本文介绍传统机器学习算法中的“支撑向量机（SVM）”，在深度学习变得流行之前，SVM是许多论文中经常会用到的算法。它有非常完备的数学推导，我在刚接触它的时候也被搞得云里雾里。现在打算系统的介绍一下它。

本文一共分成以下几个部分

1. 线性可分支撑向量机及其对偶问题
2. 线性不可分支撑向量机及其对偶问题
3. 非线性支撑向量机（核技巧）
4. SMO算法

文章重点参考了《统计学习方法》的支撑向量机一章。

声明：文章中用到的上下标，上标的标示是括号加数字，例如$\mathbf{x}^{(i)}$表示样本集中第i个样本点，加粗的$\mathbf{x}$表示这是一个向量。下标x_i表示向量$\mathbf{x}$的第i个维度。这种标记方式是follow ng的课程的。

3. 核函数

对于数据集无法利用线性模型进行划分的，我们要使用非线性支撑向量机，通过核技巧(kernel trick)来实现从线性模型向非线性模型转变。使得在变换前需要用一个分离超曲面能够分开的数据集，在变换后，可以用一个分离超平面来分离数据。

3.1 问题描述

给定一组训练数据集

$$T = \{(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), \cdots, (\mathbf{x}^{(N)}, y^{(N)})\}$$

该数据集包含N个样本点。每个样本点的$\mathbf{x}$加粗显示，表示这是一个向量，$\mathbf{x} \in \mathbb{R}^n$，当然如果n=1，则$\mathbf{x}$是一个标量。

在$\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)})$中的每一个维度，表示该样本点的一个特征，样本集中的每个样本点有n个维度或特征。

$y^{(i)}$表示第i个样本点的类别，$y \in \{+1, -1\}$，当$y^{(i)} = 1$，则表示$x^{(i)}$是正例。当$y^{(i)} = -1$，则表示$x^{(i)}$是负例。

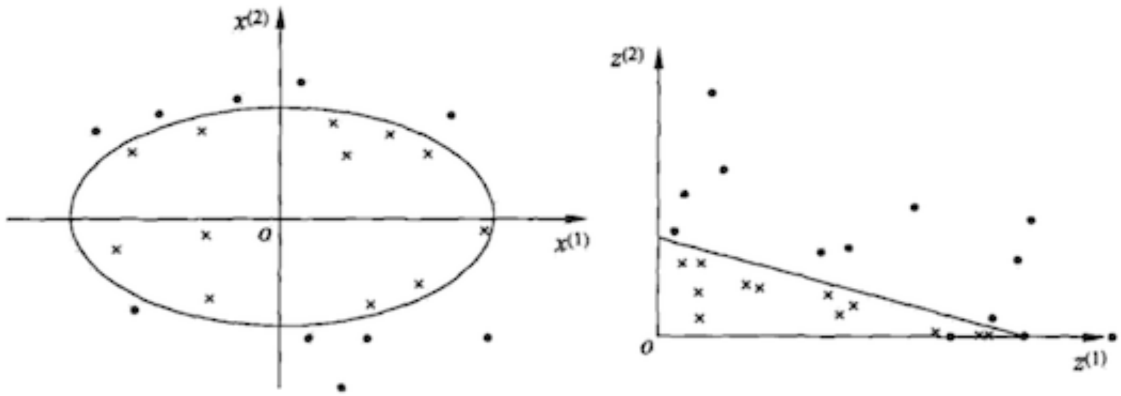

图 7.7　非线性分类问题与核技巧示例

左图表示原始空间，由二维特征(x_1, x_2)表述。图中包含一个非线性分类的数据集，无法用一个线性模型很好的划分它们。图中需要用一个椭圆来划分数据集。在原始的用(x_1, x_2)描述的空间中，椭圆的方程式为

$$w_1 x_1^2 + w_2 x_2^2 + b = 0$$

一般的做法是将左图中的原始空间，变换到右图中的映射空间。映射空间由变换后的二维特征(z_1, z_2)表述，从而使得原始空间中的每个点，在映射空间中都有一一对应。在映射空间中，o和x的数据可以被一个线性模型划分。

具体来看，原始空间为$\mathcal{X} \subset \mathbb{R}^2$，样本点$\mathbf{x} = (x_1, x_2)^T \in \mathcal{X}$。映射空间$\mathcal{Z} \subset \mathbb{R}^2$，映射后的样本点$\mathbf{z} = (z_1, z_2)^T \in \mathcal{Z}$。

定义映射函数$\phi(\mathbf{x}) : \mathbf{x} \rightarrow \mathbf{x}^2$，即$\mathbf{z} = \mathbf{x}^2 = (x_1^2, x_2^2)$。这样在原始空间中的椭圆的表达式变成由$\mathbf{z}$来描述

$$w_1 z_1 + w_2 z_2 + b = 0$$

可以看到在映射空间中的分离方程式是线性的，此时变成一个线性可分问题。

3.2 核函数

对于以上非线性可分的问题求解步骤是，先把数据集所在的原始空间，经过一次非线性变换，变换成映射空间。然后在映射空间中做线性可分的支撑向量机求解过程。回顾支撑向量机求解过程，以第一节介绍的线性可分支撑向量机为例， 分离超平面经过一番推导后可以用以下式子来表达

$$f(\mathbf{x}) = \sum_{i=1}^{N} (\alpha^{(i)})^{\star} y^{(i)} (\mathbf{x}^{(i)} \cdot \mathbf{x}) + b^{\star} = 0$$

第一项中涉及到向量内积$\mathbf{x}^{(i)} \cdot \mathbf{x}$。 假如说现在是一个非线性可分问题，按照在本节一开始的求解过程，我们现在要求的是一个分离超曲面的函数，要把数据集转变到映射空间中，那么假设映射函数为$\phi(\mathbf{x})$，则分离函数可以表达成

$$f(\mathbf{x}) = \sum_{i=1}^{N} (\alpha^{(i)})^{\star} y^{(i)} (\phi(\mathbf{x}^{(i)}) \cdot \phi(\mathbf{x})) + b^{\star} = 0$$

但是上面的求解过程，在实际的工程实现上，存在效率的问题。接下来定义核函数。

$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \phi(\mathbf{x}^{(i)}) \cdot \phi(\mathbf{x}^{(j)})$$

称$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$为核函数。

除了存在计算效率的问题，有些情况是只使用映射函数$\phi(\mathbf{x})$无法解决的，例如在映射过后要求数据处于一个无穷多维度的空间，这种情况无法显示的写出映射函数$\phi(\mathbf{x})$来。后面会具体提到这种情况。

有了核函数的定义，则支撑向量机的分离函数可以写成

$$f(\mathbf{x}) = \sum_{i=1}^{N} (\alpha^{(i)})^{\star} y^{(i)} K(\mathbf{x}^{(i)}, \mathbf{x}) + b^{\star} = 0$$

$b^{\star}$可以写成

$$b^{\star} = y^{(j)} - \sum_{i=1}^{N} (\alpha^{(i)})^{\star} y^{(i)} K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$$

对于有间隔松弛的场景，选定适当的参数C，套用上述的表达式即可。

关于核函数的性质，下面应用《统计学习方法》上的一个例题描述。 对于一个给定的核函数$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$，其表达的映射函数$\phi(\mathbf{x})$和映射后的空间不是唯一的。用以下例子说明。

假设原始空间是二维的$\mathbb{R}^2$，核函数$K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z})^2$。那么有以下的$\phi(\cdot)$和映射空间对。

因为原始空间是二维的，所以把输入的$\mathbf{x}$和$\mathbf{z}$分别记做$(x_1, x_2)^T$和$(z_1, z_2)^T$。（加向量转置是因为数学上习惯用列向量表达）

1. 取映射空间是三维的$\mathbb{R}^3$，因为

   $$K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z})^2 = (x_1 z_1 + x_2 z_2)^2 = (x_1 z_1)^2 + 2x_1 z_1 x_2 z_2 + (x_2 z_2)^2$$

   所以此时可以取映射函数$\phi(\mathbf{x}) = \phi(x_1, x_2) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2)^T$。利用该映射函数可以验证

   $$\phi(x_1, x_2) \cdot \phi(z_1, z_2) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2) \cdot (z_1^2, \sqrt{2} z_1 z_2, z_2^2) = x_1^2 z_1^2 + 2x_1 x_2 z_1 z_2 + x_2^2 z_2^2 = K(\mathbf{x}, \mathbf{z})$$

2. 取映射空间是三维的$\mathbb{R}^3$，同时取映射函数$\phi(\mathbf{x}) = \phi(x_1, x_2) = \frac{1}{\sqrt{2}}(x_1^2 - x_2^2, 2x_1 x_2, x_1^2 + x_2^2)^T$。利用该映射函数可以验证

   $$\begin{aligned}\phi(x_1, x_2) \cdot \phi(z_1, z_2) &= \frac{1}{\sqrt{2}}(x_1^2 - x_2^2, 2x_1 x_2, x_1^2 + x_2^2) \cdot \frac{1}{\sqrt{2}}(z_1^2 - z_2^2, 2z_1 z_2, z_1^2 + z_2^2) \\ &= \frac{1}{2}(x_1^2 z_1^2 - x_1^2 z_2^2 - x_2^2 z_1^2 + x_2^2 z_2^2 + x_1^2 z_1^2 + x_1^2 z_2^2 + x_2^2 z_1^2 + x_2^2 z_2^2 + 4x_1 x_2 z_1 z_2) \\ &= x_1^2 z_1^2 + 2x_1 x_2 z_1 z_2 + x_2^2 z_2^2 \\ &= K(\mathbf{x}, \mathbf{z})\end{aligned}$$

3. 取映射空间是四维的$\mathbb{R}^4$，同时取映射函数$\phi(\mathbf{x}) = \phi(x_1, x_2) = (x_1^2, x_1 x_2, x_1 x_2, x_2^2)^T$，利用该映射函数可以验证

$$
\begin{aligned}
\phi(x_1,x_2)\cdot\phi(z_1,z_2) &= (x_1^2,x_1x_2,x_1x_2,x_2^2)\cdot(z_1^2,z_1z_2,z_1z_2,z_2^2)\\
&= x_1^2z_1^2+x_1x_2z_1z_2+x_1x_2z_1z_2+x_2^2z_2^2\\
&= x_1^2z_1^2+2x_1x_2z_1z_2+x_2^2z_2^2\\
&= K(\mathbf{x},\mathbf{z})
\end{aligned}
$$

3.3 核函数的种类

常见的核函数种类有多项式核函数、高斯核函数、字符串核函数等

1. 多项式核函数

   $$K(\mathbf{x},\mathbf{z})=(\mathbf{x}\cdot\mathbf{z}+1)^p$$

   当$p=1$时，也称为线性核函数

2. 高斯核函数

   $$K(\mathbf{x},\mathbf{z})=\exp\left(-\frac{\|\mathbf{x}-\mathbf{z}\|}{2\sigma^2}\right)$$

   此时的支撑向量机是高斯径向基函数(radial basis function)分类器。

字符串核函数不作为本文的重点，具体可以参考《统计学习方法》第122页