免且激活函数引起的梯度消失，爆炸

(d) 否. 在两层的优化方式不同时，会损失灵活性和性能

d.1 第一个FC层做降维，第二个FC层做分类。合并后无法拿到第一个FC的结果 无法评估维度选择是否正确（是否线性可分）

d.2 $FC_1$ 和 $FC_2$ 采用不同的正则化/初始化方式，合并后失去灵活性

(e) 证明：

$\because y_1 = W_1 X_1 + b$

$$\therefore y_2 = \frac{y_1-\mu}{\sqrt{\sigma}+\varepsilon}\cdot\Gamma + \beta = \underbrace{\left(\frac{\Gamma}{\sqrt{\sigma}+\varepsilon}\cdot W\right)}_{\text{新权重矩阵 }\tilde{W}}\cdot y_1 + \underbrace{\left(\frac{\Gamma\cdot(b-\mu)}{\sqrt{\sigma}+\varepsilon}+\beta\right)}_{\text{新偏置向量 }\tilde{b}}$$

，其中 $\begin{cases} \mu = \frac{1}{n}\sum_{j=1}^{n} y_1 \\ \sigma = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2} \end{cases}$

情况：在训练后.推理时.

此时BN中 $\mu$.$\sigma$ 固定（全局均值，方差），不再使用 mini-batch 中变化的统计数据.

↑在训练时一定不可以 $\mu$.$\sigma$ 影响 $\partial(F)$，必须以单独的单元存在，才可在反向传播中正确计算 $\partial F$

进而此时BN也即对每个维度做线性变换了

好处：① 无需求 $\mu$.$\sigma$，减少运算次数&访存次数

② 数值更稳定，避免因 $\mu$.$\sigma$.缩放而带来的数值不稳定问题

习题四.

(a) (1) maxpooling 中选择 4x4 的池化窗口 / 从 4x4 中选择一个代表元素

(2) 降维方法 如 PCA，从 400x400 → 100x100

(b). 每个 2x2 的区域中选一个元素 / 平均数 / 中位数 (依情况而定)

100x100 → 50x50，存储开销降低了 75%

(c)

$acc_{train} = \frac{9900}{10000} = 99\% = 0.99$

$acc_{test} = \frac{5000}{10000} = 50\% = 0.5$

(d)

$macro\_P = \frac{1}{n} \cdot \sum_{i=1}^{n} P_i$ …每个类别的准确率求平均

$micro\_P = \frac{\sum_{i=1}^{n} l_i}{\sum_{i=1}^{n} m_i}$

$l_i$：预&真属 i 类

$m_i$：预属 i 类

c) 中是 micro_P

P ⋯ 预

| 真 \ 预 | 正 | 负 |
|---|---|---|
| 正 | TP | FN |
| 负 | FP | TN |

⋯ R

$P = \frac{TP}{TP+FP}$

$R = \frac{TP}{TP+FN}$

(e). 应当采用 macro_P

由于 $|D_A| >> |D_B|$，从训练集中随机选 A 类数据（不全取），和 B 类全部样本 ⇒ 训练集

交叉验证.

$D_{A1} + D_B = D_{train1} \to model 1$

⋮

$D_{An} + D_B = D_{train n} \to model\ n$

→ $D_{test}$ 看误差

选择 $D_{test}$ 上 loss 最小的做 $D_{Ai}$ 的数量 $|D_{Ai}|$ 做为未来取 $D_{Ai}$ 的标准

再多取几个 $D_{Ai} + D_B$ 做 $D_{traini}$ 训练 model i，最后 model 1, …, model n 投票决策

习题五.

a) $z_1$ 分为 A

$z_2$ 分为 A

b) $z_1$ 分为 A

$z_2$ 分为 B

c) 离 $z_1$ 最近的 1 个或 3 个“训练”样本全是 A 类

离 $z_2$ 最近的 1 个是 B 类，但最近的 3 个是 A.B.B 类

d) 可能，$x_7$ 可能是离群点

K-NN 采噪声影响小，更稳定，可在一定程度避免由于采样误差产生的错误决策

**习题六：感想**

1. **问题**：Windows 系统规定所有文件必须具有扩展名，而 Linux 则不是，因此我所下载的数据集作为 txt 文件而不能直接被字符串‘svmguide1’ 匹配。故我在框架代码中修改了 dataset 的命名（加入扩展）。
2. **感想 1**：应该复习 numpy 中的库函数精简编程的代码量，且提高运行效率。
3. **感想 2**：本次编程的框架代码具有一定的一般性，包括了数据的导入（及预处理）、模型的评估（通过准确率或是查全率、查准率、F1 度量）与选择（通过交叉验证）、模型的测试等，整体上具有一个机器学习系统所应当具备的部件。之后遇到现实模型的设计，可以借鉴这个很好的范本哇！