# 察异辨花——监督学习

陕西师范大学附中 李靖

# 情景再现

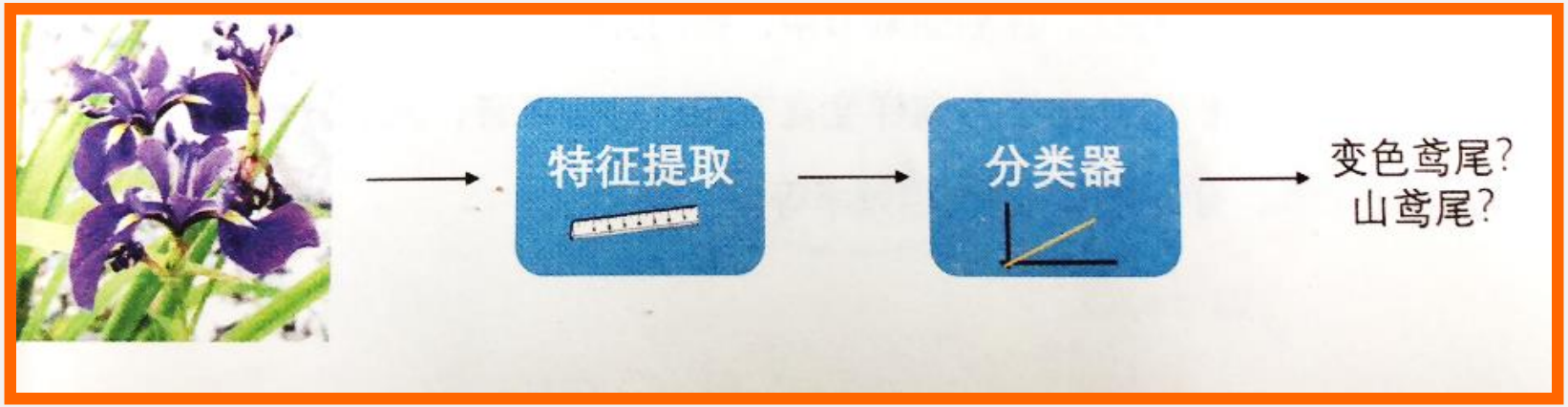
每当看到一张图片，我们就能分清图片上的事物，在生活中，我们经常会判断一个事物的类型，这样的过程在人工智能领域被称为分类。

# 情景再现

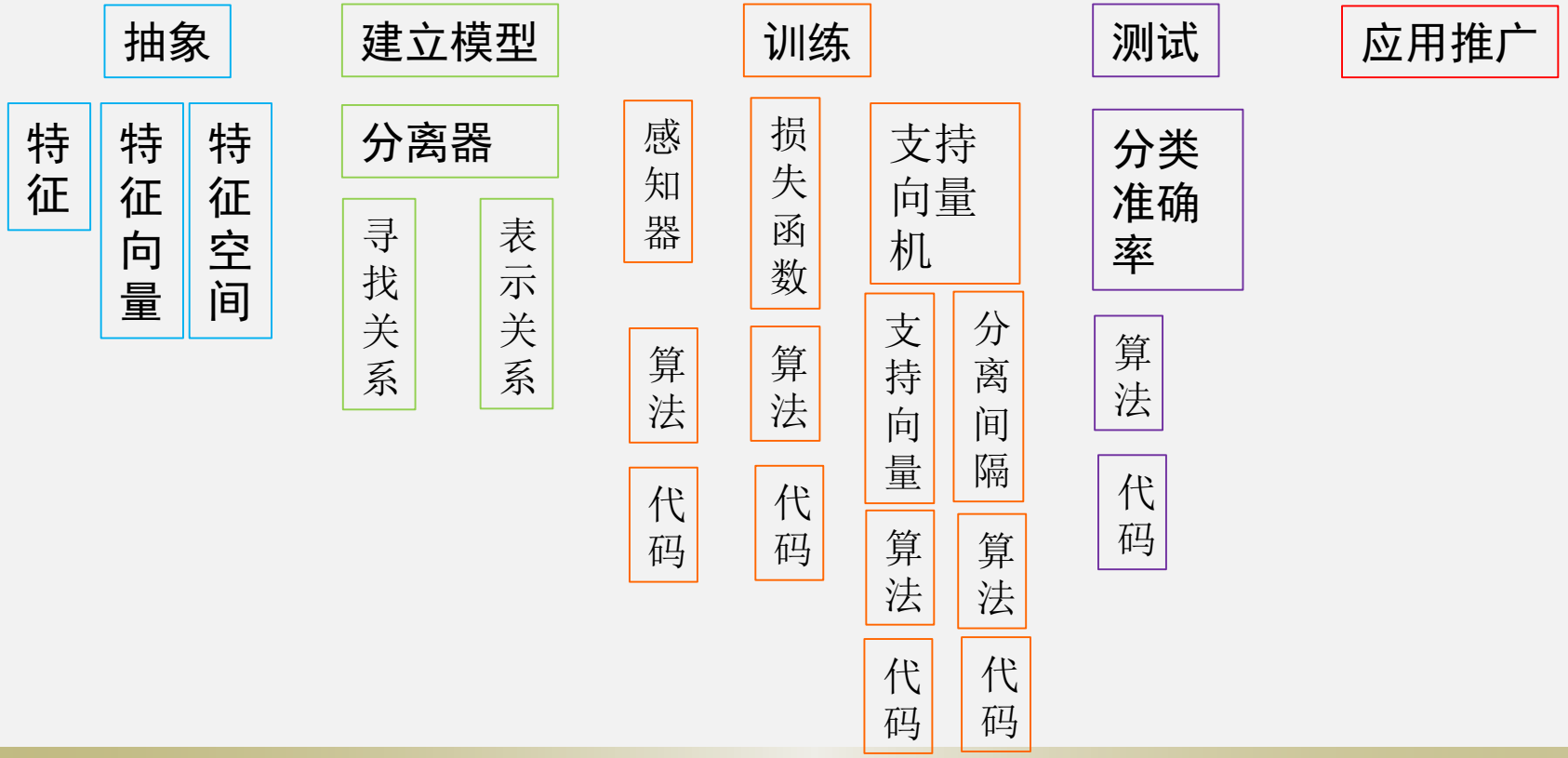
这是同一种花吗？你是如何判断的？

# 情景再现

怎样构建一个人工智能系统，能够区分变色鸢尾和山鸢尾？

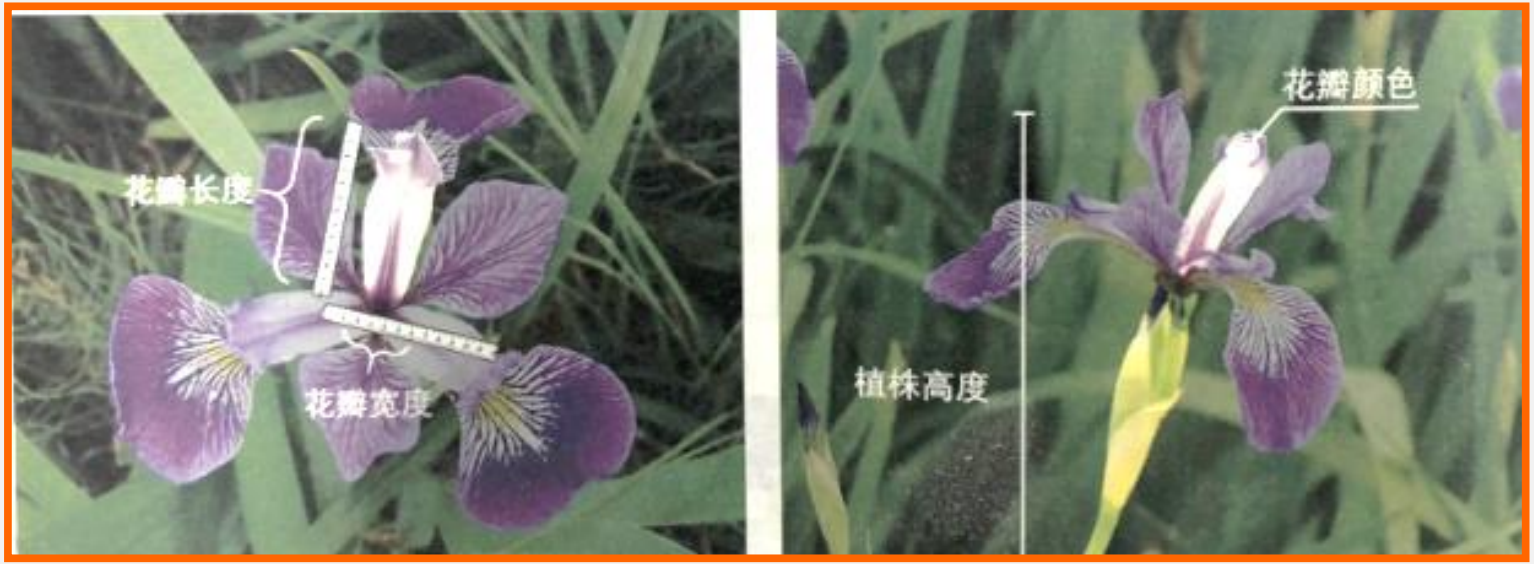


# 建立一个二分智能系统的过程



# 提取特征

怎样的特征能够区分变色鸢尾和山鸢尾？



# 提取特征

在考虑数据特点，类别差异的基础上设计有效的特征。

对于图像，利用方向梯度直方图；

对于声音，设计了梅尔频率倒谱系数；

对于视频，有光流直方图；

对于文本，有词频率—逆文档频率。

# 提取特征

## 特征向量

通过实际测量，可以得到鸢尾花的特征（花瓣长度和宽度），但是在数学上如何表示它们呢？

长度x1，宽度x2，进一步把两个数字放在一起（x1,x2）作为一组数据。

（x1,x2）二维向量

# 提取特征

## 特征向量

向量还可以进行运算

加减法：两个相同维度数的向量相加减，对应数字相加减

数量乘法：一个数和一个向量相乘，这个数和向量中的每一个数字相乘

内积：两个具有相同维数的向量做内积，每个 数字对应相乘并求和

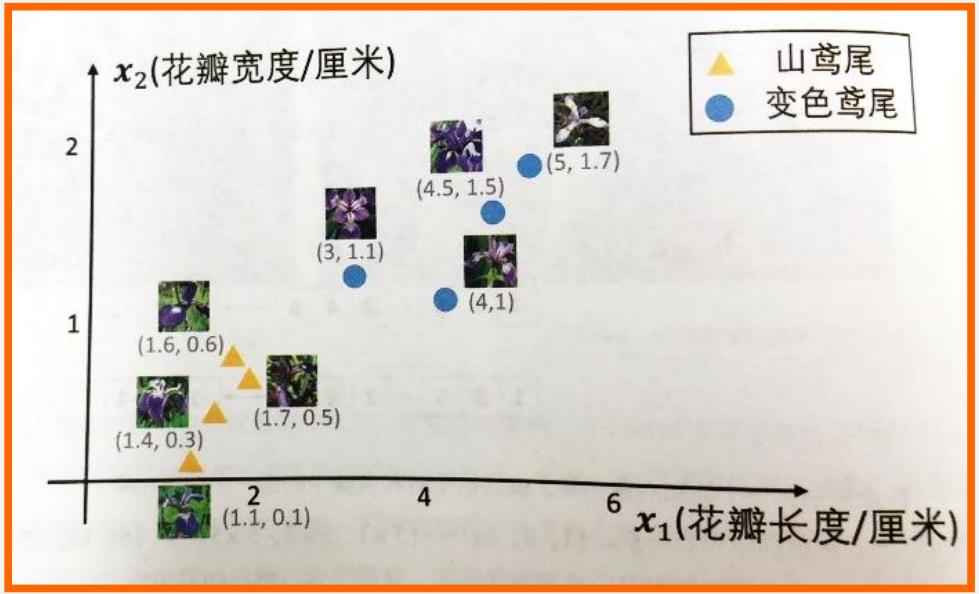
# 提取特征

## 特征点和特征空间

将数据用特征向量表示后，就可以把特征向量表示在直角坐标系中。（1.1,0.1）就可以看做直角坐标系中的一个点。

怎么来衡量两点之间的距离呢？

$$D=\sqrt{(x1-z1)^2+(x2-z2)^2}$$



# 分类器

## 分类器

就是有特征向量到预测类别的函数。

$0.5x1 + x2 - 2 = 0$

$$g(x1,x2) = \begin{cases} 1, & 0.5x1 + x2 - 2 = 0 \\ -1, & 0.5x1 + x2 - 2 = 0 \end{cases}$$


F(x1,x2...xn)=a1x1+a2x2+...+anxn+b

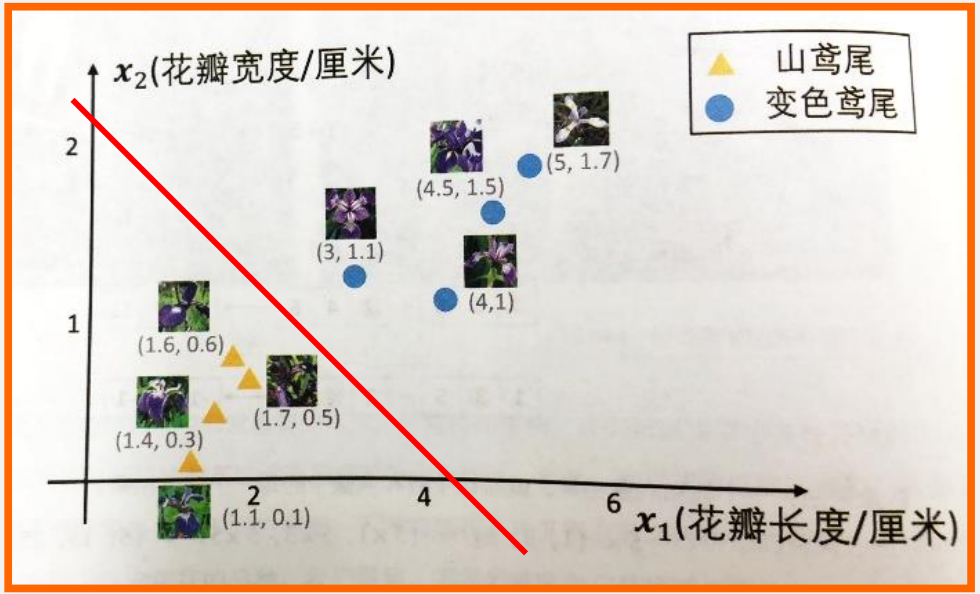
# 分类器

## 分类器

就是有特征向量到预测类别的函数。

$0.5x1 + x2 - 2 = 0$

$$g(x1,x2) = \begin{cases} 1, & 0.5x1 + x2 - 2 = 0 \\ -1, & 0.5x1 + x2 - 2 = 0 \end{cases}$$




F(x1,x2...xn)=a1x1+a2x2+...+anxn+b

线性分类器

# 训练分类器

学习 ⟷ 训练

数据

考试 ⟷ 测试

数据

解决问题 ⟷ 应用

让分类器习得到适合参数的过程叫做分类器的训练

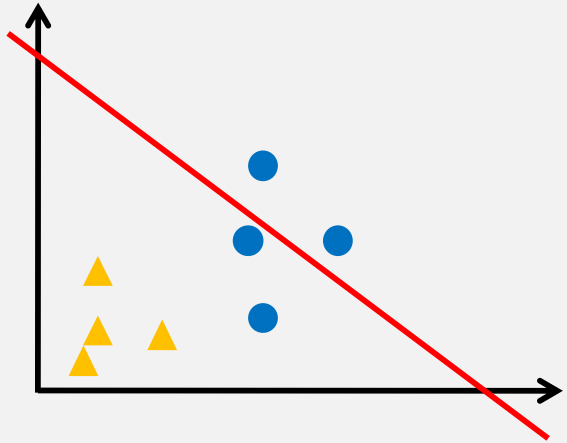
# 训练分类器

这个过程是由一系列判断和计算的步骤是组成的，也就是算法。在一个训练集上，使用不同的算法，可能会获得不同的分类器。

f（x1,x2）=a1x1+a2x2+b,目的就是找到适合的参数a1、a2、b。常见的训练线性分类器的算法——感知器和支持向量机

# 感知器

感知器，是一种训练线性分类器的算法，主要想法就是利用被错误分类的训练数据调整现有的分类器的参数，使得调整后的分类器判断得更加准确。



# 感知器

**感知器学习算法**

第一步：选取初始分类器参数 $a_1$，$a_2$，$b$；

第二步：在训练集中选取一个训练数据，如果这个训练数据被误分类，即 $y \times (a_1 x_1 + a_2 x_2 + b) \leqslant 0$，则按照以下规则更新参数（将箭头右边更新后的值赋给左边的参数）：

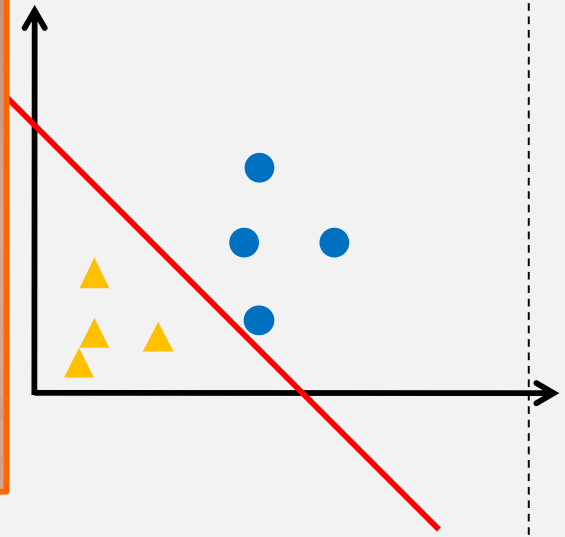
$a_1 \leftarrow a_1 + \eta y x_1$

$a_2 \leftarrow a_2 + \eta y x_2$

$b \leftarrow b + \eta y$

第三步：回到第二步，直到训练数据中没有误分类数据为止。

其中，$\eta$ 是学习率（learning rate），学习率是指每一次更新参数的程度大小。



如何衡量分类器对数据错误分类程度？

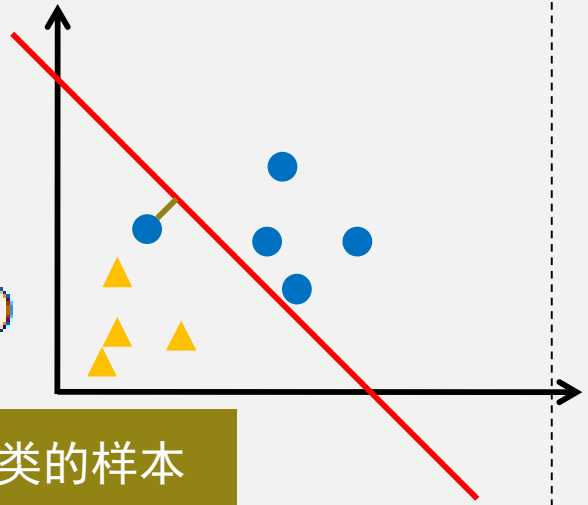
如何利用错误分类数据调整函数参数？

# 感知器

损失函数，这是在训练过程中用来度量分类器输出错误程度的数学化表示。预测错误越大，损失函数的值就越大。

感知器的损失函数l就定义为：

$$l(a1,a2,b)=\sum_{i=1}^{n}\max\left(0,-y^{(i)}\left(a1x1^{(i)}+a2x2^{(i)}+b\right)\right)$$



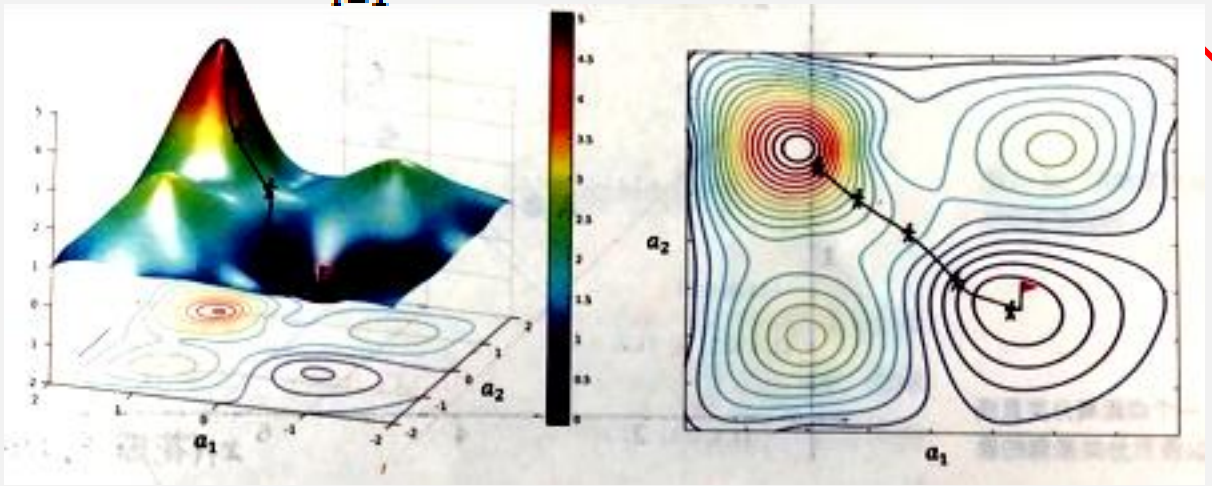
$-y^{(i)}\left(a1x1^{(i)}+a2x2^{(i)}+b\right)\geq 0$ 的样本是被误分类的样本

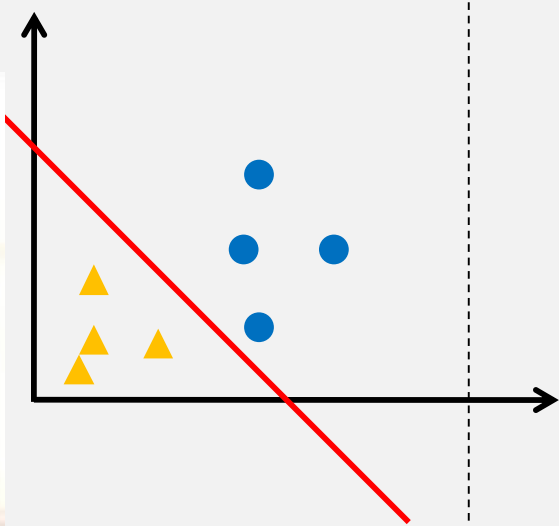
被误分类的数据点离直线远远，损失函数值就越大。

# 感知器

优化，调整分类器参数，是的损失函数值最小的过程。

$$l(a1, a2, b) = \sum_{i=1}^{n} \max\left(0, -y^{(i)}\left(a1x1^{(i)} + a2x2^{(i)} + b\right)\right)$$



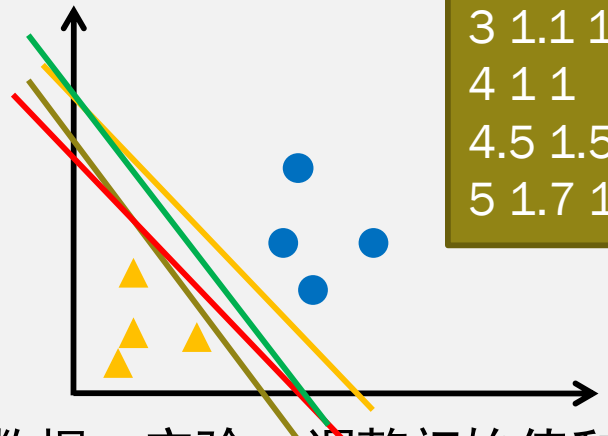



# 感知器

测试数据
8
1.6 0.6 -1
1.4 0.3 -1
1.7 0.5 -1
1.1 0.1 -1
3 1.1 1
4 1 1
4.5 1.5 1
5 1.7 1

思考：

（1）修改感知器中初始值的参数，训练多个分类器，得到的分类曲线是否相等？



（2）修改感知器学习算法中的学习率，训练多个分类器，得到的分类曲线是否相等？

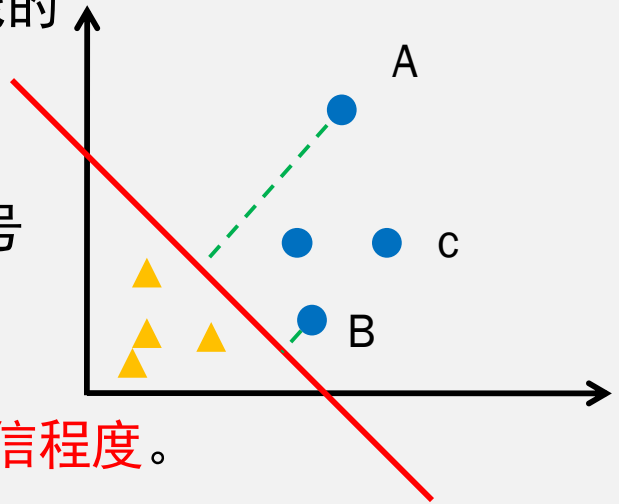
（3）理科班的同学可以尝试编写程序和测试数据，实验：调整初始值和学习率，观察其得到的分类曲线。

# 支持向量机

分类预测的确信程度:一个点距离分类直线的远近可以表示对分类的确信程度。

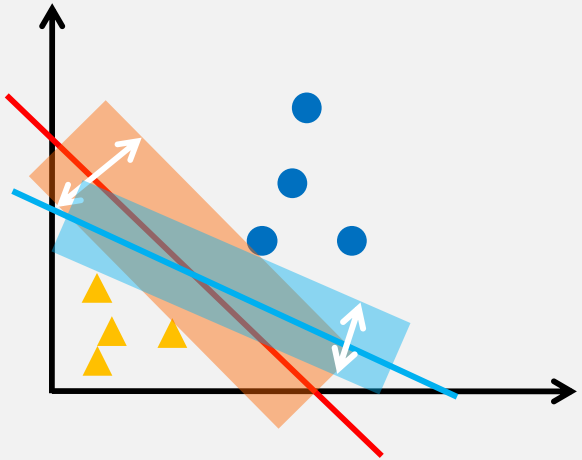
可以通过|a1x1+a2x2+b|表示相对距离，y符号是否一致，表示分类的正确性。

因此，y（a1x1+a2x2+b）表示正确性和可信程度。



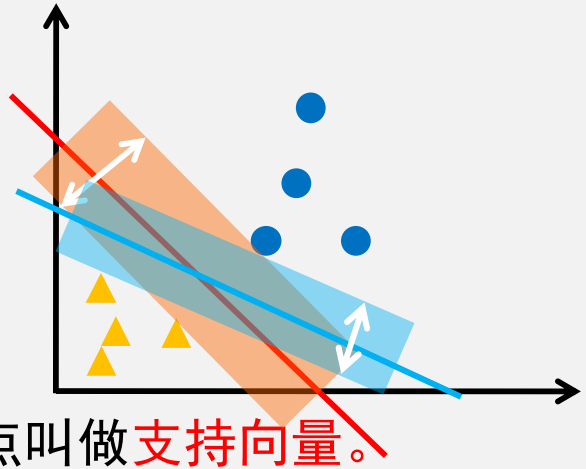
# 支持向量机

分类器间隔:只需要关注与分类直线最近点的距离，使点离分类直线越远越好。将两个类别中距离分类直线最近的点到直线的距离和，称为分类间隔。



# 支持向量机

支持向量机:是在特征空间上分类间隔最大的分类器。这里的线性支持向量机是支持向量机的一种。



最粗的直线和哪些数据有关呢？

图中和阴影部分相接触的点的数据有关，这些点叫做支持向量。

支持向量是最难被分类的数据，在求解分类任务中最富有信息的数据

# 支持向量机

问题：对于a1x1+a2x2+b=0,a1,a2,b

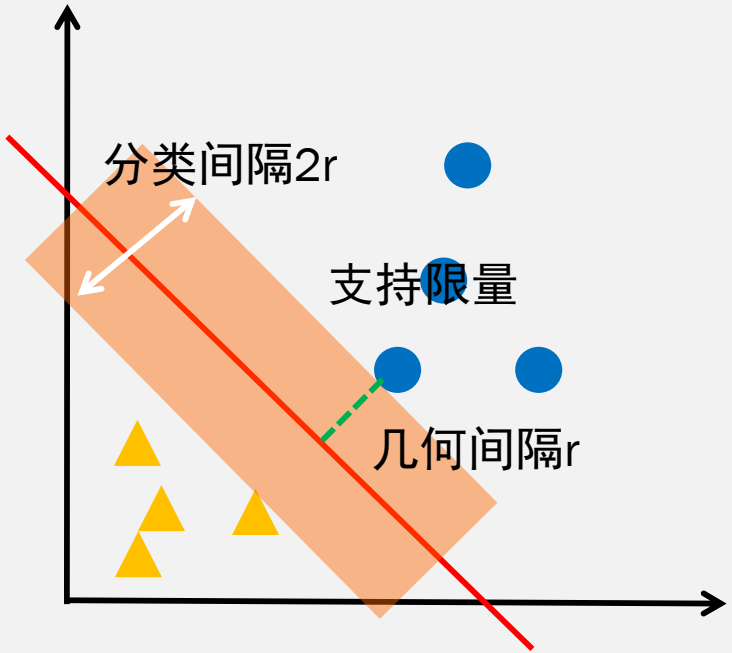
为参数，最大化分类间隔。

如果（x1,x2）被直线正确分类。

$$r^{(i)} = y^{(i)} \times \frac{a1x1^{(i)} + a2x2^{(i)} + b}{\sqrt{a1^2 + a2^2}} \quad ...(1)$$

对于全部训练数据，要找到最小值

$$r = \min_{i=1,\ldots,n} r^{(i)} \quad ...(2)$$



# 支持向量机

问题：对于a1x1+a2x2+b=0,a1,a2,b

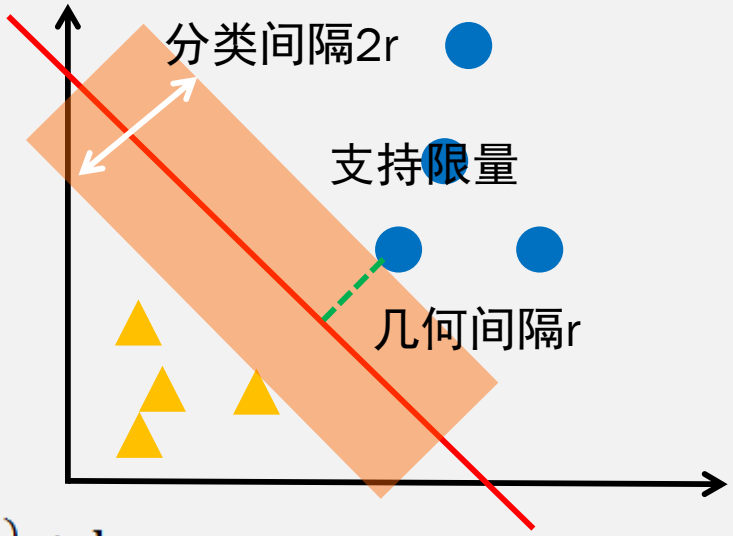
为参数，最大化分类间隔。

最大化分类间隔——

$$\max_{a1,a2,b} 2r$$

等价于最小化2/r——

$$\min_{a1,a2,b} \frac{2}{r} \quad \cdots (3)$$




对于每个i，满足

$$y^{(i)} \times \frac{a1x1^{(i)} + a2x2^{(i)} + b}{\sqrt{a1^2 + a2^2}} \geq r \quad \cdots (4)$$

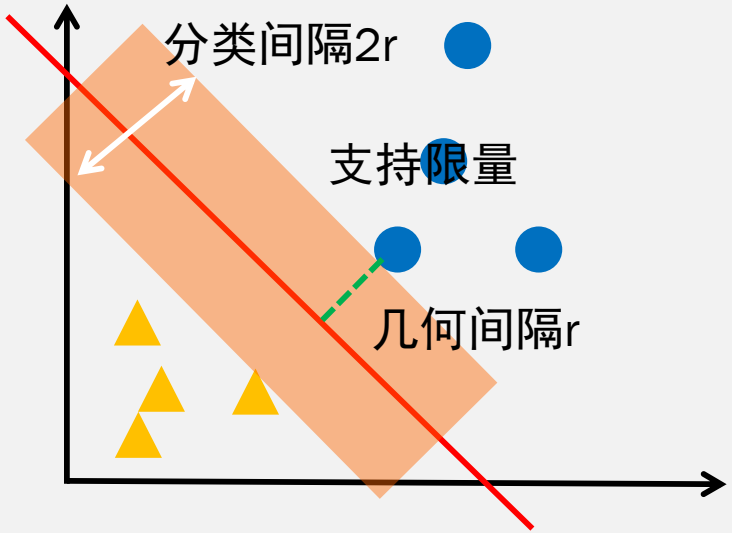
# 实践出真知：测试和应用

经过感知器与向量机的学习算法后，

我们希望知道得到分类器的效果怎么样。

于是.....我们需要测试。

学习——测试——打分

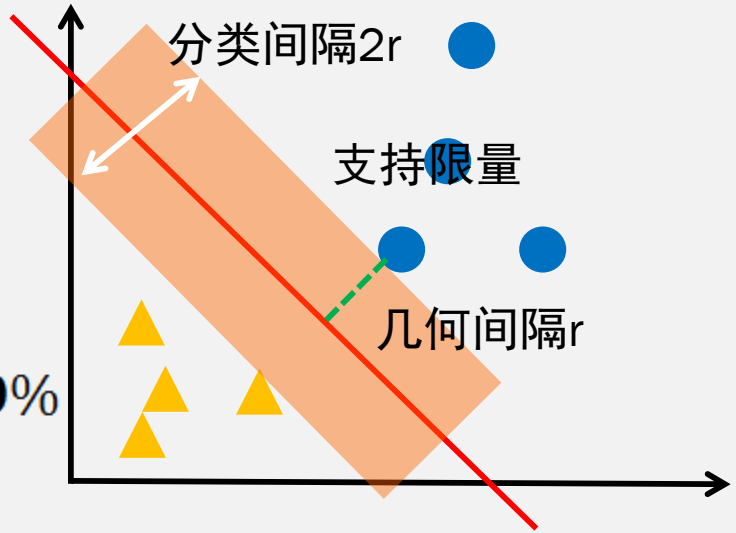


# 实践出真知：测试和应用

学习——测试——打分

训练数集　测试数集　分类准确率

分类准确率= $\frac{\text{分类正确的样本数}}{\text{测试总体样本数}} \times 100\%$




问题：测试和应用有什么区别呢？

# 二分在生活中的应用

在生活中遇到很多问题，“是不是”都属于二分类的问题范畴。

Eg：这是不是一张人脸？

这是不是肿瘤？

这是不是一处可能有矿藏的地方？

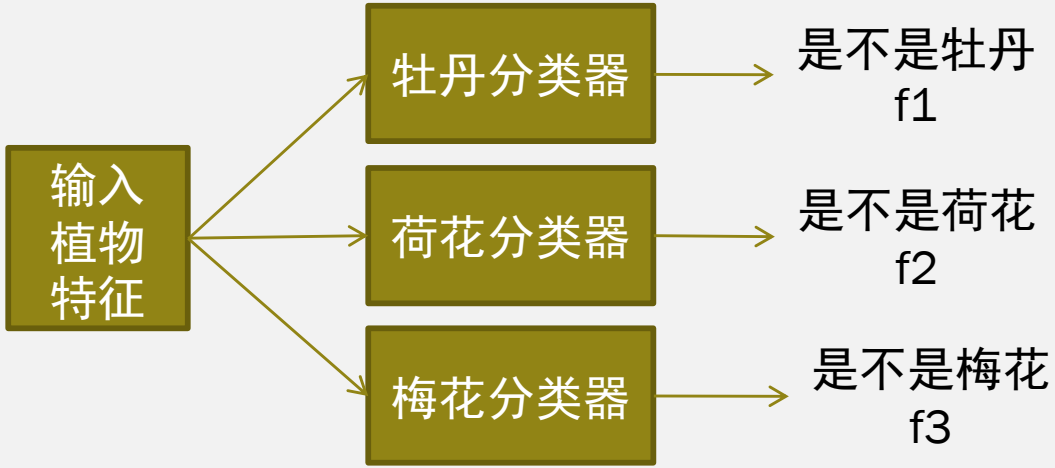
.....

你还能举出哪些二分类的问题吗？

# 多类别分类

问题：现实生活中需要进行多类别的分类。

多分转换为二分，每一个二分函数都具有自己的分类参数。



# 多类别分类

将数据分别输入三个分类其中，得到f1、f2、f3三个函数值，通过函数值得比较能够得出特征数据代表的植物属于哪个类别。

例如系统中有三个分类器，输入（x1,x2）——得到三个值，形成（压缩到）一个向量（f1,f2,f3）。

问题：怎么通过向量值清晰判断出类别呢？

# 多类别分类

归一化指数函数：将一个向量压缩到另一个向量中，使得每一个元素范围在（0,1）之间，且所有元素和为1。

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

# 多类别分类

归一化指数函数 $\sigma(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$

思考：为什么要用指数函数进行归一化？指数函数有什么特点？

Eg（0.3,0.5）—f1=-1,f2=2,f3=3—(-1,2,3)利用归一化指

| 输入z | -1 | 2 | 3 | 求和 |
| --- | --- | --- | --- | --- |
| 指数变化 $e^z$ | $e^{-1} \approx 0.368$ | $e^2 \approx 7.389$ | $e^3 \approx 20.086$ | 27.843 |
| 归一化指数函数 | $\frac{0.386}{27.843} \approx 0.013$ | $\frac{7.389}{27.843} \approx 0.265$ | $\frac{20.086}{27.843} \approx 0.722$ | 1 |

可能性1.3%　可能性26.5%　可能性72.2%