北京交通大学硕士研究生《机器学习》课件

第16章 多类数据升维：核方法

一花一世界，一叶一如来。

——《益州嵩山野竹禅师后录·卷二》

北京交通大学《机器学习》课程组

如果复杂模型比简单模型有效，就选择复杂的。

合适的特征空间不一定是原空间

更不一定是原空间的降维空间

有时可能是原空间的升维空间。

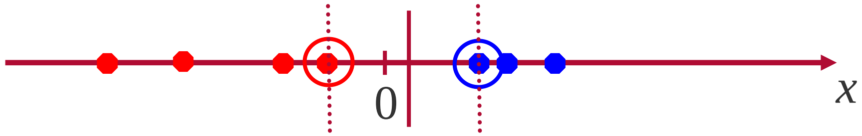
不识庐山真面目，

只缘身在此山中。

——苏轼

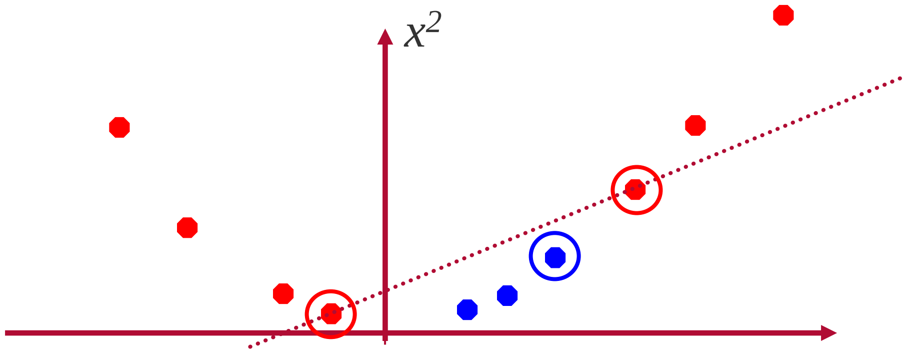
换一个角度，从更高维的空间更容易发现原空间的性质。

- Datasets that are linearly separable (with some noise) work out great:



- But what are we going to do if the dataset is just too hard?

- How about … mapping data to a higher-dimensional space:



目录

- 16.1 核方法
- 16.2 非线性支持向量机
  - 16.2.1 特征空间
  - 16.2.2 核函数
  - 16.2.3 常用核函数
  - 16.2.4 非线性支持向量机
- 16.3 多核方法

16.1 核方法

问题

现实应用中数据通常是非线性可分的

目标概念通常不能用简单线性函数进行类认知表示

在原空间表达过于复杂的类认知表示导致黑箱、不易解释

需要寻找更加抽象的非线性可分的特征空间

方法

在比输入空间更高维的空间表达类认知表示

优势

如果低维到高维的变换映射是非线性映射

则在**低维空间中需要复杂非线性函数才能表示的类**

到**高维空间中**以后可能**用线性函数就可以表示**

核方法

类输入表示

$(X, U, \underline{X}, Sim_X)$

类输出表示

$(Y, V, \underline{Y}, Sim_Y)$

通过空间变换映射$\varphi()$, $\forall k, y_k = \varphi(x_k)$

实现数据升维 $p = \dim(X) < \dim(Y) = d$

无穷维

$\varphi()$是非线性映射，有时难于直接计算

并不需要直接知道$\varphi()$, 只需知道$\forall k \forall l, \varphi(x_k)^T \varphi(x_l)$即可

$\forall k \forall l, k(x_k, x_l) = \varphi(x_k)^T \varphi(x_l)$，$k(\cdot,\cdot)$已知

核方法

目录

- 16.1 核方法
- 16.2 非线性支持向量机
  - 16.2.1 特征空间
  - 16.2.2 核函数
  - 16.2.3 常用核函数
  - 16.2.4 非线性支持向量机
- 16.3 多核方法

16.2 非线性支持向量机

对非线性可分两分类问题，对支持向量机引入核方法，就可以得到非线性可分的支持向量机。

选择合适的核函数将训练数据非线性映射到高维空间，不增加参数个数，提高了线性表示机制的分类能力。

16.2.1 特征空间

预处理$x = (x_1, x_2, \cdots, x_p)^T \rightarrow \varphi(x) = (\phi_1(x), \phi_2(x), \cdots, \phi_d(x))^T$

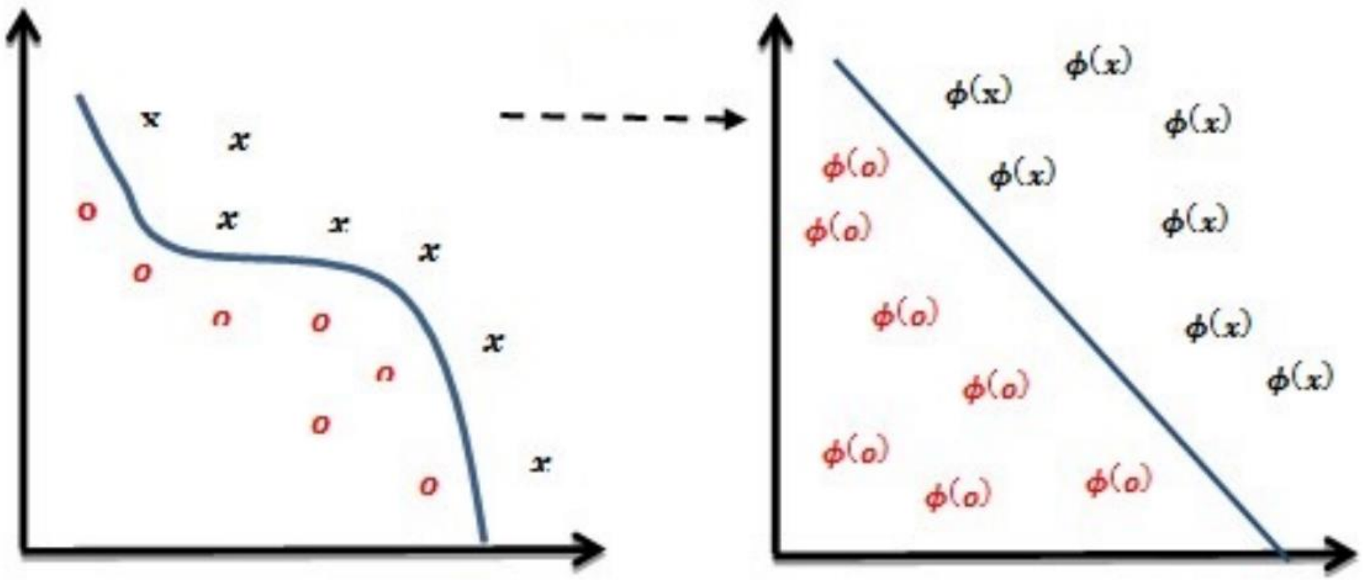


图 16.1　可以简化分类问题的非线性映射

在输入特征空间中不能使用线性函数分开的数据
在输出空间中就可能线性分开

16.2.2 核函数

假设数据输出特征空间上的线性判别函数为

$$f(x) = \sum_{i}^{d} w_i \phi_i(x) + b = w^T \varphi(x) + b \quad 16.1$$

代入$w = \sum_{k=1}^{N} \alpha_k u_k \varphi(x_k)$，则

$$f(x) = \sum_{k}^{N} \alpha_k u_k \langle \varphi(x_k) \cdot \varphi(x) \rangle + b \quad 16.2$$

核函数定义：对于所有的 $x, z \in X$,K 满足:

$$K(x, z) = \langle \varphi(x) \cdot \varphi(z) \rangle \quad 16.3$$

核方法

φ是从数据输入特征空间到输出特征空间的映射

核函数

特征映射很难构造，核函数容易构造，需要判断给定的函数是否是核函数。

定理16.1 （核函数的充要条件） 设K：$X \times X \to R$是对称函数，则$K(x,z)$为核函数的充要条件是对于任意$x_k \in X, k = 1,2,\cdots,N$， $K(x,z)$对应的Gram矩阵：

$$K = [K(x_k, x_l)]_{N\times N} \qquad (16.4)$$

是半正定矩阵。

根据任意有限训练样本集合计算Gram矩阵并判断其是否为半正定，这不是简单的计算。

常选用已有的核函数

16.2.3 常用核函数

(1)q次多项式：

$$K(x_k, x) = (x^T x_k + 1)^q$$

其中q为参数，当q=2,p=2时有：

$$\begin{aligned} K(x,z) &= (z^T x + 1)^2 \\ &= ((x)_1 (z)_1 + (x)_2 (z)_2 + 1)^2 \\ &= 1 + 2(x)_1 (z)_1 + 2(x)_2 (z)_2 + \\ &\quad 2(x)_1 (x)_2 (z)_1 (z)_2 + ((x)_1 (z)_1)^2 + ((x)_2 (z)_2)^2 \end{aligned}$$

它对应的特征映射函数为：

$$\phi(x) = [1, \sqrt{2}(x)_1, \sqrt{2}(x)_2, \sqrt{2}(x)_1 (x)_2, ((x)_1)^2, ((x)_2)^2]^T$$

常用核函数

(2)径向基函数：

$$K(x_k, x) = exp\left(-\frac{\left\|x_k - x\right\|^2}{\delta^2}\right)$$

球形核， x_k 为球形中心，参数 δ 为球形半径可以自由定义。

(3)S形函数：

映射函数写不出来

$$K(x_k, x) = \tanh\left(2x^T x_k + 1\right)$$

该函数中的 $\tanh(\cdot)$ 与S形函数形状类似，区别在于这里取值[-1，+1]。

16.2.4 非线性支持向量机

利用核函数可以将非线性可分问题转化为线性可分的支持向量机问题。

$$\max_{\alpha} W(\alpha) = \sum_{k=1}^{N} \alpha_k - \frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N} u_k u_l \alpha_k \alpha_l \boxed{K(x_k, x_l)}$$

$$\text{s.t. } \sum_{k=1}^{N} u_k \alpha_k = 0$$

$$0 \leq \alpha_k \leq C; K = 1, 2, \ldots, N$$

非线性支持向量机

求得最优解 α^* ，选择 α^* 的一个正分量，计算

$$w^* = \sum_{K=1}^{N} u_k \alpha_k^* x_k \qquad 16.8$$

$$b^* = u_l - \sum_{K=1}^{N} u_k \alpha_k^* K(x_{k,} x_l) \qquad 16.9$$

超平面函数为

$$f(x) = \sum_{K=1}^{N} u_k \alpha_k^* K(x, x_k) + b^* \qquad 16.10$$

补：非线性支持向量回归

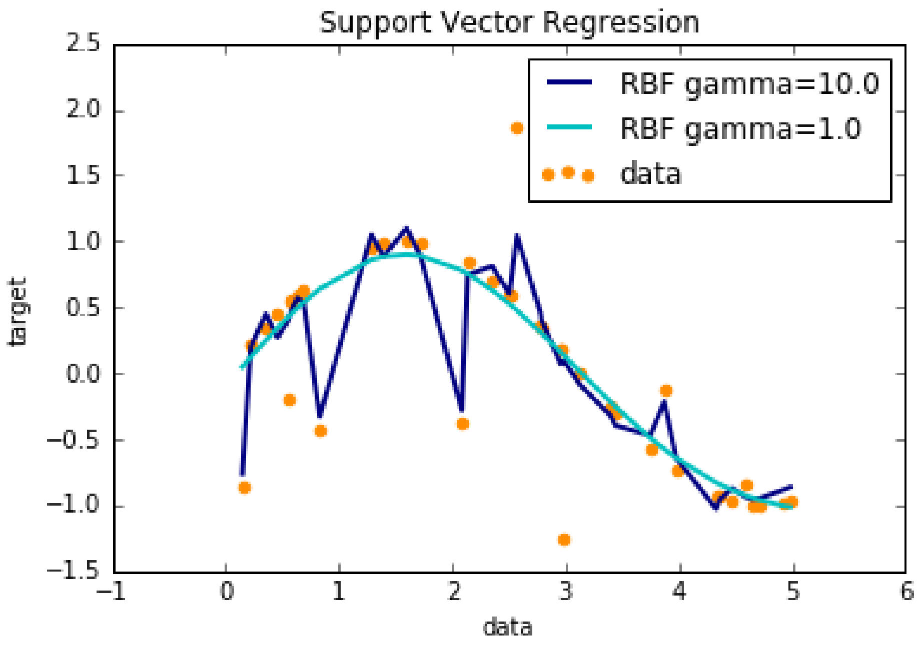
利用核函数可以将线性SVR的回归模型推广为非线性回归函数。

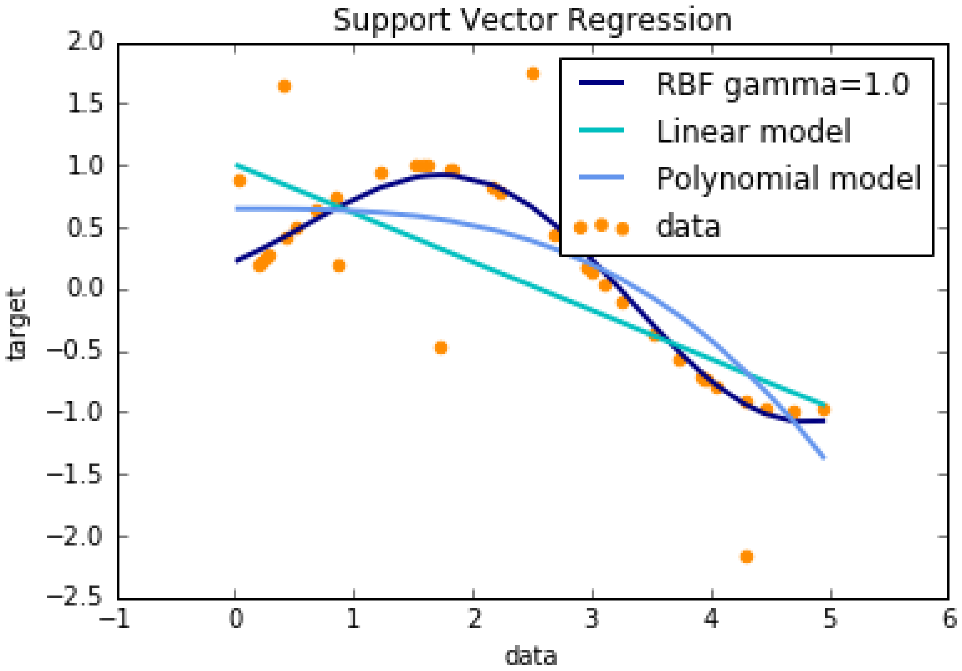
$$f(x) = w^T\varphi(x) + b$$

$$\min_{\alpha,\hat{\alpha}} W(\alpha,\hat{\alpha}) = \sum_{k=1}^{N}(\hat{\alpha_k} - \alpha_k)y_k - \varepsilon\sum_{k=1}^{N}(\hat{\alpha_k} + \alpha_k)$$
$$-\frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N}(\hat{\alpha_k} - \alpha_k)(\hat{\alpha_l} - \alpha_l)K(x_k, x_l)$$
$$\text{s.t. } \sum_{k=1}^{N}(\hat{\alpha_k} - \alpha_k) = 0,\ \ 0 \le \alpha_k, \hat{\alpha_k} \le C;\ k = 1,2,\ldots,N$$

$$f(x) = \sum_{K=1}^{N}(\hat{\alpha_k} - \alpha_k)K(x, x_k) + b^*$$

补：非线性支持向量回归举例





目录

- 16.1 核方法
- 16.2 非线性支持向量机
  - 16.2.1 特征空间
  - 16.2.2 核函数
  - 16.2.3 常用核函数
  - 16.2.4 非线性支持向量机
- 16.3 多核方法

16.3 多核方法：两种策略

简单核函数有时不太有效，可以合成简单核函数构造新的核函数。

1、假设两个有效的简单核函数 $K_1(x,z)$ 和 $K_2(x,z)$

$$K(x,z)=\begin{cases} cK_1(x,z) \\ K_1(x,z)+K_2(x,z) \\ K_1(x,z)\cdot K_2(x,z) \end{cases}$$

为什么可以构造?

合成后仍是半正定矩阵

多核方法：两种策略

2、使用训练集的不同子集，可以融合不同数据域的信息

将训练集合x表示为x_A，x_B，通过简单核函数构造新的核函数为

$$K_A(x_A,z_A)+K_B(x_B,z_B)=\varphi_A(x_A)^T\varphi_A(z_A)+\varphi_B(x_B)^T\varphi_B(z_B)$$
$$=\varphi(x)^T\varphi(z)$$
$$=K(x,z)$$

其中 $x=[x_A，x_B]$为两种表示相连接，对应于x_A，x_B的两个核函数相加相当于连接后的特征向量之间点乘。

多核方法

两个简单核函数→多个简单核函数

$$K(x,z)=\sum_{t=1}^{m}K_t(x,z)$$

$$K(x,z)=\sum_{t=1}^{m}\eta_t K_t(x,z)$$

避免了挑选最优核函数的过程

限制 $\eta_t \geq 0, \sum_t \eta_t = 1$

支持向量机：多核情况

$$\max_{\alpha} W(\alpha)=\sum_{k=1}^{N}\alpha_k-\frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N}u_k u_l \alpha_k \alpha_l \sum_{t=1}^{m}\eta_t K_t(x_k,x_l)$$

$$f(x)=\sum_{k=1}^{N}\alpha_k^* u_k \sum_{t=1}^{m}\eta_t K_t(x_k,x_l)+b^*$$

多核学习

讨论

一旦数据给定，现有数据特征不一定完全符合知识提取的需要

特征再生

- 数据特征冗余 → 降维
- 数据特征过于复杂 → 空间变换简化特征
  如果新空间较原空间维数高，即为特征升维
  核方法是常见的特征升维方法
- 数据特征缺失 → 补齐，多源数据、矩阵填充
- 样本缺失 → 迁移学习、非平衡数据学习

本章总结

多类数据升维
- 核方法
- 非线性支持向量机
  - 特征空间
  - 核函数
  - 常用核函数
  - 非线性支持向量机
- 多核方法 → 简单核构造新的核函数

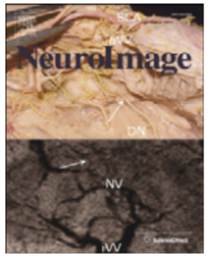
NeuroImage 55 (2011) 856–867

ELSEVIER

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



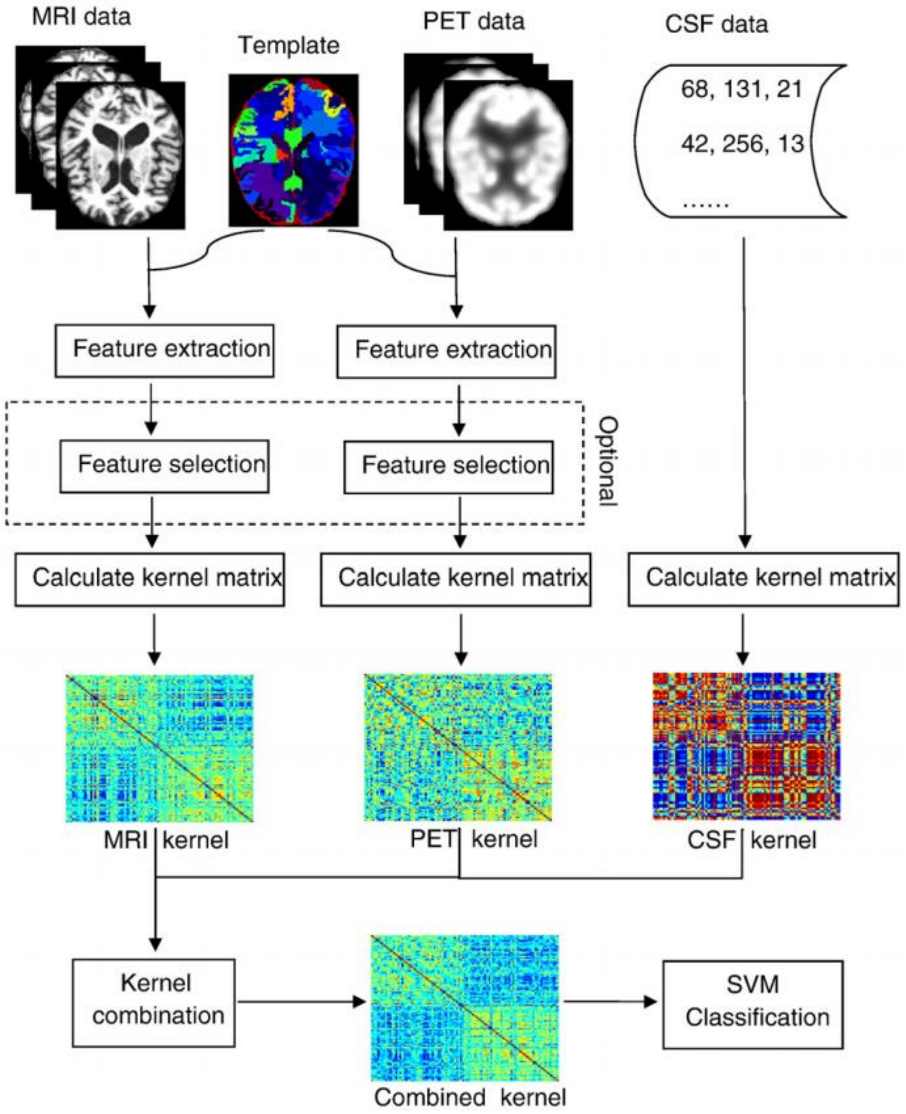
Multimodal classification of Alzheimer's disease and mild cognitive impairment

Daoqiang Zhang [a], Yaping Wang [a,b], Luping Zhou [a], Hong Yuan [a], Dinggang Shen [a,*]
and the Alzheimer's Disease Neuroimaging Initiative [1]

[a] *Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA*
[b] *Department of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi Province, China*

Dosenbach et al., Science, 2010

应用实例



■Dosenbach et al., Science, 2010

应用实例

$$\min_{\mathbf{w}^{(m)},b,\xi_i} \frac{1}{2}\sum_{m=1}^{M}\beta_m\left\|\mathbf{w}^{(m)}\right\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$s.t. \quad y_i\left(\sum_{m=1}^{M}\beta_m\left(\left(\mathbf{w}^{(m)}\right)^T\phi^{(m)}\left(\mathbf{x}_i^{(m)}\right)+b\right)\right)\geq 1-\xi_i$$

$$\xi_i \geq 0, i=1,\ldots,n.$$

$$\max_{\alpha} \sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j\sum_{m=1}^{M}\beta_m k^{(m)}\left(\mathbf{x}_i^{(m)},\mathbf{x}_j^{(m)}\right)$$

$$s.t. \quad \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$0\leq\alpha_i\leq C, i=1,\ldots,n.$$

$$f\left(\mathbf{x}^{(1)},\mathbf{x}^{(2)},\ldots,\mathbf{x}^{(M)}\right) = sign\left(\sum_{i=1}^{n}y_i\alpha_i\sum_{m=1}^{M}\beta_m k^{(m)}\left(\mathbf{x}_i^{(m)},\mathbf{x}^{(m)}\right)+b\right)$$

应用实例

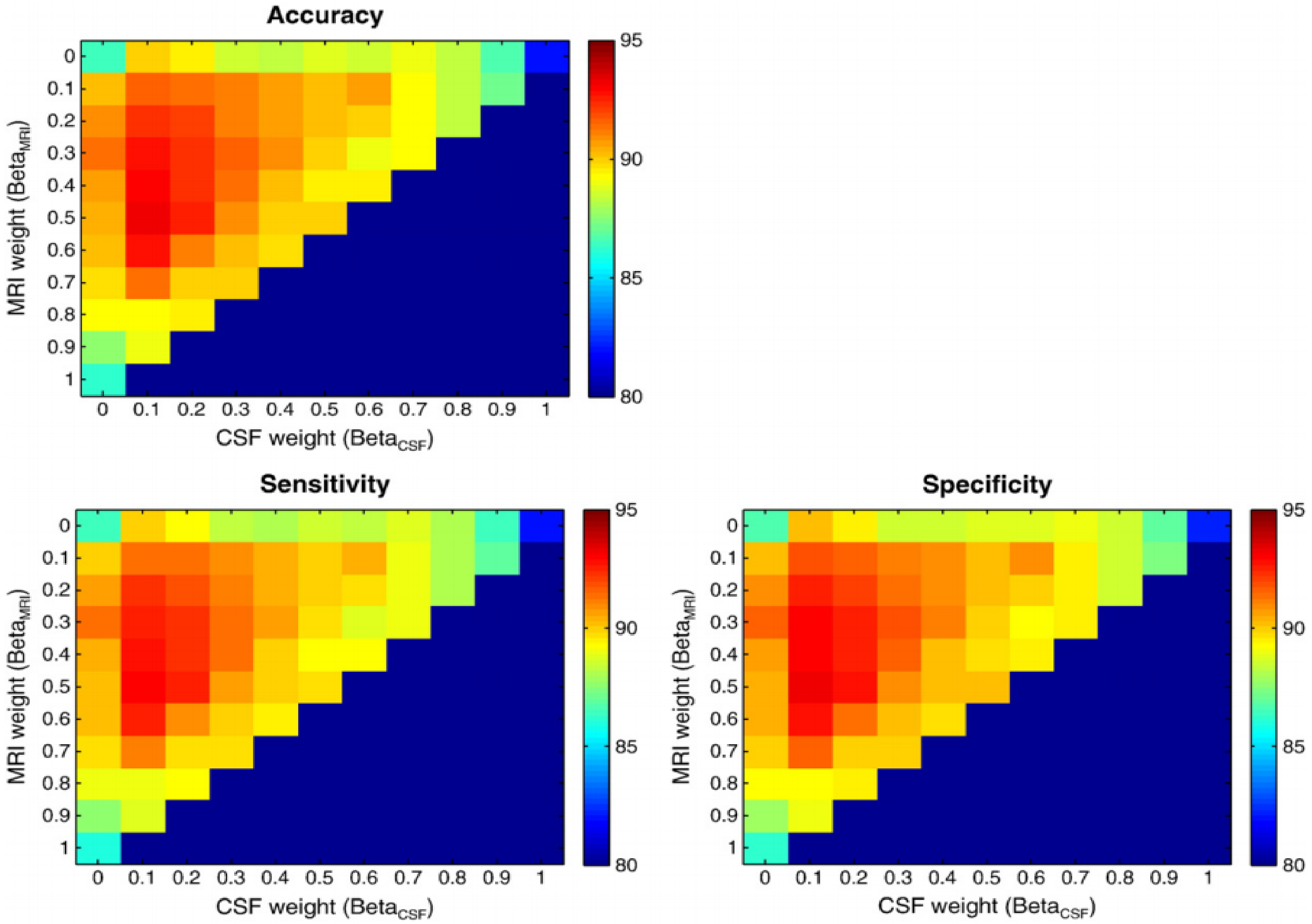


Fig. 3. AD classification results with respect to different combining weights of MRI, PET and CSF.

作业

理论部分

教材161页第2、3题

实践部分

1. 基于给定的样本特征和标签数据，采用支持向量机对样本进行分类，比较使用线性核函数和非线性核函数的效果;

2. 实现《 Multimodal classification of Alzheimer ‘s disease and mild cognitive impairment 》描述的算法，在给定数据上测试算法性能（将12720维数据均匀切割为3段，当作是3个模态特征）

北京交通大学《机器学习》课程组成员

于　剑：jianyu@bjtu.edu.cn，http://faculty.bjtu.edu.cn/6463/

景丽萍：lpjing@bjtu.edu.cn，http://faculty.bjtu.edu.cn/8249/

田丽霞：lxtian@bjtu.edu.cn，http://faculty.bjtu.edu.cn/7954/

黄惠芳：hfhuang@bjtu.edu.cn，http://faculty.bjtu.edu.cn/7418/

杨　凤：fengyang@bjtu.edu.cn，http://faculty.bjtu.edu.cn/8518/

吴　丹：wudan@bjtu.edu.cn，http://faculty.bjtu.edu.cn/8925/

万怀宇：hywan@bjtu.edu.cn，http://faculty.bjtu.edu.cn/8793/

王　晶：wj@bjtu.edu.cn，http://faculty.bjtu.edu.cn/9167/