分类：SVM（原问题加上核函数技巧与梯度下降算法）

覃雄派

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 准备工作：向量的点积

Basic operation on vectors in $\mathbb{R}^n$

5. Dot product

Consider vectors $\vec{a}=(a_1,a_2,\ldots,a_n)$ and $\vec{b}=(b_1,b_2,\ldots,b_n)$

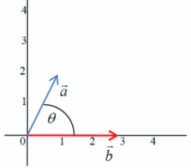
Define dot product: $\vec{a}\cdot\vec{b}=a_1b_1+a_2b_2+\ldots+a_nb_n=\sum_{i=1}^{n}a_ib_i$

The law of cosines says that $\vec{a}\cdot\vec{b}=\|\vec{a}\|_2\|\vec{b}\|_2\cos\theta$ where θ is the angle between $\vec{a}$ and $\vec{b}$. Therefore, when the vectors are perpendicular $\vec{a}\cdot\vec{b}=0$.

$\vec{a}=(1,2)$

$\vec{b}=(3,0)$

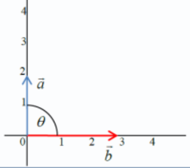
$\vec{a}\cdot\vec{b}=3$



$\vec{a}=(0,2)$

$\vec{b}=(3,0)$

$\vec{a}\cdot\vec{b}=0$



$\vec{a}\cdot\vec{b}=a_1b_1+a_2b_2+\ldots+a_nb_n=\sum_{i=1}^{n}a_ib_i$

- Property: $\vec{a}\cdot\vec{a}=a_1a_1+a_2a_2+\ldots+a_na_n=\|\vec{a}\|_2^2$
- In the classical regression equation $y=\vec{w}\cdot\vec{x}+b$ the response variable *y* is just a dot product of the vector representing patient characteristics ($\vec{x}$) and the regression weights vector ($\vec{w}$) which is common across all patients plus an offset *b*.

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 准备工作：向量的第二范数

Basic operation on vectors in $\mathbb{R}^n$

4. Euclidian length or L2-norm

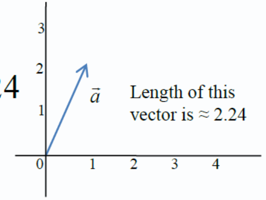
Consider a vector $\vec{a} = (a_1, a_2, \ldots, a_n)$

Define the L2-norm: $\|\vec{a}\|_2 = \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2}$

We often denote the L2-norm without subscript, i.e. $\|\vec{a}\|$

$\vec{a} = (1,2)$

$\|\vec{a}\|_2 = \sqrt{5} \approx 2.24$



L2-norm is a typical way to measure length of a vector; other methods to measure length also exist.

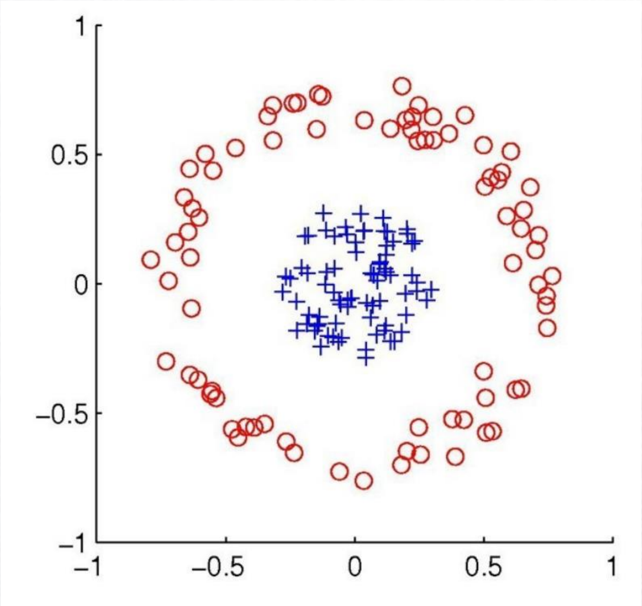
提纲

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 线性不可分问题与核函数
- 核函数及其优势
- 数据升维并线性可分实践
- 常用核函数
    - 高斯核函数的可视化
    - 理解高斯核函数
- 非线性可分数据的SVM原问题求解
    - 原理
    - 实践

分类：SVM（原问题加上核函数技巧与梯度下降算法）

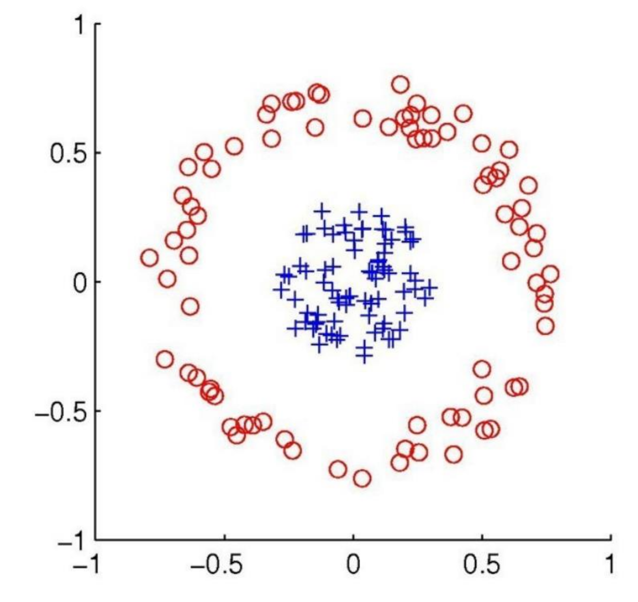
- 线性不可分问题与核函数
  - 假设平面上有两类数据点
    - 红色/蓝色
  - 我们无法通过一条直线
    - 把两类数据点分开
  - 这是一个线性不可分的问题



分类：SVM（原问题加上核函数技巧与梯度下降算法）

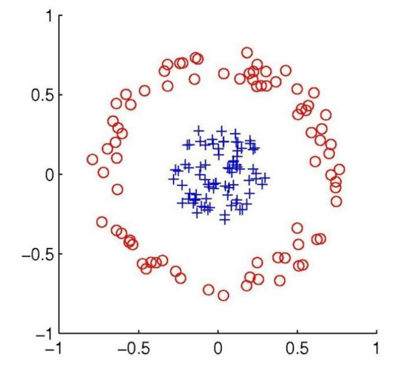
- 线性不可分问题与核函数
- 假设一个样本x（二维）
  - 记为$\mathbf{x}$={ x_{i1}, x_{i2} }
- 给出一个如下变换ϕ如下
  - 把二维样本转换为三维样本

$$\Phi(\mathbf{x}_i) = \{x_{i1}, \underbrace{x_{i2}, \sqrt{2}x_{i1}x_{i2}}_{\text{三维}}\}$$



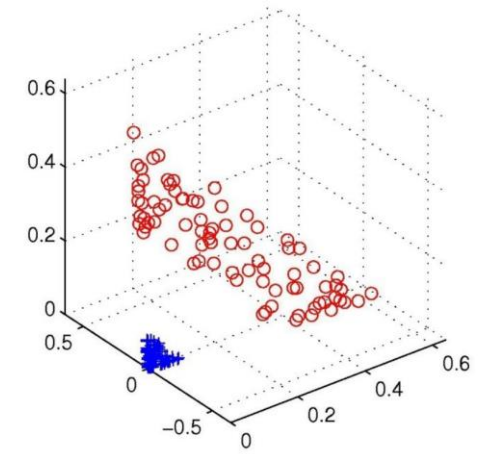
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 线性不可分问题与核函数
- 假设一个样本x（二维）
  - 记为**x**={ x_{i1}, x_{i2} }
- 给出一个如下变换ϕ如下
  - 把二维样本转换为三维样本



$$\Phi(\mathbf{x}_i) = \{x_{i1}, x_{i2}, \sqrt{2}x_{i1}x_{i2}\}$$

- 对转换后的数据进行可视化
  - 结果如右图所示



可以看到，在3维空间，数据是线性可分的，或者基本线性可分的

分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 为什么点积很重要
- SVM的原问题可以转换成对偶问题
  - 具体形式如下（在此不展开转换过程、以及对偶问题求解方法）
    - 该问题可以用二次规划方法求解
    - 可以看到样本点之间的点积是非常重要的操作

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{\mathbf{x}_i^T \cdot \mathbf{x}_j}$$

$$\text{s.t.} \sum_i \alpha_i y_i = 0$$

SVM optimization problem:
Benefits of using dual formulation

1) No need to access original data, need to access only dot products.

Objective function: $\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \boxed{\vec{x}_i \cdot \vec{x}_j}$

Solution: $f(\vec{x}) = sign(\sum_{i=1}^{N} \alpha_i y_i \boxed{\vec{x}_i \cdot \vec{x}} + b)$

2) Number of free parameters is bounded by the number of support vectors and not by the number of variables (beneficial for high-dimensional problems).

E.g., if a microarray dataset contains 20,000 genes and 100 patients, then need to find only up to 100 parameters!

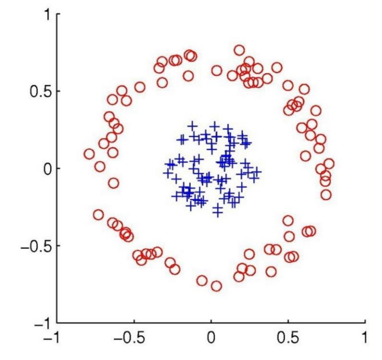
对偶问题有成熟解法SMO算法

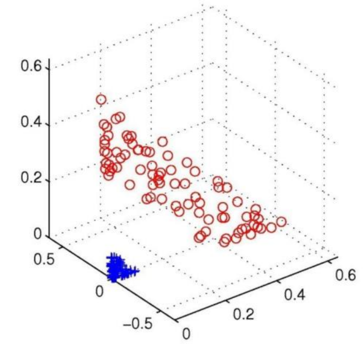
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 我们可以
  - 1.把所有样本，利用某种函数Φ映射到高维空间，使其基本线性可分
  - 2.在高维空间，构建如下优化问题，进行求解

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \ \Phi(\ \mathbf{x}_i\ )^T \cdot \ \Phi(\ \mathbf{x}_j\ )$$

$$\text{s.t.} \sum_i \alpha_i y_i = 0$$





分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 核函数及其优势
- 假设有如下转换函数

$$\Phi(\mathbf{x}_i) = \{x_{i1}, x_{i2}, \sqrt{2}x_{i1}x_{i2}\}$$

- 为了计算两个样本升维以后的点积，有两种办法
  - 办法1
  - 1.每个样本先升维
  - 2.再进行点积

$$\langle \Phi(\mathbf{x}i), \Phi(\mathbf{x}j) \rangle$$

$$\begin{aligned}\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle &= \langle \{x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}\}, \{x_{j1}^2, x_{j2}^2, \sqrt{2}x_{j1}x_{j2}\} \rangle \\ &= x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 核函数及其优势
- 假设有如下转换函数

$$\Phi(\mathbf{x}_i) = \{x_{i1}, x_{i2}, \sqrt{2}x_{i1}x_{i2}\}$$

- 为了计算两个样本升维以后的点积，有两种办法
  - 办法2
  - 1.两个样本点积
  - 2.调用核函数进行转换

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

称为核函数

$$\begin{aligned}\langle \mathbf{x}_i, \mathbf{x}_j \rangle^2 &= \langle \{x_{i1}, x_{i2}\}, \{x_{j1}, x_{j2}\} \rangle^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\ &= \boxed{x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}}\end{aligned}$$

两个方法的结果是相等的

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 比较两者的运算量

$$\Phi(\mathbf{x}_i) = \{x_{i1}, x_{i2}, \sqrt{2}x_{i1}x_{i2}\}$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

$$\begin{aligned}\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle &= \boxed{\langle \{x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}\}, \{x_{j1}^2, x_{j2}^2, \sqrt{2}x_{j1}x_{j2}\} \rangle} \\ &= \boxed{x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}}\end{aligned}$$

$$\begin{aligned}\langle \mathbf{x}_i, \mathbf{x}_j \rangle^2 &= \langle \{x_{i1}, x_{i2}\}, \{x_{j1}, x_{j2}\} \rangle^2 \\ &\boxed{= (x_{i1}x_{j1} + x_{i2}x_{j2})^2} \\ &= x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

1.两个平方+一个乘法
2.两个平方+一个乘法
3.三个乘法+两个加法

共11个操作

1.两个乘法
2.一个加法
3.一个平方

共4个操作

节省计算开销，达到同样效果

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 比较两者的运算量

$$\Phi(\mathbf{x}_i) = \{x_{i1}, x_{i2}, \sqrt{2}x_{i1}x_{i2}\}$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

$$\begin{aligned}\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle &= \langle \{x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}\}, \{x_{j1}^2, x_{j2}^2, \sqrt{2}x_{j1}x_{j2}\} \rangle \\ &= x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

$$\begin{aligned}\langle \mathbf{x}_i, \mathbf{x}_j \rangle^2 &= \langle \{x_{i1}, x_{i2}\}, \{x_{j1}, x_{j2}\} \rangle^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\ &= x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

- 高维问题的求解，无须对向量进行升维

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{\Phi(\mathbf{x}_i)^T \cdot \Phi(\mathbf{x}_j)}$$

$$\text{s.t.} \sum_i \alpha_i y_i = 0$$

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{K(\mathbf{x}_i, \mathbf{x}_j)}$$

$$\text{s.t.} \sum_i \alpha_i y_i = 0$$

节省计算开销，达到同样效果

甚至无需构造转换函数ϕ

分类：SVM（原问题加上核函数技巧与梯度下降算法）

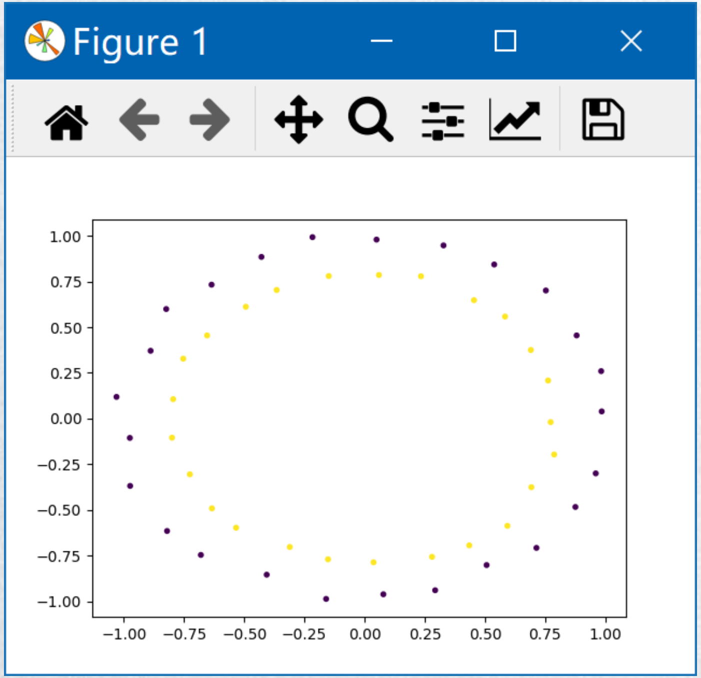
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 数据升维与线性可分实践

名称	类型	大小	修改日期
03kernel_2d_3d(2).py	Python File	2 KB	2021/10/31 12:35

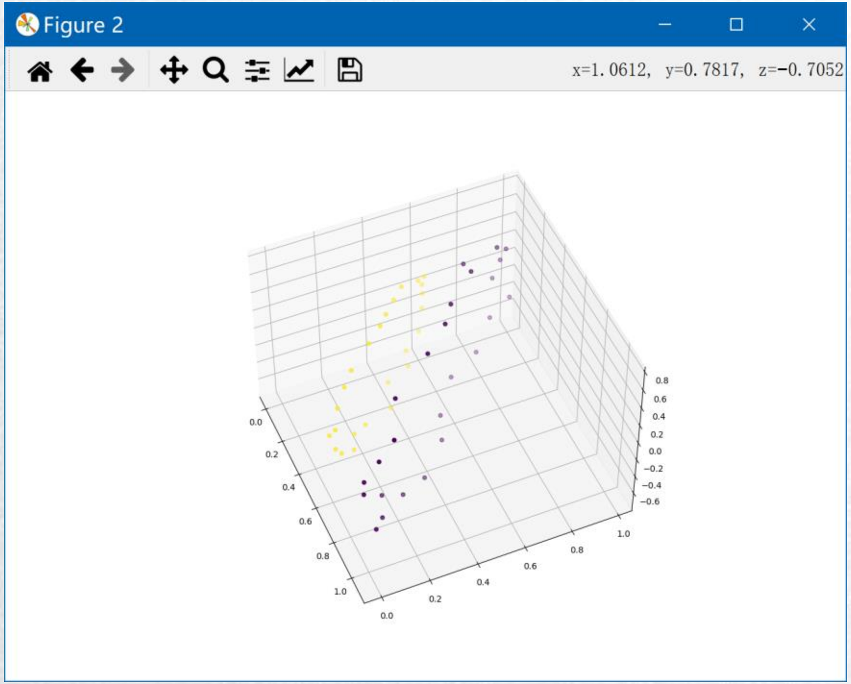
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 数据升维与线性可分实践
  - 2维空间数据点



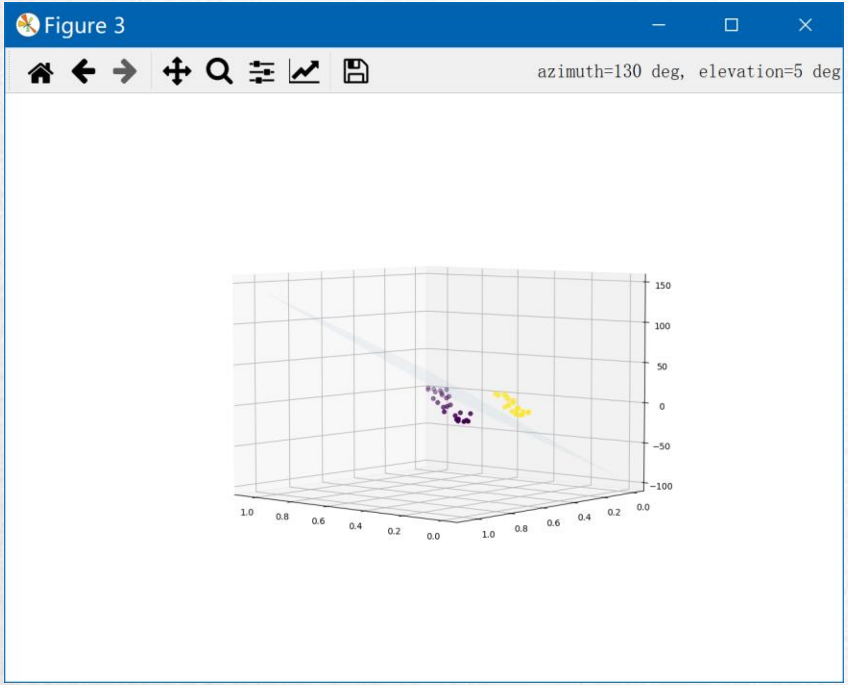
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 数据升维与线性可分实践
  - 2维空间数据点
  - 升维到3维空间



分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 数据升维与线性可分实践
  - 2维空间数据点
  - 升维到3维空间
  - 线性可分和分类超平面
  - （拖动鼠标转动才能看清楚）



分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 常用的核函数
  - 如右图所示
    - 核函数一般用K表示，K为Kernel的缩写
  - 注意和前文所述转换函数Φ不是一回事
  - （后文继续）
  - 部分核函数的具体形式列举如下

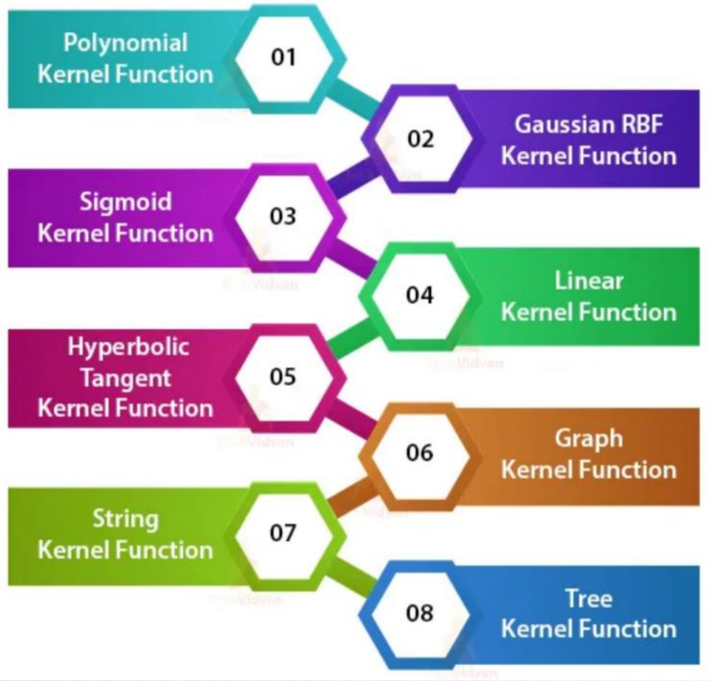
Popular kernels

A kernel is a dot product in *some* feature space:

$$K(\vec{x}_i,\vec{x}_j)=\Phi(\vec{x}_i)\cdot\Phi(\vec{x}_j)$$

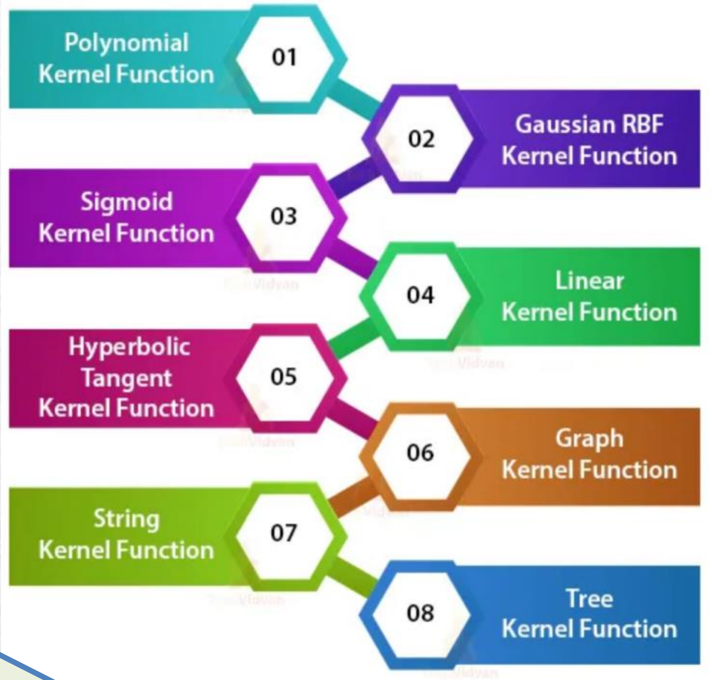
<u>Examples:</u>

$K(\vec{x}_i,\vec{x}_j)=\vec{x}_i\cdot\vec{x}_j$	Linear kernel
$K(\vec{x}_i,\vec{x}_j)=\exp(-\gamma\lVert\vec{x}_i-\vec{x}_j\rVert^2)$	Gaussian kernel
$K(\vec{x}_i,\vec{x}_j)=\exp(-\gamma\lVert\vec{x}_i-\vec{x}_j\rVert)$	Exponential kernel
$K(\vec{x}_i,\vec{x}_j)=(p+\vec{x}_i\cdot\vec{x}_j)^q$	Polynomial kernel
$K(\vec{x}_i,\vec{x}_j)=(p+\vec{x}_i\cdot\vec{x}_j)^q\exp(-\gamma\lVert\vec{x}_i-\vec{x}_j\rVert^2)$	Hybrid kernel
$K(\vec{x}_i,\vec{x}_j)=\tanh(k\vec{x}_i\cdot\vec{x}_j-\delta)$	Sigmoidal



分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 常用的核函数
- 核函数都具有这样的性质



- 两个向量$\mathbf{x}_i$和 $\mathbf{x}_j$

$$\Phi(\mathbf{x}_i) = \{x_{i1}, x_{i2}, \sqrt{2}x_{i1}x_{i2}\}$$

  - 经过转换函数升维以后的点积

$$\begin{aligned}\langle\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)\rangle &= \langle\{x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}\}, \{x_{j1}^2, x_{j2}^2, \sqrt{2}x_{j1}x_{j2}\}\rangle \\ &= x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

  - 等于
  - 两个向量$\mathbf{x}_i$和 $\mathbf{x}_j$的点积代入核函数的结果

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle\mathbf{x}_i, \mathbf{x}_j\rangle^2$$

核函数的作用是平方

$$\begin{aligned}\langle\mathbf{x}_i, \mathbf{x}_j\rangle^2 &= \langle\{x_{i1}, x_{i2}\}, \{x_{j1}, x_{j2}\}\rangle^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\ &= x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 高斯核函数（Radial Basis Function，径向基核函数）
  - 具体形式如下

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

  - 计算过程是两个数据点的距离的平方
  - 除以$2\sigma^2$
  - 取负数
  - 计算自然数的幂

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 高斯核函数（Radial Basis Function，径向基核函数）
  - 可视化效果
  - 变形(gamma)

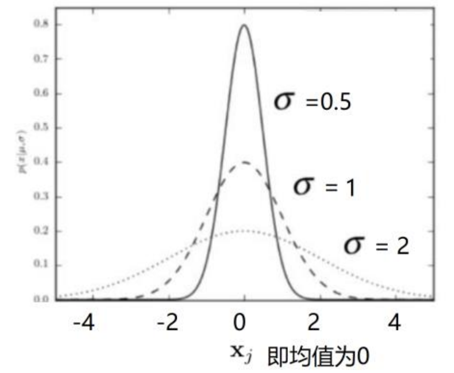
$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}\right) = \exp\left(-\gamma\, ||\mathbf{x}_i - \mathbf{x}_j||^2\right)$$

  - 一维高斯函数

$$g(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2\sigma^2}}$$

  - 可视化

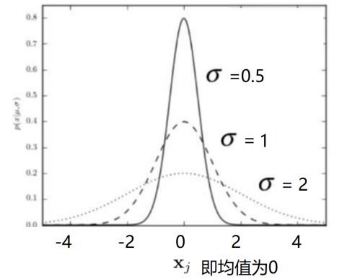
1维高斯函数
$\mathbf{x}_j$ =0



总结：
- σ越大，γ越小，分布越宽
- σ越小，γ越大，分布越窄

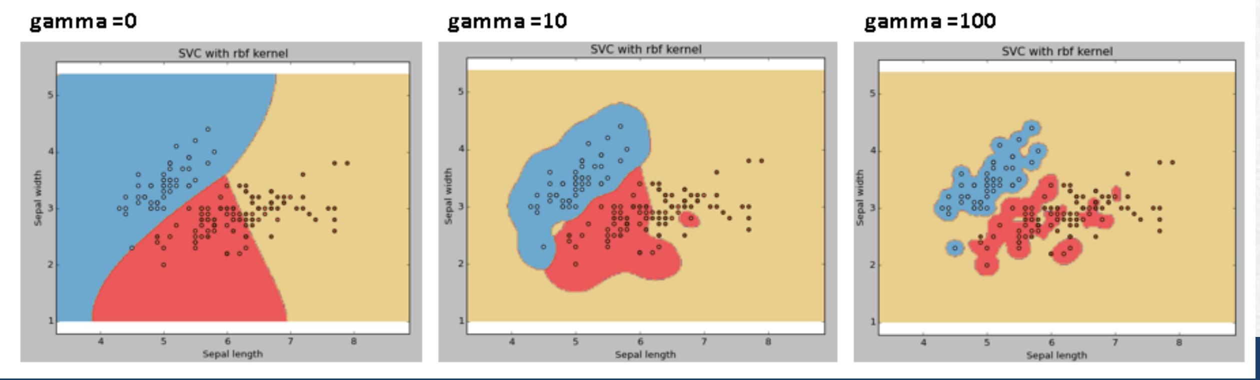
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 高斯核函数（Radial Basis Function，径向基核函数）
  - 可视化效果
  - 一维高斯函数可视化



总结：
- σ越大，γ越小，分布越宽
- σ越小，γ越大，分布越窄

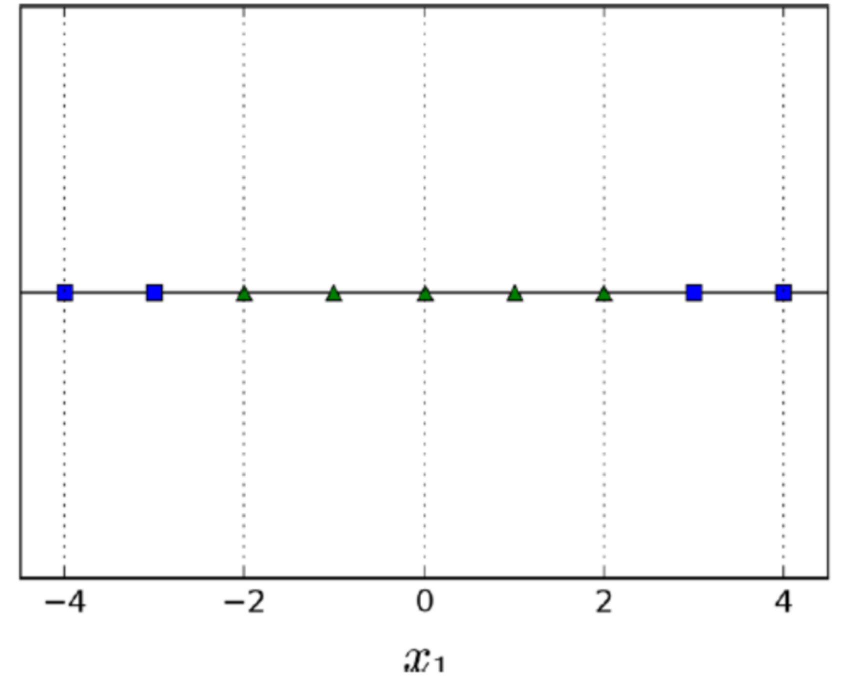
  - 不同大小的gamma值
  - 的分类效果如下图



分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 理解高斯核函数
  - 假设有如下的一维数据
  - 数据共有两类
    - 三角形一类
    - 正方形一类
    - 特征就一个x1
  - 可见，它们不是线性可分的



分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 理解高斯核函数
- 简单起见
  - 以-2和1这两个数据点做Landmark，即l1和l2
  - 创建核函数
    - l表示landmark
    - 核函数有两个
    - 图像如绿色和蓝色虚线所示

$$\exp\left(-\gamma \| \mathbf{x} - \ell \|^2\right)$$

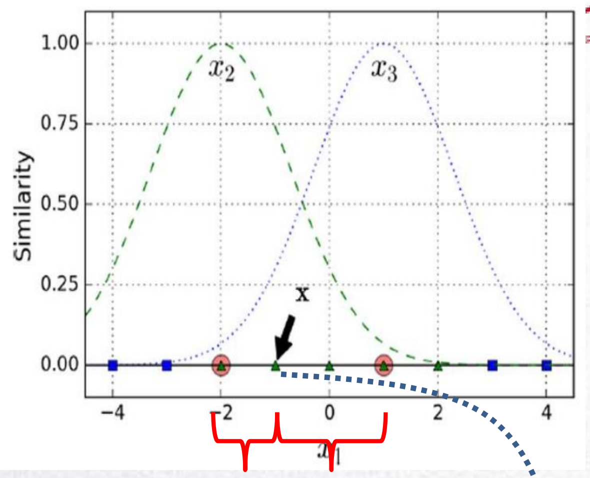
$$\gamma = 0.3$$

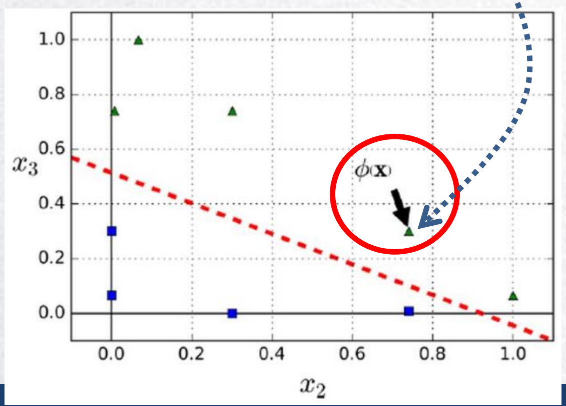
  - 把-1这个数据点代入核函数
    - 注意有两个landmark
    - 有两个核函数
    - 得到x2和x3

$$x_2 = \exp\left(-0.3 \times 1^2\right) \approx 0.74$$

$$x_3 = \exp\left(-0.3 \times 2^2\right) \approx 0.30$$

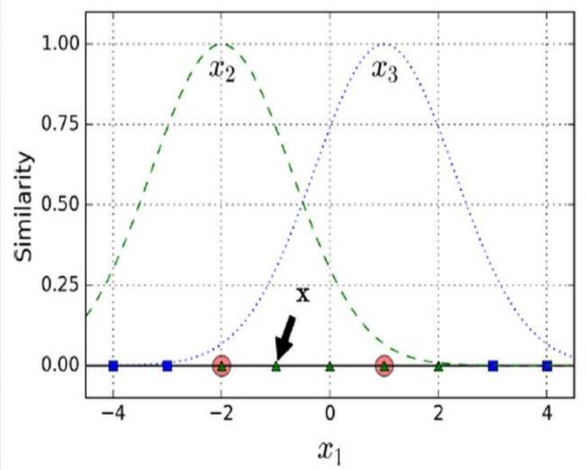
    - {x2, x3}作为-1这个一维数据点的
    - 升维到二维的坐标系，绘制如右下图所示

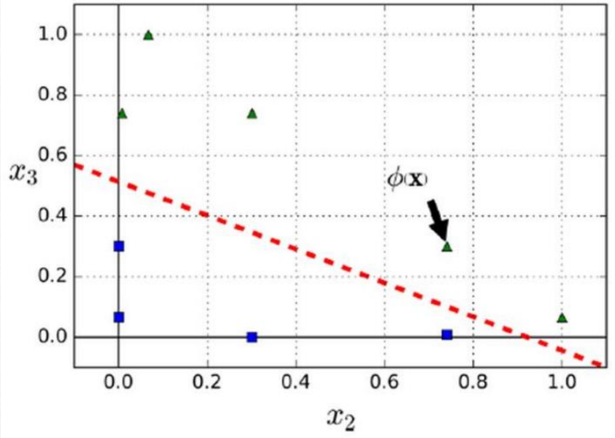




分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 理解高斯核函数
  - 所有的一维数据点
    - 都利用同样的方法来进行升维

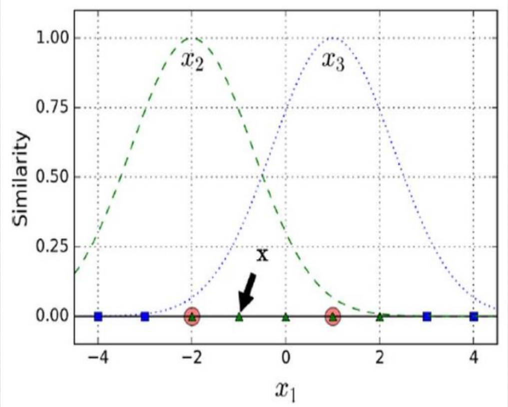


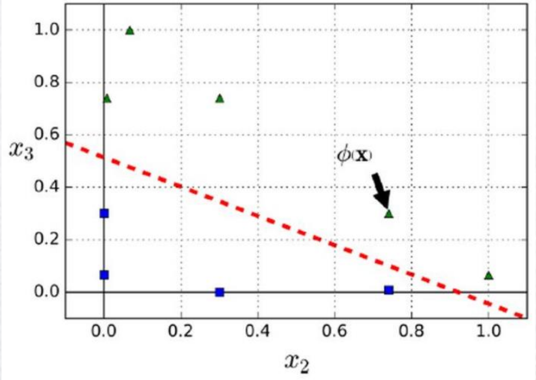


- 低维空间线性不可分的问题
- 转换为高维**线性可分**的问题

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 理解高斯核函数
  - 所有的一维数据点
    - 都利用同样的方法来进行升维





- 本实例，仅仅用了两个样本来做landmark（注意有几个landmark映射后就有几个维度）
- 在实际应用中，最简单的选取landmark的方法是选取样本个数，假设样本有N个，那么有N个landmark，我们映射后的维度就是N维
- 新建了更多的维度，使得数据被线性可分的机会大大增加

分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

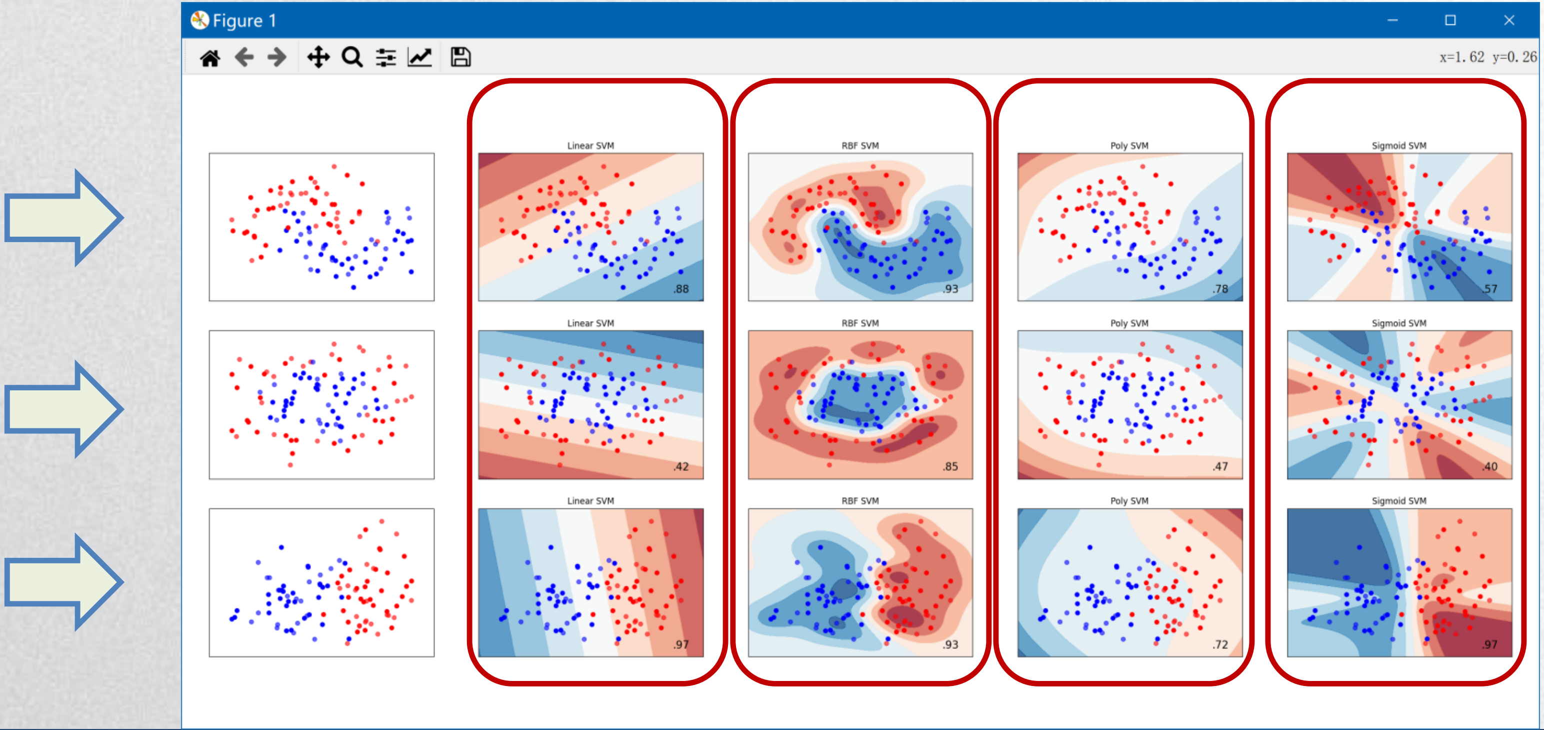
- 不同核函数效果比较（不同数据集）

Application (D:) › 2021-07-18《数据科学概论》new plan › 2022newPPT › 0303-分类：SVM（原问题加上核函数技巧与梯度下降算法） 搜索"030

名称	类型	大小	修改日期
01compare-svm-kernels.py	Python File	4 KB	2021/11/10 12:50

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 不同核函数效果比较（不同数据集）



分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 非线性可分数据的SVM原问题求解
  - 回顾软间隔问题(线性可分或者基本线性可分问题)的目标函数和约束条件如下（请参考上一个PPT）：
  - min($\frac{1}{2}||w|| + C\sum_{i=1}^{n}\xi_i$)
  - s.t. $y_i(w^T.x_i + b) \geq 1 - \xi_i$, i=1,2,...,n
  - $\xi_i \geq 0$, i=1,2,...,n

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 非线性可分数据的SVM原问题求解
  - 回顾当我们采用Hinge Loss的时候，问题转化为如下优化问题（请参考上一个PPT）：
  - $min \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \max(0, 1 - y_i f(x_i))$
  - s.t. $y_i(w^T . x_i + b) \geq 1 - \xi_i$, i=1,2,…,n
  - $\xi_i \geq 0$, i=1,2,…,n

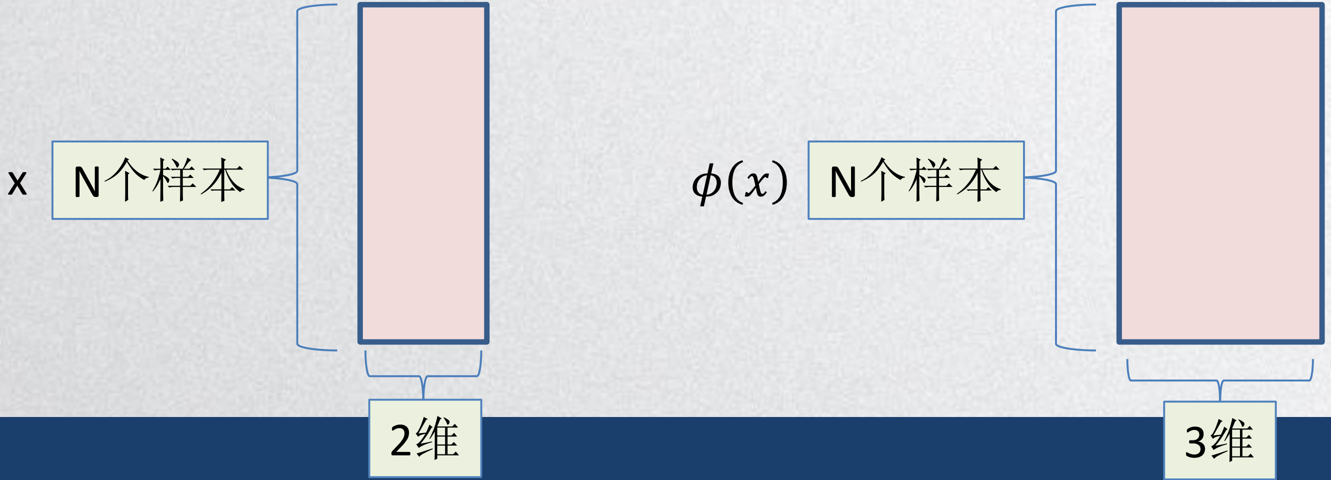
分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 转换函数ϕ
  - 假设现在已经把数据转换到高维，即有低维空间的$x_1,x_2,\ldots,x_n$等样本，对这些样本升维以后为$\phi(x_1),\phi(x_2),\ldots,\phi(x_n)$，注意n为样本个数
  - 把高维的各个特征的系数记为β
    - β表达为$\phi(x_1),\phi(x_2),\ldots,\phi(x_n)$的线性组合，即有$\beta=\sum_{i=1}^{n} w_i\,\phi(x_i)$
    - 注意，这里的w_i不是软间隔问题的建模里面的w的各个分量，而是由$\phi(x_1),\phi(x_2),\ldots,\phi(x_n)$构成$\beta$的权重，共有n个
    - 在不至于引起混淆的情况下，姑且这么使用

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 转换函数φ
  - 假设低维空间为2维，高维空间为3维
  - $x_1,x_2,\ldots,x_n$都是1*2向量，那么$\phi(x_1)$为1*3向量，$\phi(x_1)^T$为3*1向量
  - x为n*2向量，$\phi(x)$为n*3向量，$\phi(x)^T$为3*n向量



分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 已知有$\beta = \phi(\mathrm{x})^T \mathrm{w} = (\phi(\mathrm{x}_1)^T, \phi(\mathrm{x}_2)^T, \ldots, \phi(\mathrm{x}_n)^T)\begin{pmatrix} w_1 \\ w_2 \\ \ldots \\ w_n \end{pmatrix}$
  - 此时，高维空间的目标函数为
  - $\mathrm{L} = \frac{1}{2}\beta^T \beta + C\sum_{i=1}^{n} \max(0, 1 - y_i(\beta^T \phi(\mathrm{x}_i) + \mathrm{b})$

$min\frac{1}{2}\,||w||^2 + C\sum_{i=1}^{N} \max(0,1 - y_i f(x_i))$
s.t. $y_i\left(w^T.x_i + b\right) \geq 1 - \xi_i$, i=1,2,…,n
$\xi_i \geq 0$, i=1,2,…,n

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 把$\beta = \phi(x)^T w$代入，对其进行简化如下
  - $L = \frac{1}{2}(\phi(x)^T w)^T \phi(x)^T w + C\sum_{i=1}^{n} \max(0, 1 - y_i((\phi(x)^T w)^T \phi(x_i) + b)$
  - $= \frac{1}{2} w^T \phi(x) \phi(x)^T w + C\sum_{i=1}^{n} \max(0, 1 - y_i(w^T \phi(x)\phi(x_i) + b)$
  - $= \frac{1}{2} w^T K(x, x^T) w + C\sum_{i=1}^{n} \max(0, 1 - y_i(w^T K(x, x_i) + b)$

- 上述式子各个成分的维度为
- (1*n) K(n*2*2*n) (n*1) + max((1*1), (1*1)-(1*1)(1*n)K(n*2*2*1) +(1*1))
- 这里的2指的是(上文)假设低维空间的维度为2

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 把β= $\phi(x)^T w$代入，对其进行简化如下
  - $\frac{1}{2} w^T K(x, x^T) w + C \sum_{i=1}^{n} \max(0, 1 - y_i (w^T K(x, x_i) + b))$
  - 为了寻找最小值，需要针对参数w和b求偏导数
  - 然后代入梯度下降公式，不断进行迭代即可
  - $\frac{\partial L}{\partial w} = K(x, x^T) w - C \sum_{i=1, \xi_i \geq 0}^{n} y_i K(x, x_i)$

$\xi_i = 0$ 表示支持向量，
$\xi_i > 0$表示分类错误

各个成分的维度为K(n*2*2*n) (n*1)–(1*1)K(n*2*2*1)

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 把β= $\phi(x)^T w$代入，对其进行简化如下
  - $\frac{1}{2} w^T K(x, x^T) w + C \sum_{i=1}^{n} \max(0, 1 - y_i(w^T K(x, x_i) + b)$
  - 为了寻找最小值，需要针对参数w和b求偏导数
  - 然后代入梯度下降公式，不断进行迭代即可。
  - $\frac{\partial L}{\partial b} = -C \sum_{i=1, \xi_i \geq 0}^{n} y_i$
  - 各个成分的维度为(1*1)

$\xi_i = 0$ 表示支持向量，
$\xi_i > 0$表示分类错误

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 通过梯度下降算法迭代求解，求得w和b参数
  - $W = w - \eta \frac{\partial L}{\partial w}$
  - $b = b - \eta \frac{\partial L}{\partial b}$
  - η为学习率

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- 对于非线性可分的SVM问题，是不能直接用上述模型求解的，需要对数据进行升维
  - 最后，可以使用如下式子，预测样本z的类别
  - $\hat{y} = sign\left(\beta^T \phi(Z) + b\right)$
  - $= sign\left((\phi(x)^T w)^T \phi(Z) + b\right)$
  - $= sign\left(w^T \phi(x)\phi(Z) + b\right)$
  - $= sign\left(w^T K(x, z) + b\right)$
  - 各个成分的维度为(1*n) K(n*2)(2*1) +(1*1)

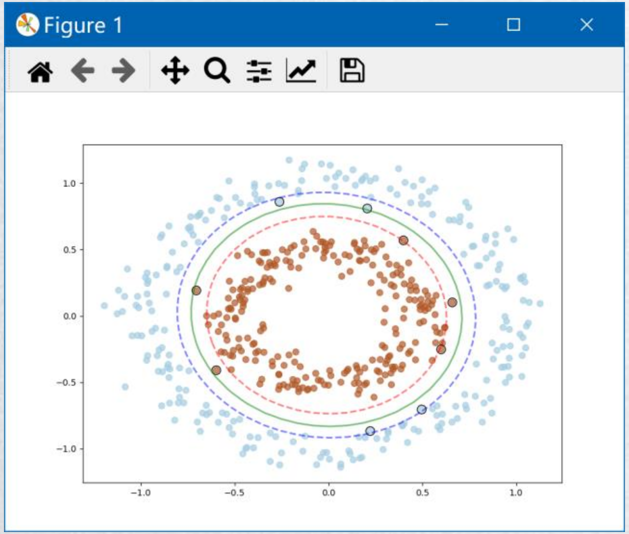
分类：SVM（原问题加上核函数技巧与梯度下降算法）

分类：SVM（原问题加上核函数技巧与梯度下降算法）

- SVM （原问题加上核函数技巧与梯度下降算法）
  - 实践（代码）

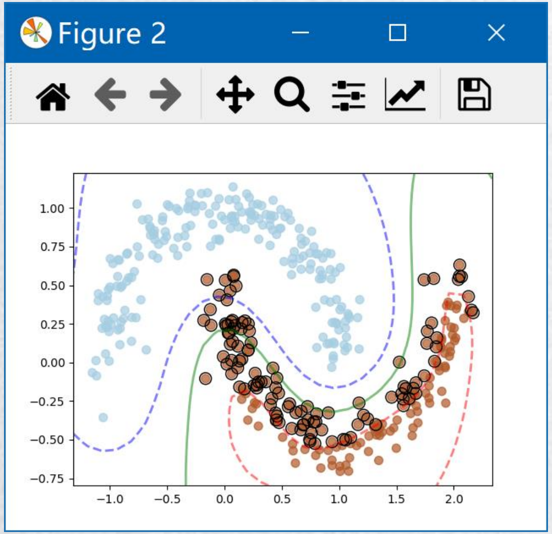
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- SVM （原问题加上核函数技巧与梯度下降算法）
  - 实践（数据集1）



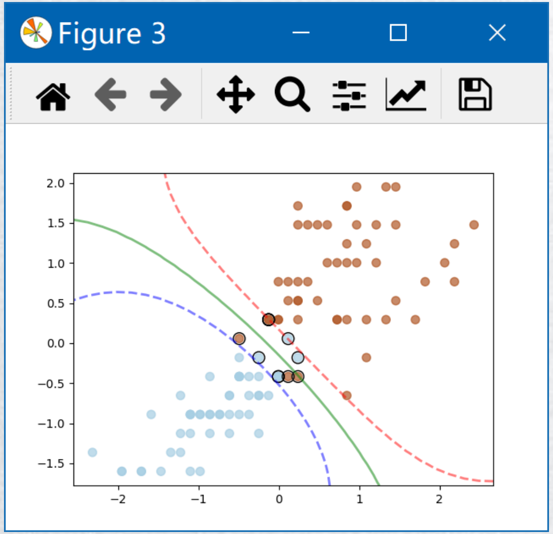
分类：SVM（原问题加上核函数技巧与梯度下降算法）

- SVM （原问题加上核函数技巧与梯度下降算法）
  - 实践（数据集2）



分类：SVM（原问题加上核函数技巧与梯度下降算法）

- SVM （原问题加上核函数技巧与梯度下降算法）
  - 实践（数据集3）



分类：SVM（原问题加上核函数技巧与梯度下降算法）