神经网络和深度学习（类脑）Week 3 作业

姓名: 王冰怡　　学号:21210850015

2021 年 10 月 18 日

题目 1.　一、某地成年人肥胖者 (A_1) 占 10%, 中等者 (A_2) 占 82%, 瘦小者 (A_3) 占 8%, 又肥胖者, 中等者, 瘦小者高血压病的概率分别为 20%,10%,5%.
(1). 求该地成年人患高血压的概率.
(2). 若知某人患高血压，他最可能属于哪种体型?

解答.　设事件 A_1, A_2, A_3 分别某地成年人为肥胖者、中等者、瘦小者；为设事件 B 为某地成年人患有高血压病。
则 $P(A_1) = 0.1, P(A_2) = 0.82, P(A_1) = 0.08, P(B|A_1) = 0.2, P(B|A_2) = 0.1, P(B|A_3) = 0.05$
(1). 由全概率公式

$$\begin{aligned} P(B) &= \sum_{i=1}^{3} P(A_i)P(B|A_i) \\ &= 0.1 * 0.2 + 0.82 * 0.1 + 0.08 * 0.05 \\ &= 10.6\% \end{aligned}$$

(2). 由贝叶斯公式

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{3} P(A_i)P(B|A_i)} \approx 18.87\%$$

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{\sum_{i=1}^{3} P(A_i)P(B|A_i)} \approx 77.36\%$$

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{\sum_{i=1}^{3} P(A_i)P(B|A_i)} \approx 3.77\%$$

这个人最有可能是中等体型的人。

题目 2. 如教材 3.54 式和 3.56 式，分别写出如下有向图和无向图对应的概率分布.

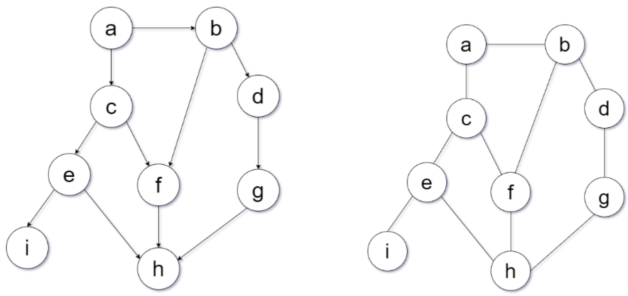


图1 有向图　　图2 无向图

解答. (1) 通过看图，先找出所有节点的父节点:

$$
\begin{aligned}
&Pa\mathcal{G}(a) = \{\} \\
&Pa\mathcal{G}(b) = \{a\} \\
&Pa\mathcal{G}(c) = \{a\} \\
&Pa\mathcal{G}(d) = \{b\} \\
&Pa\mathcal{G}(e) = \{c\} \\
&Pa\mathcal{G}(f) = \{c, b\} \\
&Pa\mathcal{G}(g) = \{d\} \\
&Pa\mathcal{G}(h) = \{e, f, g\} \\
&Pa\mathcal{G}(i) = \{e\} \\
&\text{则P(x)} = \prod_i p(x_i | Pa\mathcal{G}(x_i)) \\
&\qquad = p(a)p(b|a)p(c|a)p(d|b)p(e|c)p(f|c,b)p(g|d)p(h|e,f,g)p(i|e)
\end{aligned}
$$

(2) 通过观察无向图图的连通性，可以划分为 11 个团，正好对应于每条边的两个节点为一个团。

$$
p(a,b,c,d,e,f,g,h,i) = \frac{1}{Z}\phi^{(1)}(a,b)\phi^{(2)}(a,c)\phi^{(3)}(b,d)\phi^{(4)}(c,e)
$$
$$
\phi^{(5)}(c,f)\phi^{(6)}(d,g)\phi^{(7)}(e,i)\phi^{(8)}(e,h)\phi^{(9)}(e,h)\phi^{(10)}(e,h)\phi^{(11)}(e,h)
$$

其中归一化常数 Z 被定义为 ϕ 函数乘积的所有状态的求和或积分。

题目 3. 取值为 0,1,2,3,4,5 的概率分别为 1/2,1/4,1/8,1/16,1/32,1/32。求其香农熵.

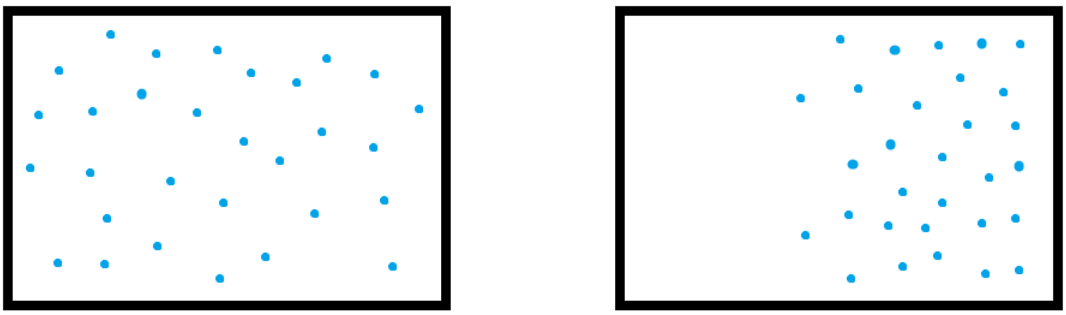
解答. 设 X 取值为 0,1,2,3,4,5 的概率分别为 1/2,1/4,1/8,1/16,1/32,1/32.

$$
\begin{aligned}
H(x) &= \mathbb{E}_{X\sim P}[I(x)] = -\mathbb{E}_{X\sim P}[logP(x)] \\
&= -\sum_{i=1}^{6} P(X=x_i)logP(X=x_i) \\
&= \frac{1}{2}log2 + \frac{1}{4}log4 + \frac{1}{8}log8 + \frac{1}{16}log16 + 2*\frac{1}{32}log32 \\
&= \frac{31}{16}log2
\end{aligned}
$$

题目 4. 叙述 KL 散度和交叉熵定义，并给出自己的理解

解答. 我并非数学专业，对于数学公式没有直观的感觉，只能通过例子去体会数学公式

首先从熵开始，第一次接触熵的概念，是在高中化学课上，化学老师表达的意思是，如果分子无序分布，则熵高；如果分子有序地分布在某一边，则熵低。



无序分布，熵高　　相对有序分布，熵低

熵大概表示着混乱程度，不确定程度，和决策困难度?

熵的定义：

$$H(X) = -\sum_{i=1}^{n} p(x_i)logp(x_i)$$

在逛大学路的时候，看到美人蹄和 1 串成名两家店，两家店门口人数差不多。小王打算从其中一家店买点东西，但是他很纠结，假设他选中每一家店的概率都是 0.5。

表 1: 夜宵种类选择概率.

x	**p(x)**
美人蹄	0.5
1 串成名	0.5

根据熵的计算公式得出：$H(X) = -\frac{1}{2}log_2\frac{1}{2} - \frac{1}{2}log_2\frac{1}{2} = 1$
这里计算熵我们都用 2 位底数。

再看看饮品店:

桂花茶和炒酸奶前的人群分布显然不一样，我们假设小王选择桂花茶和炒酸奶的概率分别为 0.1 和 0.9

表 2: 饮料种类选择概率.

x	**p(x)**
桂源铺	0.1
又炒酸奶	0.9

根据熵的计算公式得出：$H(X) = -\frac{1}{10}log_2\frac{1}{10} - \frac{9}{10}log_2\frac{9}{10} \approx 0.1954$ 跟夜宵计算出来的熵相比，决定夜宵吃什么比决定冷饮喝什么，更纠结。看看上述两组店铺的人群的分布，像不像最开始描述的分子的分布一样。

对于离散随机变量，如果概率密度是由一个 1 和若干个 0 组成，如

表 3: 离散型随机变量的特殊情况.

x	**p(x)**
1	1
2	0
⋮	⋮
n	0

根据熵的计算公式 $H(X) = 1log1 + 0log0 + \cdots + 0log0 = 0$。得到熵等于 0. 这算是一个比较重要的特点。
然后看看 KL 散度，KL 散度表示着 P 和 Q 分布的不对称性。是一种量化两种概率分布 P 和 Q 之间差异的方式

$$D_{KL}(p||q) = \sum_{i=1}^{n} p(x_i) log(\frac{p(x_i)}{q(x_i)})$$

表 4: 分布不对称的一个例子.

x	**p(x)**	**q(x)**
1	0.1	0.9
2	0.9	0.1

根据熵的计算公式有 $-\sum_{i=1}^{n} p(x_i) logp(x_i) = -\sum_{i=1}^{n} q(x_i) logq(x_i)$ 熵是相等的。

$$D_{KL}(p||q) = 0.1[log_2 0.1 - log_2 0.9] + 0.9(log_2 0.9 - log_2 0.1) \approx 2.536$$

KL 散度有一个特殊的情况，如果在某些取值上，p 有分布，而 q 没有分布，则 KL 散度会变成正无穷。

表 5: 分布不对称的一个例子.

x	**p(x)**	**q(x)**
1	0.5	1
2	0.5	0

$$D_{KL}(p||q) = 0.5log(\frac{0.5}{1}) + 0.5log(\frac{0.5}{0})) = +\infty$$
$$D_{KL}(p||q) = 1 * log_2(\frac{1}{0.5}) + 0 * log(\frac{0}{0.5})) = 1$$
$$\text{其中} \lim_{x \to 0^+} xlogx = 0$$

KL 散度表示着用 Q 分布描述 P 分布需要增加的字节数。

表 6: KL 散度表示编码位数差的一个例子.

天气	**地区 1 的概率 p(x)**	**地区 1 的编码**	**地区 2 的概率 q(x)**
晴天	$\frac{1}{2}$	1	$\frac{1}{8}$
多云	$\frac{1}{4}$	01	$\frac{1}{8}$
小雨	$\frac{1}{8}$	011	$\frac{1}{4}$
大雨	$\frac{1}{8}$	010	$\frac{1}{2}$

$$D_{KL}(p||q)=\frac{1}{2}log_2\frac{\frac{1}{2}}{\frac{1}{8}}+\frac{1}{4}log_2\frac{\frac{1}{4}}{\frac{1}{8}}+\frac{1}{8}log_2\frac{\frac{1}{8}}{\frac{1}{4}}+\frac{1}{8}log_2\frac{\frac{1}{8}}{\frac{1}{2}}=\frac{7}{8}$$

$$H(P)=-\frac{1}{2}log_2\frac{1}{2}-\frac{1}{4}log_2\frac{1}{4}-\frac{1}{8}log_2\frac{1}{8}-\frac{1}{8}log_2\frac{1}{8}=\frac{14}{8}$$

$$E(\text{用地区 1 的编码编码地区 1 的天气用的编码位数})=$$

$$\frac{1}{2}*1+\frac{1}{4}*2+\frac{1}{8}*8+\frac{1}{8}*3=\frac{14}{8}$$

$$E(\text{用地区 1 的编码编码地区 2 的天气用的编码位数})=$$

$$\frac{1}{8}*1+\frac{1}{8}*2+\frac{1}{4}*3+\frac{1}{2}*3=\frac{21}{8}$$

所以在这个例子中:

$$H(P)=E(\text{用地区 1 的编码编码地区 1 的天气用的编码位数})$$

$$E(\text{用地区 1 的编码编码地区 1 的天气用的编码位数})=$$

$$E(\text{用地区 1 的编码编码地区 2 的天气用的编码位数})-D_{KL}(p||q)$$

$$D_{KL}(p||q)=H(p,q)-H(p)$$

当 $H(p)=0$ 时，$D_{KL}(p||q)=H(p,q)$。因此在机器学习的分类任务领域中，让 p 表示真实分布的标签，比如猫狗分类，手写体数字识别。真实的标签由一个 1 和若干个 0 组成，因此 $H(p)=0$。所以在机器学习领域，交叉熵等于 KL 散度。

题目 5. 机器学习初识：学习 sklearn 中的朴素贝叶斯

Sklearn 为我们提供了这几种朴素贝叶斯的分类器：

naive_bayes.GaussianNB

naive_bayes.BernoulliNB

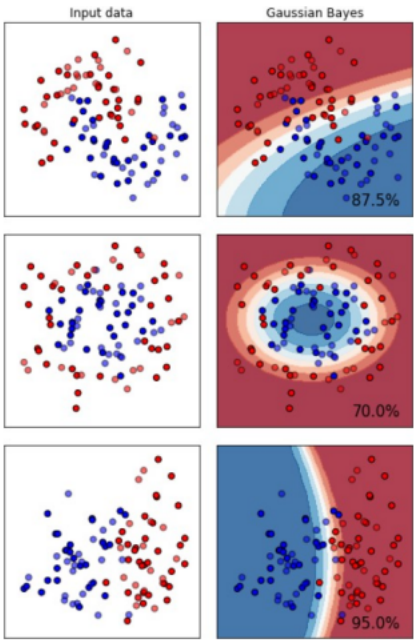
naive_bayes.MultinomialNB

naive_bayes.ComplementNB

1. 说明每种分类器的含义以及区别

2. 利用 sklearning 自带的数据集：(from sklearn.datasets import make_moons, make_circles, make_classification)，自行划分训练集和测试集（无需验证集），使用第一种分类器 (naive_bayes.GaussianNB)，并且可视化结果，说明在不同数据类型下分类器效果的差别。

举例效果如下：



解答. 1. 这四种模型的选择主要取决于输入数据的类型。

naive_bayes.GaussianNB，如果参数是连续值，而且越接近正态分布越好。

naive_bayes.BernoulliNB，适用于参数都是二值变量，如果不是二值变量，可以先对变量二值化

naive_bayes.MultinomialNB, 特征变量是离散数据，而且符合多项分布。适合处理文本数据，而且输入参数不能是负数。

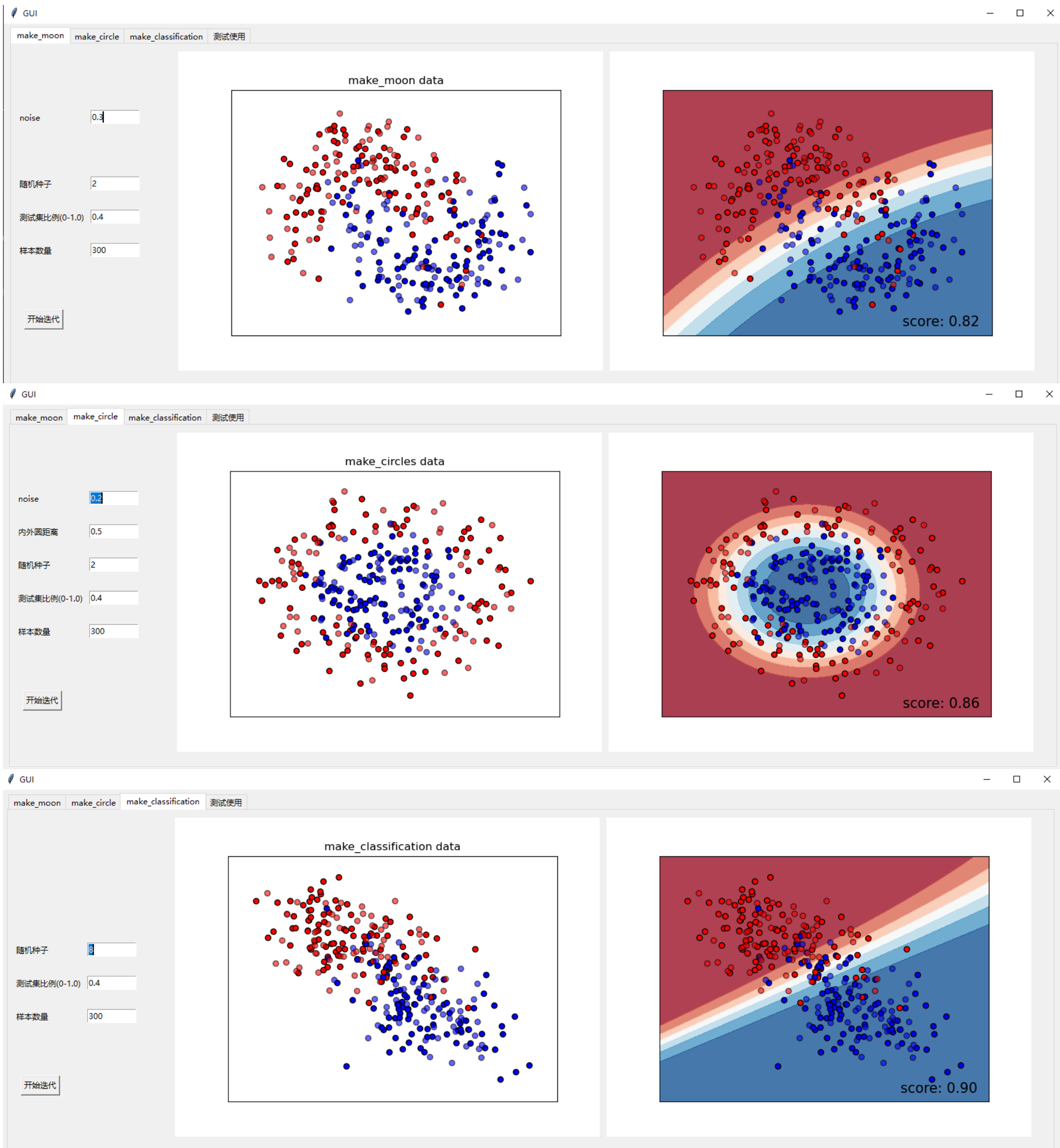
和 naive_bayes.ComplementNB 是 MultinomialNB 模型的一个变种，实现了补码朴素贝叶斯 (CNB) 算法。CNB 是标准多项式朴素贝叶斯 (MNB) 算法的一种改进，比较适用于不平衡的数据集，在文本分类上的结果通常比 MultinomialNB 模型好。

2.

从结果上来看，并不是想象中拟合出两个山峰一样的概率密度，首先这个决策边界的形状，是由两个不相关的正态分布组成，可以是椭圆形，可以是"X"型，或者说椭圆很大，以至于显示的部分看起来像一个直线。显然在 make_circles 数据集里面拟合的最好，但是如果内外圆距离太近，就会过拟合，样本多一些就没事了。

以 make_moon 数据为例，在上图看到接近直线的决策边界，其实是一个大椭圆的一部分。这个大椭圆就是二维正态分布峰值部分。

应用不同的数据集，能看到决策边界可能是"X"型，这说明两个正态分布可能是一正一负。


GUI
make_moon
make_circle
make_classification
noise
0.3
随机种子
2
测试集比例(0-1.0)
0.4
样本数量
300
开始迭代
make_moon data
score: 0.82
GUI
make_moon
make_circle
make_classification
noise
0.2
内外圆距离
0.5
随机种子
2
测试集比例(0-1.0)
0.4
样本数量
300
开始迭代
make_circles data
score: 0.86
GUI
make_moon
make_circle
make_classification
随机种子
测试集比例(0-1.0)
0.4
样本数量
300
开始迭代
make_classification data
score: 0.90

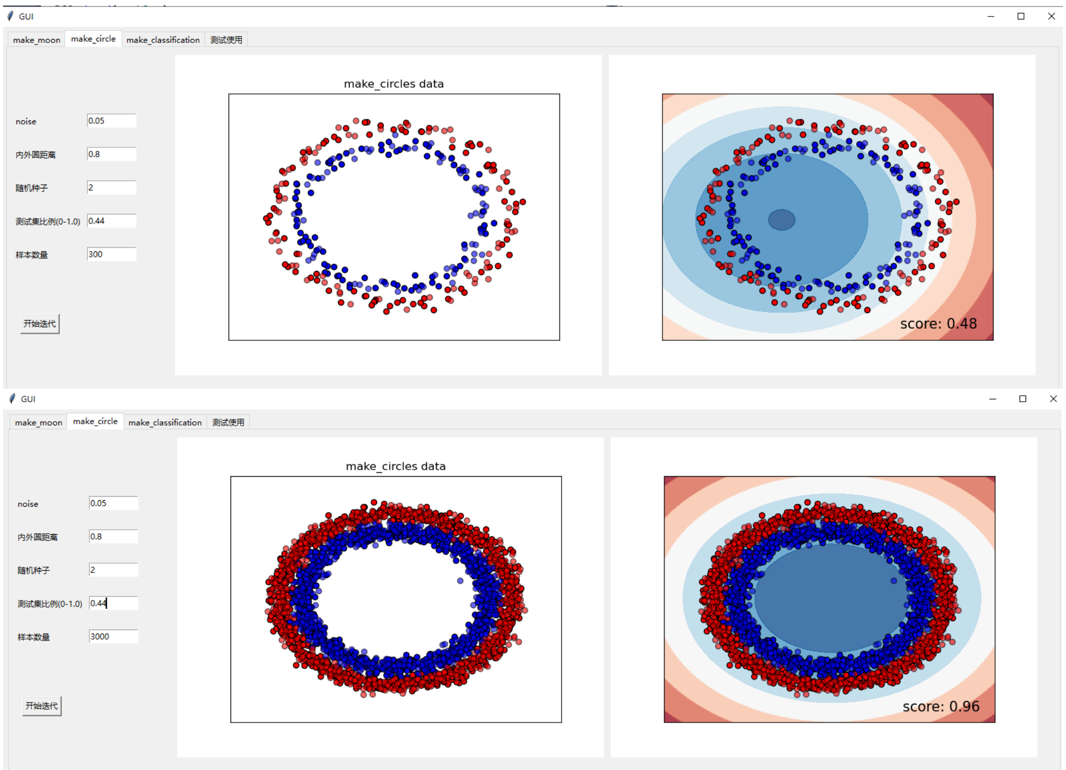


图 1: 样本太少导致过拟合

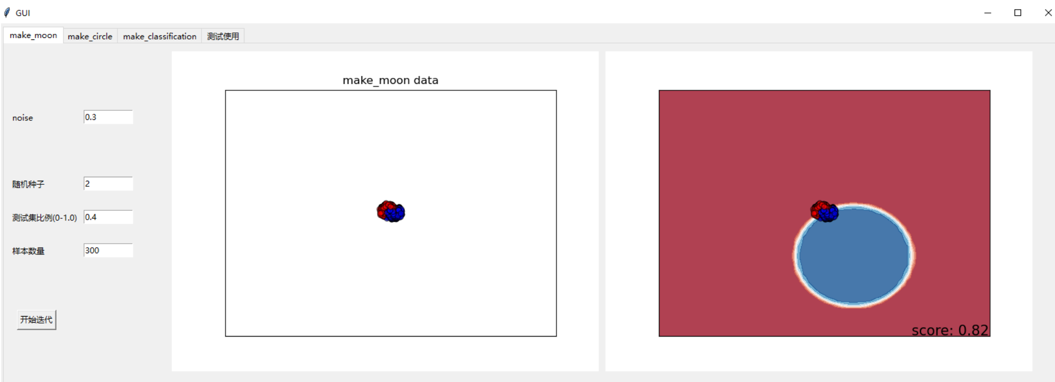


图 2: 在 make_moon 数据集观察到椭圆形决策边界

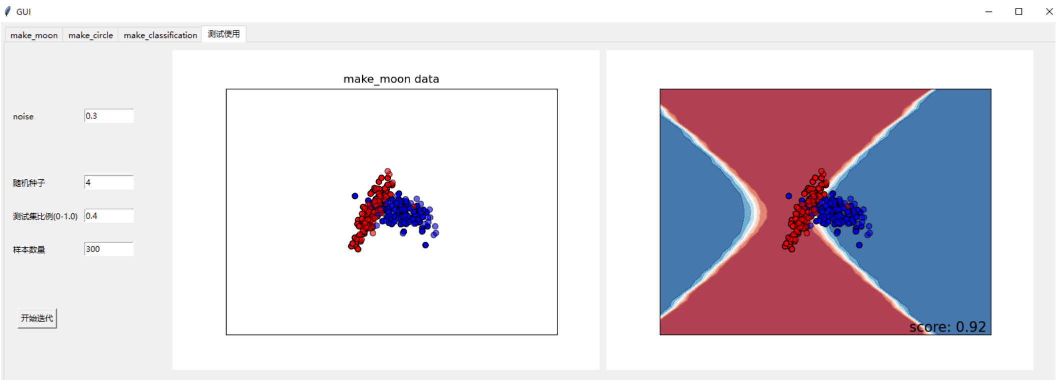


图 3: 在 make_clasification 数据集观察到“X”形决策边界