

智能之门

神经网络和深度学习入门

(基于Python的实现)

STEP 5 非线性分类

第 11 章

多入多出的双层神经网络

非线性多分类

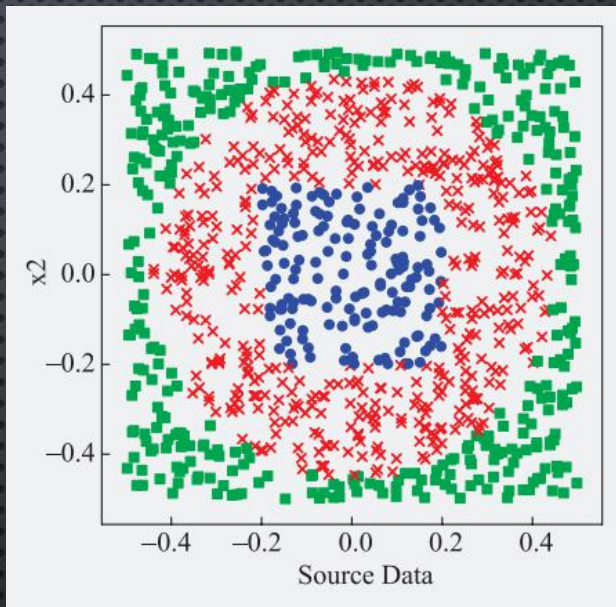
- 11.1 双变量非线性多分类
- 11.2 非线性多分类的实现
- 11.3 非线性多分类的原理
- 11.4 分类样本不平衡问题

解决完二分类问题，我们将学习如何解决更复杂的三分类问题，由于样本的复杂性，必须在隐层使用多个神经元才能完成分类任务。

11.1 双变量非线性多分类

前面用异或问题和弧形样本学习了二分类，本章将讲解非线性多分类。部分数据样本和可视化结果如下。样本组成了一个貌似铜钱的形状，因此把这个问题叫“铜钱孔形分类”问题。

问题：如何用两层神经网络实现这个铜钱孔三分类问题？



样本	x_1	x_2	y
1	0.228 251 11	-0.345 870 97	2
2	0.209 826 06	0.433 884 47	3
...
1 000	0.382 301 43	-0.164 553 77	2

11.1 双变量非线性多分类

➤ 评估标准

- 以三分类问题举例，假设每类有 100 个样本，一共 300 个样本，最后的分类结果如下表：

样本所属类别	分到类 1	分到类 2	分到类 3	各类样本总数	准确率
类 1	90	4	6	100	90%
类 2	9	84	5	100	84%
类 3	1	4	95	100	95%
总数	101	93	106	300	89.67%

- 总体的准确率是89.67%。三类的精确率是90%、84%、95%。实际上上表也是混淆矩阵在二分类基础上的扩展形式，其在对角线上的值越大越好。
- 当然也可以计算每个类别的Precision和Recall，但是只在需要时才去做具体计算。

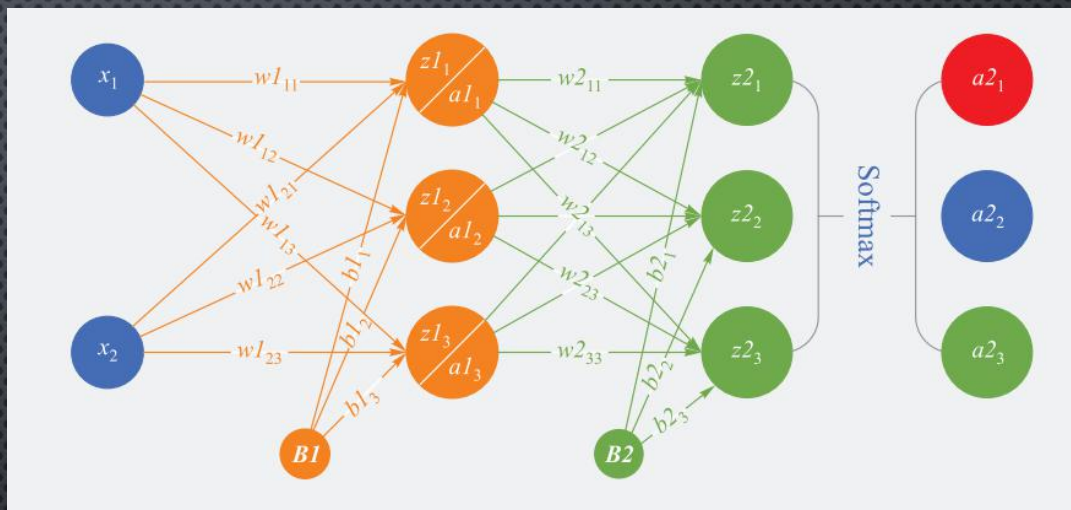
11.2 非线性多分类的实现

➤ 神经网络结构

- 输入层: $X = (x_1 \ x_2)$ 。
- 隐层: $Z1 = (z1_1 \ z1_2 \ z1_3)$,
 $A1 = (a1_1 \ a1_2 \ a1_3)$
- 输出层: $Z2 = (z2_1 \ z2_2 \ z2_3)$,
 $A2 = (a2_1 \ a2_2 \ a2_3)$ 。

- 隐层权重和偏置:

$$W1 = \begin{pmatrix} w1_{11} & w1_{12} & w1_{13} \\ w1_{21} & w1_{22} & w1_{23} \end{pmatrix}$$
$$B1 = (b1_1 \ b1_2 \ b1_3)$$



- 输出层权重和偏置:

$$W2 = \begin{pmatrix} w2_{11} & w2_{12} & w2_{13} \\ w2_{21} & w2_{22} & w2_{23} \\ w2_{31} & w2_{32} & w2_{33} \end{pmatrix}$$
$$B2 = (b2_1 \ b2_2 \ b2_3)$$

11.2 非线性多分类的实现

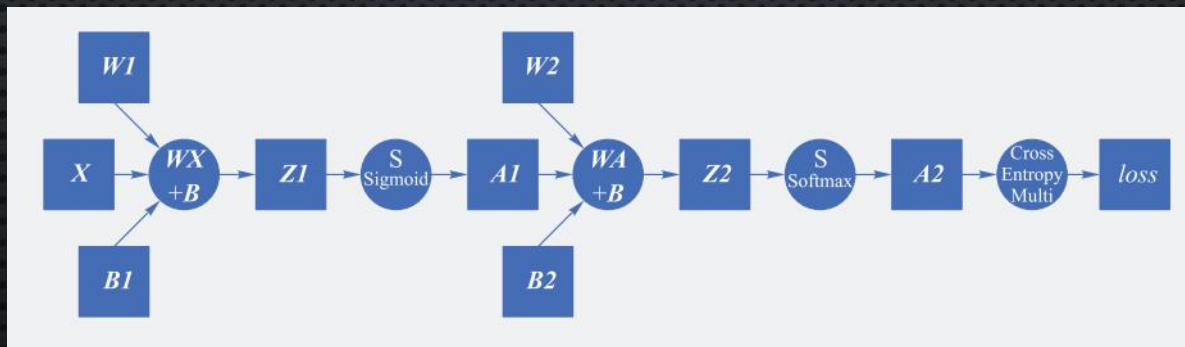
➤ 前向计算

- 层间计算

$$Z1 = X \cdot W1 + B1, \quad A1 = \text{Sigmoid}(Z1), \quad Z2 = A1 \cdot W2 + B2, \quad A2 = \text{Softmax}(Z2)$$

- 损失函数

$$\text{loss} = -(y_1 \ln a2_1 + y_2 \ln a2_2 + y_3 \ln a2_3), \quad J(W, B) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n y_{ij} \ln a2_{ij}$$



11.2 非线性多分类的实现

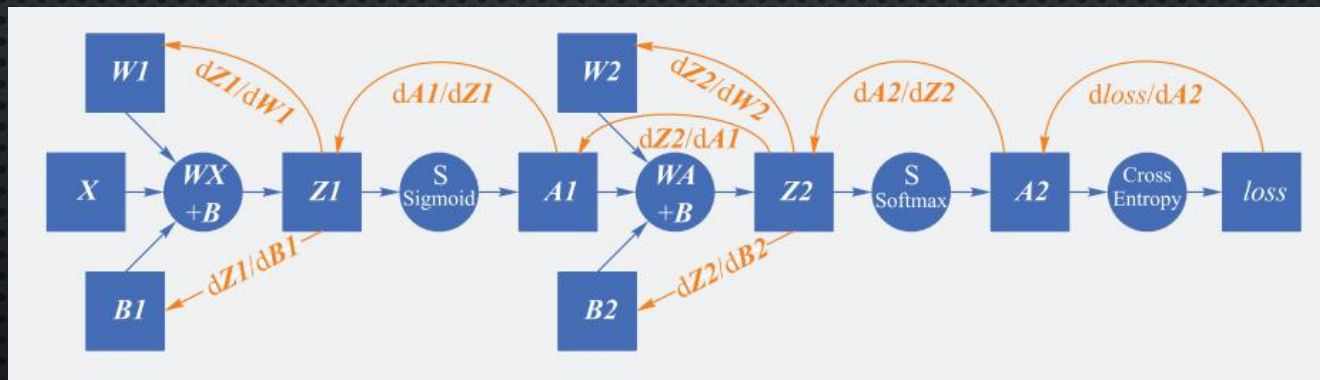
➤ 反向传播

- 链式法则求导，结果类似于此前章节的推导结果：

$$dZ2 = \frac{\partial loss}{\partial Z2} = A2 - Y, \quad dW2 = \frac{\partial loss}{\partial W2} = A1^T \cdot dZ2, \quad dB2 = \frac{\partial loss}{\partial B2} = dZ2$$

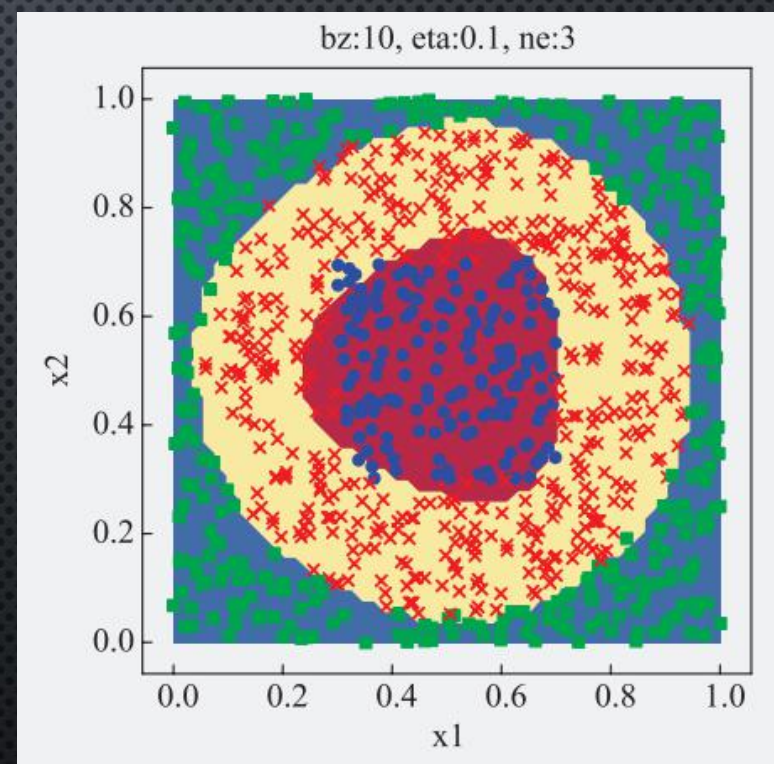
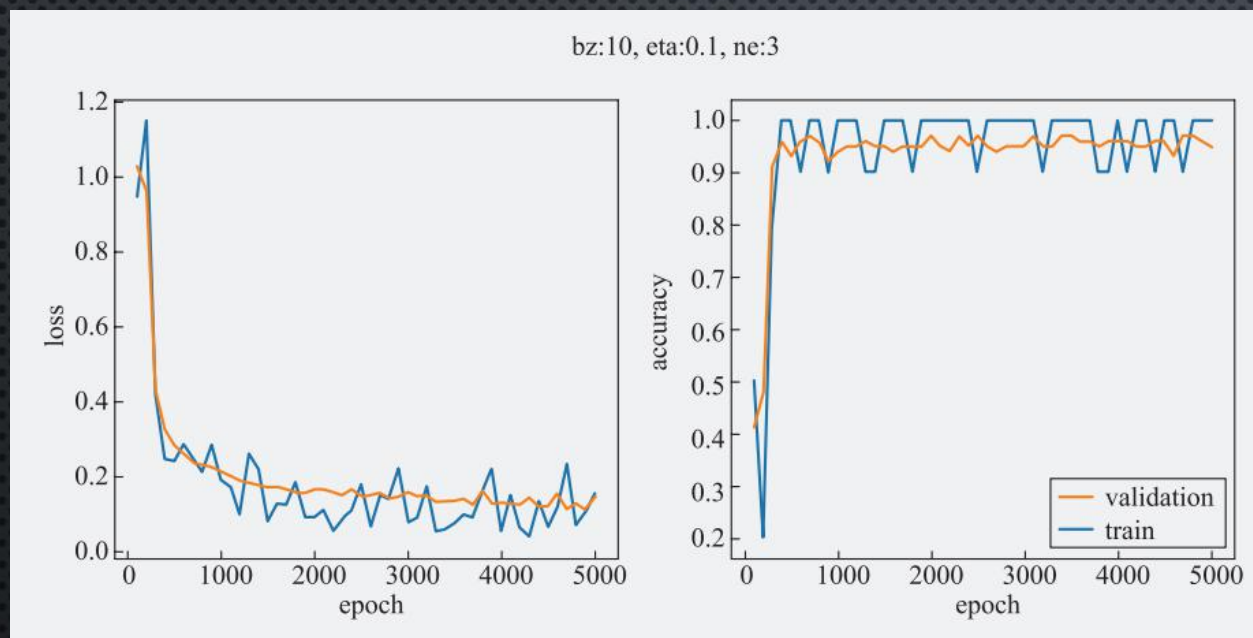
$$dA1 = \frac{\partial loss}{\partial A1} = dZ2 \cdot W2^T, \quad dZ1 = \frac{\partial loss}{\partial Z1} = dZ2 \cdot W2^T \odot dA1,$$

$$dW1 = X^T \cdot dZ1, \quad dB1 = dZ1$$



11.2 非线性多分类的实现

➤ 迭代运行结果



11.3 非线性多分类的原理

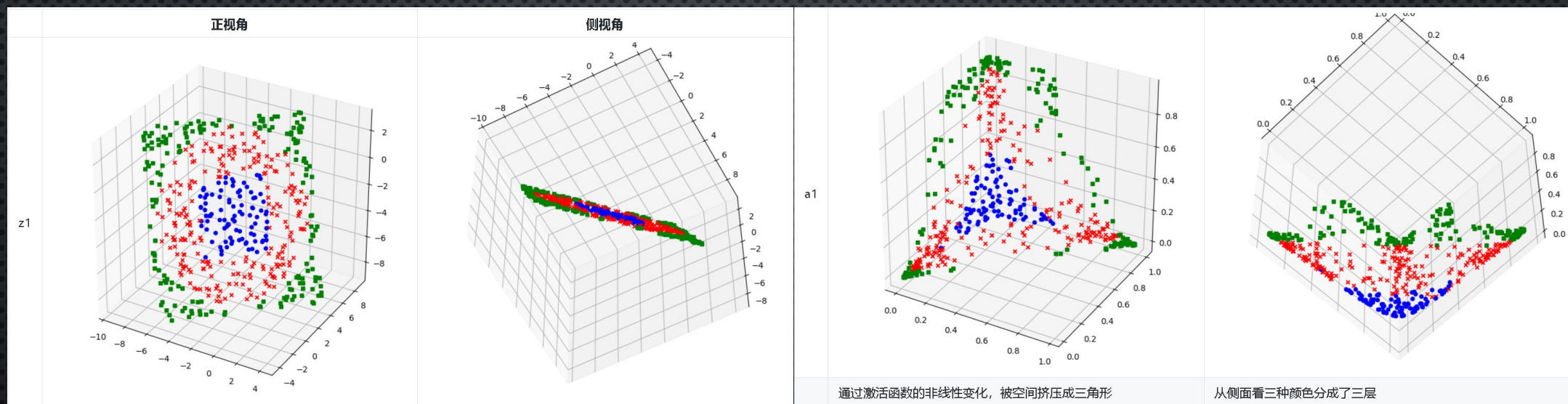
➤ 隐层神经元数量的影响

- 2个神经元：没有完成分类任务。
- 4、8个神经元：基本完成，但是边缘不够清晰。
- 16、32、64个神经元：较好地完成了分类任务。

11.3 非线性多分类的原理

➤ 三维空间内的变换过程

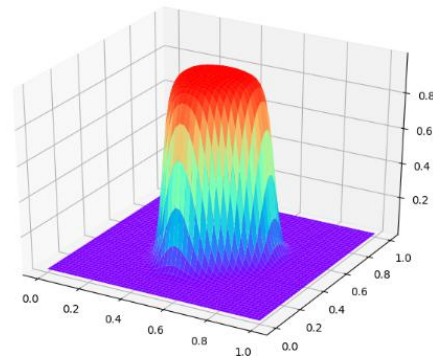
- 观察net.A1的侧视图，似乎是已经分层了，蓝点沉积下去，绿点浮上来，红点在中间，像鸡尾酒一样分成了三层，这就给第二层神经网络创造了做一个线性三分类的条件，只需要两个平面，就可以把三者轻松分开了。



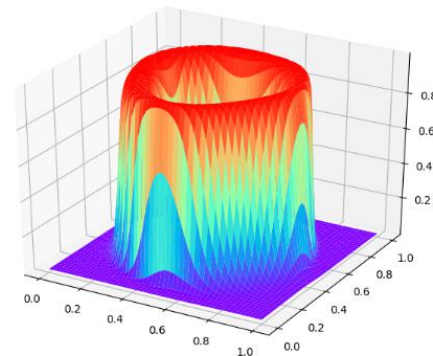
11.3 非线性多分类的原理

➤ 3D分类结果图

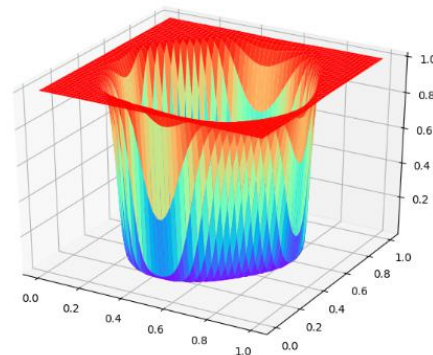
- 我们只能一次显示一个类别的分类效果图，如右图。
- 最后一行的图片显示类别1和2的累加效果。



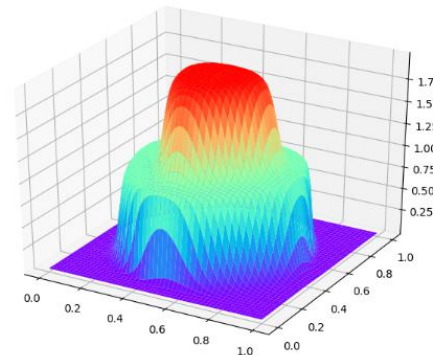
红色：类别1样本区域



红色：类别2样本区域



红色：类别3样本区域



红色：类别1，青色：类别2，紫色：类别3

11.4 分类样本不平衡问题

➤ 样本不平衡问题

- 在一般的分类学习方法中都有一个假设，就是不同类别的训练样本的数量相对平衡。
- 以二分类为例，比如正负例都各有1000个左右。如果是1200:800的比例，也是可以接受的，但是如果是1900:100，就需要有些措施来解决不平衡问题了，否则最后的训练结果很大可能是忽略了负例，将所有样本都分类为正类了。
- 如果是三分类，假设三个类别的样本比例为：1000:800:600，这是可以接受的；但如果是1000:300:100，就属于不平衡了。它带来的结果是分类器对第一类样本过拟合，而对其它两个类别的样本欠拟合，测试效果一定很糟糕。

11.4 分类样本不平衡问题

➤ 如何解决？

- **平衡数据集：**有一句话叫做“更多的数据往往战胜更好的算法”。所以一定要先想办法扩充样本数量少的类别的数据，比如目前的正负类样本数量是1000:100，则可以再搜集2000个数据，最后得到了2800:300的比例，此时可以从正类样本中丢弃一些，变成500:300，就可以训练了。
- **尝试其它评价指标：**从前面的分析可以看出，准确度这个评价指标在类别不均衡的分类任务中并不能有效工作，甚至进行误导（但是从这个指标来看，该分类器有着很好的评价指标得分）。因此在类别不均衡分类任务中，需要使用更有说服力的评价指标来对分类器进行评价。常规的分类评价指标可能会失效，比如将所有的样本都分类成大类，那么准确率、精确率等都会很高。这种情况下，AUC是最好的评价指标。

11.4 分类样本不平衡问题

- **尝试产生人工数据样本：**一种简单的人工样本数据产生的方法便是，对该类下的所有样本每个属性特征的取值空间中随机选取一个组成新的样本，即属性值随机采样。你可以使用基于经验对属性值进行随机采样而构造新的人工样本，或者使用类似朴素贝叶斯方法假设各属性之间互相独立进行采样，这样便可得到更多的数据，但是无法保证属性之前的线性关系（如果本身是存在的）。
- **尝试一个新的角度理解问题：**我们可以从不同于分类的角度去解决数据不均衡性问题，我们可以把那些小类的样本作为异常点，因此该问题便转化为异常点检测与变化趋势检测问题。将小类样本作为异常点这种思维的转变，可以帮助考虑新的方法去分离或分类样本。这两种方法从不同的角度去思考，让你尝试新的方法去解决问题。

11.4 分类样本不平衡问题

- **修改现有算法：**设超大类中样本的个数是极小类中样本个数的 L 倍，那么在随机梯度下降算法中，每次遇到一个极小类中样本进行训练时，训练 L 次。将大类中样本划分到 L 个聚类中，然后训练 L 个分类器，每个分类器使用大类中的一个簇与所有的小类样本进行训练得到。最后对这 L 个分类器采取少数服从多数对未知类别数据进行分类，如果是连续值（预测），那么采用平均值。设小类中有 N 个样本。将大类聚类成 N 个簇，然后使用每个簇的中心组成大类中的 N 个样本，加上小类中所有的样本进行训练。
- **集成学习：**该方法是一个boosting算法，它递归地训练三个弱学习器，然后将这三个弱学习器结合起形成一个强的学习器。我们可以使用这个算法的第一步去解决数据不平衡问题。
 - ✓ 假设是一个二分类问题，大部分的样本都是true类。让 $L1$ 输出始终为true。使用50%在 $L1$ 分类正确的与50%分类错误的样本训练得到 $L2$ ，即从 $L1$ 中学习错误的样本集与学习正确的样本集中，循环一边采样一个。因此， $L2$ 的训练样本是平衡的。再使用 $L1$ 与 $L2$ 分类不一致的那些样本训练得到 $L3$ ，即在 $L2$ 中分类为false的那些样本。最后，结合这三个分类器，采用投票的方式来决定分类结果，因此只有当 $L2$ 与 $L3$ 都分类为false时，最终结果才为false，否则true。

THE END

谢谢！