模式识别与计算机视觉 **人工智能学院**

Homework 1

Instructor: 吴建鑫 *Name:* 方盛俊, *StudentId:* 201300035

1. 习题一

我们不妨设公式:

$$f(a)=\sqrt[3]{a+\frac{a+1}{3}\sqrt{\frac{8a-1}{3}}}+\sqrt[3]{a-\frac{a+1}{3}\sqrt{\frac{8a-1}{3}}} \tag{1.1}$$

(a)

由于只考虑实数的情况, 虚部为零, 因此有

$$\frac{8a-1}{3}\geq 0$$

则我们可推出对输入的要求:

$$a\geq\frac{1}{8} \tag{1.2}$$

(b)

当 $a=\frac{1}{8}$ 时, 带入式 (1.1) 有

$$f(\frac{1}{8})=\sqrt[3]{\frac{1}{8}}+\sqrt[3]{\frac{1}{8}}=1 \tag{1.3}$$

(c)

我们带入方便计算的特殊样例 $a=\frac{1}{2}$ 可得

$$f(\frac{1}{2})=\sqrt[3]{\frac{1}{2}+\frac{1}{2}}+\sqrt[3]{\frac{1}{2}-\frac{1}{2}}=1$$

同理带入方便计算的特殊样例 $a=\frac{13}{8}$ 可得

$$f(\frac{13}{8})=\frac{3}{2}+\frac{\sqrt[3]{-1}}{2}=1$$

我们发现两者结果均为 1.

(d)

这条命令的返回值为 $1.2182+0.1260i$.

(e)

由于 $(\cdot)^{1/3}$ 在 MATLAB 中等价于 $\mathrm{power}(\cdot, 1/3)$, 该函数是在复数域内计算, 最终计算结果的误差会累计增大, 得到一个错误的结果, 我们应该使用在实数域计算的函数 $\mathrm{nthroot}(\cdot, n)$, 即使用

```
a = 3 / 4
f = nthroot(a + (a+1)/3 * sqrt((8*a-1)/3), 3) + ...
    nthroot(a - (a+1)/3 * sqrt((8*a-1)/3), 3)
```

可以算出结果为 1.0.

给 $a \geq 0.125$ 带入不同的值, 依然等于这个结果.

(f)

由于 $a \geq \frac{1}{8}$, 我们不妨令 $a = \frac{3x^2}{8} + \frac{1}{8}$, 其中 $x \geq 0$, 则有

$$\begin{aligned}
f(\frac{3x^2}{8} + \frac{1}{8}) &= \sqrt[3]{a + \frac{a+1}{3}\sqrt{\frac{8a-1}{3}}} + \sqrt[3]{a - \frac{a+1}{3}\sqrt{\frac{8a-1}{3}}} \\
&= \frac{\sqrt[3]{3x^2 - (x^2+3)\sqrt{x^2}+1}}{2} + \frac{\sqrt[3]{3x^2 + (x^2+3)\sqrt{x^2}+1}}{2} \\
&= \frac{\sqrt[3]{-(x-1)^3}}{2} + \frac{\sqrt[3]{(x+1)^3}}{2} \\
&= \frac{1-x}{2} + \frac{1+x}{2} \\
&= 1
\end{aligned}$$

可见当 $a \geq \frac{1}{8}$ 时有 $f(a) = 1$.

(g)

令 $a = 2$ 则有

$$\begin{aligned}
f(2) &= \sqrt[3]{2 + \frac{2+1}{3}\sqrt{\frac{16-1}{3}}} + \sqrt[3]{2 - \frac{2+1}{3}\sqrt{\frac{16-1}{3}}} \\
&= \sqrt[3]{2+\sqrt{5}} + \sqrt[3]{2-\sqrt{5}} \\
&= 1
\end{aligned}$$

即该表达式结果为 1.

(h)

查阅资料后, 得知 Cardano 证明了三次方程

$$z^3 + pz + q = 0$$

其中 p, q 是实数, 且 $\Delta = \frac{q^2}{4} + \frac{p^3}{27} > 0$ 时, 方程有实根

$$\sqrt[3]{-\frac{q}{2}+\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}$$

因此我们推测式子

$$f(a)=\sqrt[3]{a+\frac{a+1}{3}\sqrt{\frac{8a-1}{3}}}+\sqrt[3]{a-\frac{a+1}{3}\sqrt{\frac{8a-1}{3}}}$$

是某个三次方程的根.

由于 $f(a)=1$ 在 $a>0.125$ 时恒成立, 我们可以猜测存在该三次方程存在一个根 $z=1$.

使用待定系数法, 可得

$$(z-1)(z^2+bz+c)=z^3+(b-1)z^2+(c-b)z-c$$

令 $b-1=0, -c=q$ 可得

$$(z-1)(z^2+z-q)=z^3+(-q-1)z+q$$

再观察求根公式与 $f(a)$ 的差异, 我们可以令 $a=-\dfrac{q}{2}$, 则有

$$\begin{cases} p=2a-1 \\ q=-2a \end{cases}$$

则我们可以知道, $f(a)$ 是三次方程

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

的一个根, 在 $a>0.125$ 时恒等于 1.

并且经过检验, 该结论成立.

2. 习题二

(a)

由于 $X\sim\mathcal{N}(0,1)$, 我们有

$$\begin{aligned} P(X\geq\epsilon) &= \int_{\epsilon}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\mathrm{d}t \\ &= \int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-(x+\epsilon)^2/2}\mathrm{d}x \\ &\leq e^{-\epsilon^2/2}\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\mathrm{d}x \\ &= \frac{1}{2}e^{-\epsilon^2/2} \end{aligned}$$

(b)

由于 X 的概率密度函数为 $f(x)=\dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$, 因此求导有 $f'(x)=-xf(x)$, 则有

$$
\begin{aligned}
P(|X| \geq \epsilon) &= 2\int_{\epsilon}^{+\infty} f(x)\mathrm{d}x = 2\int_{\epsilon}^{+\infty} \frac{xf(x)}{x}\mathrm{d}x \\
&\leq 2\int_{\epsilon}^{+\infty} \frac{xf(x)}{\epsilon}\mathrm{d}x = -2\int_{\epsilon}^{+\infty} \frac{f'(x)}{\epsilon}\mathrm{d}x \\
&= -\frac{2}{\epsilon}[f(x)]_{\epsilon}^{+\infty} \\
&= \sqrt{\frac{2}{\pi}}\frac{e^{-\epsilon^2/2}}{\epsilon}
\end{aligned}
$$

因此我们有

$$
P(|X| \geq \epsilon) \leq \min\{1, \sqrt{\frac{2}{\pi}}\frac{e^{-\epsilon^2/2}}{\epsilon}\}
$$

3. 习题三

(a)

带入 $x = 0$ 则有

$$
f^*(0) = \sup_{x \in \mathrm{dom}\, f}(y^{\mathrm{T}}x - f(x)) = \sup_{x \in \mathrm{dom}\, f} -f(x)
$$

两边取负号则有

$$
\inf_x f(x) = -f^*(0)
$$

(b)

当 $x \notin \mathrm{dom}(f)$ 时, 由于 $f(x) = \infty$,

$$
f(x) + f^*(y) = \infty \geq x^{\mathrm{T}}y
$$

恒成立.

当 $x \in \mathrm{dom}(f)$ 时,

要证

$$
f(x) + f^*(y) \geq x^{\mathrm{T}}y
$$

即证

$$
f^*(y) \geq y^{\mathrm{T}}x - f(x)
$$

我们知道

$$
f^*(y) = \sup_{x \in \mathrm{dom}\, f}(y^{\mathrm{T}}x - f(x)) \geq y^{\mathrm{T}}x - f(x)
$$

因此原式成立.

(c)

证明如下:

$$
\begin{aligned}
f^{**}(x) &= \sup_y (y^{\mathrm{T}} x - f^*(y)) \\
&= \sup_y (y^{\mathrm{T}} x - \sup_x (y^{\mathrm{T}} x - f(x))) \\
&= \sup_y (y^{\mathrm{T}} x + \inf_{\hat{x}} (f(\hat{x}) - y^{\mathrm{T}} \hat{x})) \\
&= \sup_y \inf_{\hat{x}} (f(\hat{x}) + y^{\mathrm{T}} (x - \hat{x})) \\
&= \inf_{\hat{x}} \sup_y (f(\hat{x}) + y^{\mathrm{T}} (x - \hat{x})) \\
&= \inf_{\hat{x}} (f(\hat{x}) + \sup_y y^{\mathrm{T}} (x - \hat{x})) \\
&\leq f(x) + \sup_y y^{\mathrm{T}} (x - x) \\
&= f(x)
\end{aligned}
$$

或者使用 (b) 中的结论 $f(x) + f^*(y) \geq x^{\mathrm{T}} y$,

则有对任意 x, y 有

$$
y^{\mathrm{T}} x - f^*(y) \leq f(x)
$$

因此有

$$
\begin{aligned}
f^{**}(x) &= \sup_y (y^{\mathrm{T}} x - f^*(y)) \\
&= y^{*T} x - f^*(y^*) \\
&\leq f(x)
\end{aligned}
$$

其中 y^* 是使得 $(y^{\mathrm{T}} x - f^*(y))$ 取得其中一个上界的 y 值.

4. 习题四

(a)

1. 最近邻插值: 将拍摄图像中的 $(4i+1, 4j+1)$ 的像素点 $f(4i+1, 4j+1)$ 像素值作为最近邻插值, 插值成为存储图像的 (i, j) 像素点的像素值.
2. 双线性插值: 将拍摄图像中的均值 $[f(4i+1, 4j+1) + f(4i+2, 4j+1) + f(4i+1, 4j+2) + f(4i+2, 4j+2)]/4$ 的像素值作为双线性插值, 插值成为存储图像的 (i, j) 像素点的像素值.
3. 均值插值: 将拍摄图像中的 4×4 像素点, 类似双线性插值一般取取均值, 插值成为存储图像的 (i, j) 像素点的像素值.

(b)

将每 2×2 像素格取均值进行插值, 变为一个 1×1 的像素点. 存储开销能降为原来的 25%.

(c)

在训练集上的准确率 $acc_{train} = \dfrac{9900 + 0}{9900 + 100} \times 100\% = 99\%$

在测试集上的准确率 $acc_{test} = \dfrac{5000 + 0}{5000 + 5000} \times 100\% = 50\%$

(d)

对于在 n 个二分类混淆矩阵上综合考察查准率, 查全率以及准确率等指标的情况, 我们有两种不同的方法.

第一种是 micro 方法, 将自身类作为正类，其他所有类作为反类, 先计算每一类正例反例的样本数, 其中 $i \in \{A, B\}$, 得到 TP_i, FP_i, TN_i, FN_i, 则 micro 准确率为:

$$\text{micro}-Acc = \frac{\sum_i TP_i}{\text{total examples}}$$

第二种是 macro 方法, 先对各类别求出准确率, 得到 Acc_i, 再取平均值计算出 macro 准确率:

$$\text{macro}-Acc = \frac{1}{N}\sum_i \frac{TP_i}{TP_i + FP_i}$$

按照 micro 方法, (c) 中训练集的准确率结果为 $\dfrac{9900+0}{9900+100} \times 100\% = 99\%$

按照 macro 方法, (c) 中训练集的准确率结果为 $\dfrac{1}{2} \times (\dfrac{9900}{9900} + \dfrac{0}{100}) \times 100\% = 50\%$

因此我们在 (c) 中采用的是 micro 方法.

或者我们通过 F1 来分析, 其中 micro-F1 为先对混淆矩阵的对应元素进行平均, 再进行计算:

$$\text{micro}-P = \frac{\overline{TP}}{\overline{TP} + \overline{FP}}$$

$$\text{micro}-R = \frac{\overline{TP}}{\overline{TP} + \overline{FN}}$$

$$\text{micro}-F1 = \frac{2 \times \text{micro}-P \times \text{micro}-R}{\text{micro}-P + \text{micro}-R}$$

其中 macro-F1 为先在各个混淆矩阵上算出查准率和查全率, 再算平均值:

$$\text{macro}-P = \frac{1}{N}\sum_{i=1}^{N} P_i$$

$$\text{macro}-R = \frac{1}{N}\sum_{i=1}^{N} R_i$$

$$\text{macro}-F1 = \frac{2 \times \text{macro}-P \times \text{macro}-R}{\text{macro}-P + \text{macro}-R}$$

经过计算我们可以算出, 在多分类问题下, 准确率 (accuracy), 查准率 (precision), 查全率 (recall) 以及 F1 的值都是相同的.

这也可以说明我们在 (c) 中采用的是 micro 方法.

(e)

我们应该采用 macro 方法来评估准确率. 因为 macro-F1 是计算每一类的 F1 score, 然后再求算术平均, 如果模型在小样本上表现不好, 小样本的 F1 会极大程度上拉低 macro-F1, 这样就能对长尾识别问题中类别不平衡问题进行一定的改善.

并且我们知道, 按照 macro 方法, (c) 中训练集的准确率结果为

$$\frac{1}{2} \times (\frac{9900}{9900} + \frac{0}{100}) \times 100\% = 50\%$$

可以看出是通过对各个类别的准确率都赋予了相同的权重, 避免了类别不平衡导致的问题.

为了长尾识别问题中的类别不平衡问题, 我们可以采用以下方法:

1. **重采样**: 对样本少的类别进行又放回的随机采样, 并加入训练集中. 例如此时 A 类有 9900 个样本, B 类有 100 个样本, 我们就可以随机在 B 类上重采样 9800 个样本, 来平衡不同类别的样例.
2. **欠采样**: 在样本多的类别中取出样本少的类别数目一样的样本用于训练.
3. **代价敏感矩阵**: 给不同的类别的样本赋予不同的权重, 以增加样本少的类别对结果的影响.

5. 习题五

(a)

$z_1 = (0, -2)$ 的最近邻分类结果为 $x_3 = (0, -1)$ 对应的类别 A.

$z_2 = (8, 2)$ 的最近邻分类结果为 $x_7 = (8, 1)$ 对应的类别 A.

(b)

$z_1 = (0, -2)$ 的 k-近邻分类结果为 k-近邻 x_1, x_3, x_4 投票得到的类别 A.

$z_2 = (8, 2)$ 的 k-近邻分类结果为 k-近邻 x_6, x_7, x_8 投票得到的类别 B.

(c)

z_1 附近都是类别 A 的样本, 因此仍然是分类为类别 A 不变, 但是 z_2 附近只是偶然有一个类别 A 的样本 x_7, 但是还有更多的类别为 B 的临近样本 x_6, x_8, 因此被分类成类别 B.

(d)

x_7 可能属于类别 B, 可能是在采集数据的时候数据不小心打错了标签. 因此 k-NN 相比于 1-NN 的一个很大的优势就是容错率高, 不容易被偶然的错误样本影响到分类结果.