第3讲:支持向量机对偶，核与回归

IMAGE

C19机器学习希拉里2015Zisserman

IMAGE

•原始形式和双重形式

•重新考虑线性可分性

•特征图

支持向量机的内核

•回归

•岭回归

•基函数

SVM -评论

IMAGE

•我们已经看到，对于SVM学习，线性分类器f(x) = w>x + b

表示为解决w上的优化问题:XN wmin∈Rd||w|| 2+ C max(0,1−yif(xi)) i

•这个二次优化问题被称为原始问题。

•相反，支持向量机可以被表述为学习线性分类器XN f(x) = αiyi(xi>x) + b

我

通过求解在αi上的优化问题。

•这就是对偶问题，我们会看看这个公式的优点。

对偶形式的草图推导

IMAGE

代表定理表明，解w总是可以写成训练数据的线性组合:

XN j = 1

w =

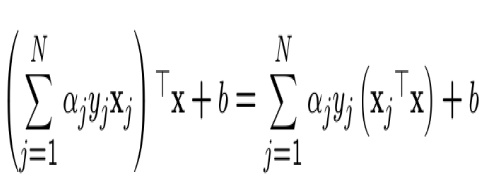
αjyjxj

证明:见范例表。

现在，把w代入f(x) = w>x + b

f(x) =⎝x´和w在成本函数minw ||w||>服从于yi w xi + b´≥1，∀i

³



||w|| 2=⎩⎧⎨X αjyjxj⎫⎬>⎧⎨Xαkykxk⎬⎫= j⎭⎩k⎭X αjαkyjyk(xj>xk) jk

因此，在yi⎝的约束下，在αj X⎛XN min αjαkyjyk(xj>xk)上存在一个等价的优化问题

αj jkαjyj(xj>xi) + b⎠⎞≥1，∀i j=1

并且需要更多的步骤来完成这个推导。

原始和对偶公式

IMAGE

N为训练点个数，d为特征向量x的维数

XN wmin∈Rd||w|| 2+ C max(0,1−yif(xi)) i

对偶问题:对于α∈RN(未证明的表述):maxX αi−1X αjαkyjyk(xj >xk)，且满足∀i≤αi≤C，且x αi≥0 2 i jk i

α I yi = 0

•需要学习d个原始参数，N个对偶参数

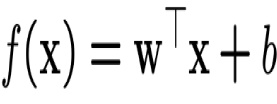
•如果N << d，那么求解α比求解w更有效

•对偶形式只涉及(xj> xk）.当我们研究内核时，我们将回过头来讨论为什么这是一个优势。

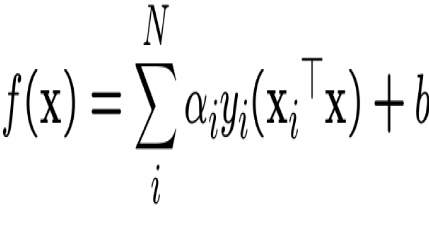
原始和对偶公式

IMAGE

分类器的原始版本:

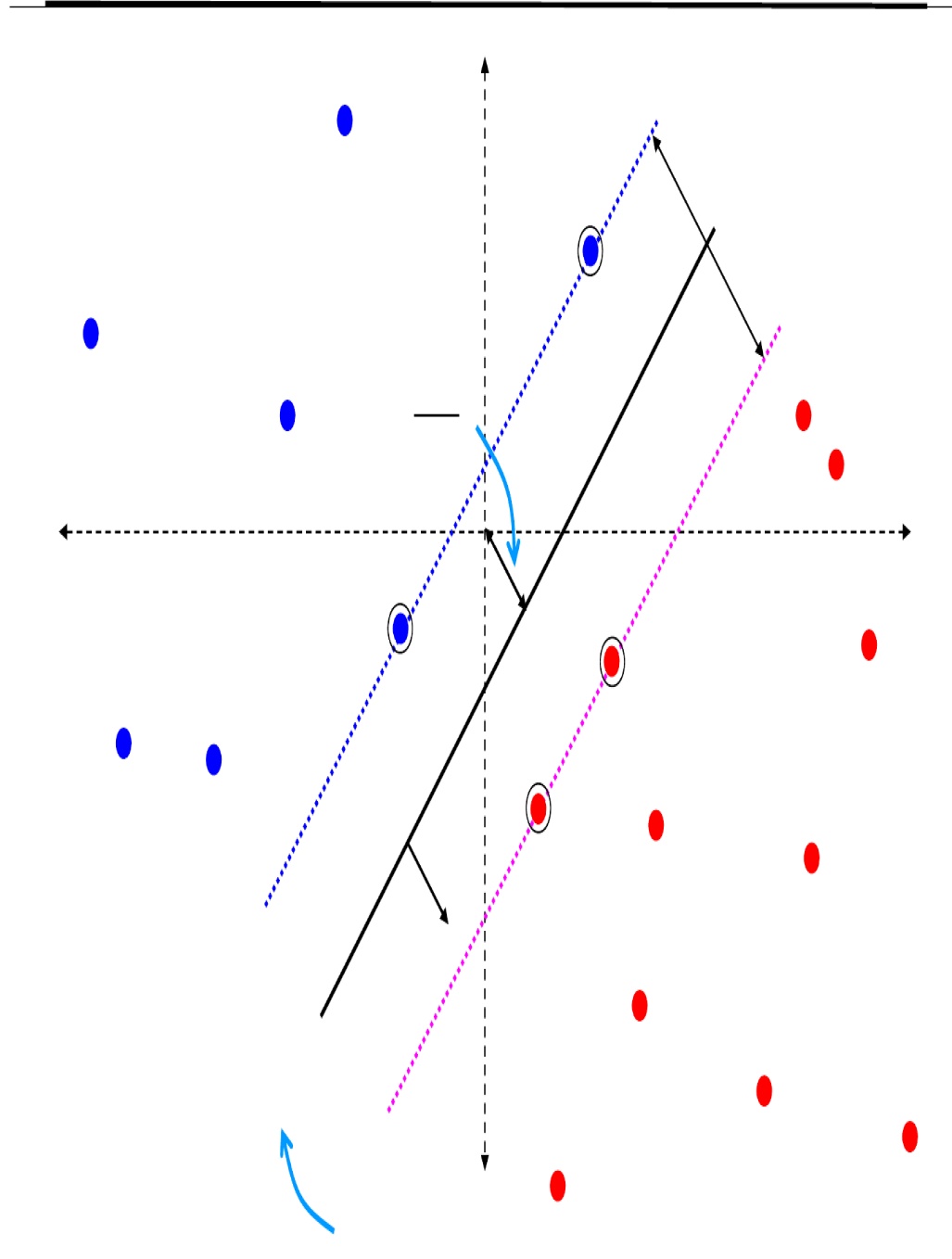


分类器的双版本:



乍一看，双重形式似乎有K-NN分类器的缺点-它需要训练数据点xi。但是，很多αi是零。那些非零的定义了支持向量xi。

支持向量机



**支持向量**

**支持向量**

f (x) =

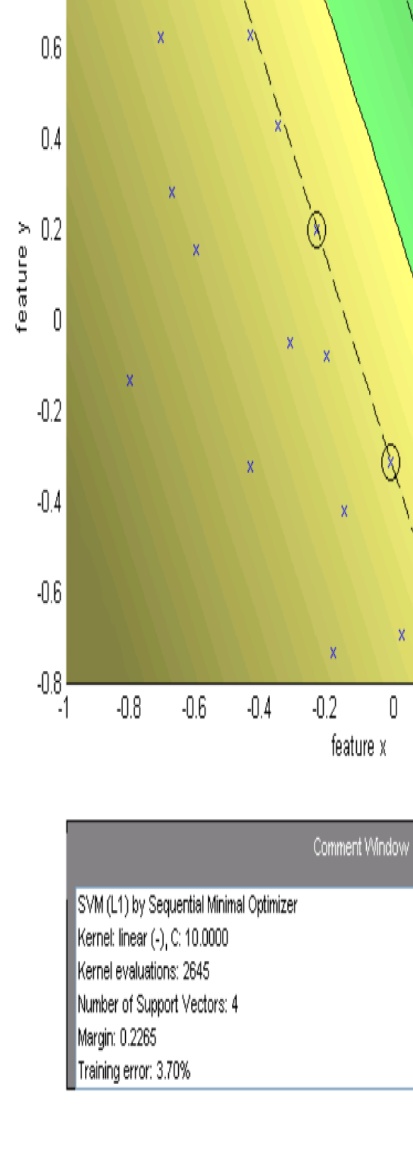
支持向量

**w**

X   
αiyi(ξ> x) + b  
我

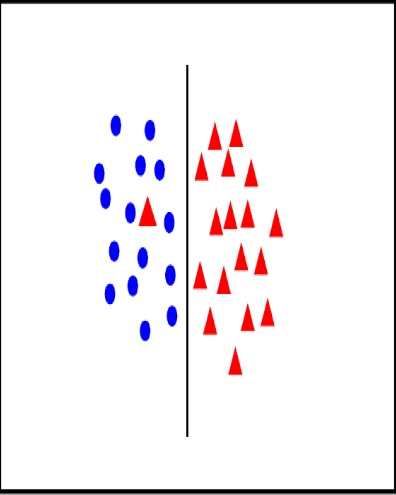
软边缘

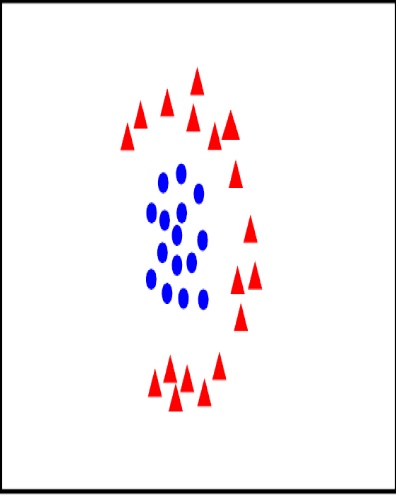
C = 10



处理不能线性可分的数据

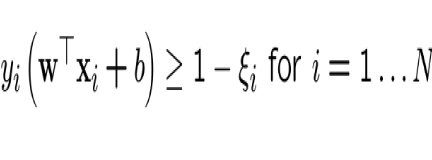
IMAGE





引入松弛变量XN min||w|| 2+ C ξi w∈Rd，ξ i∈R+ i

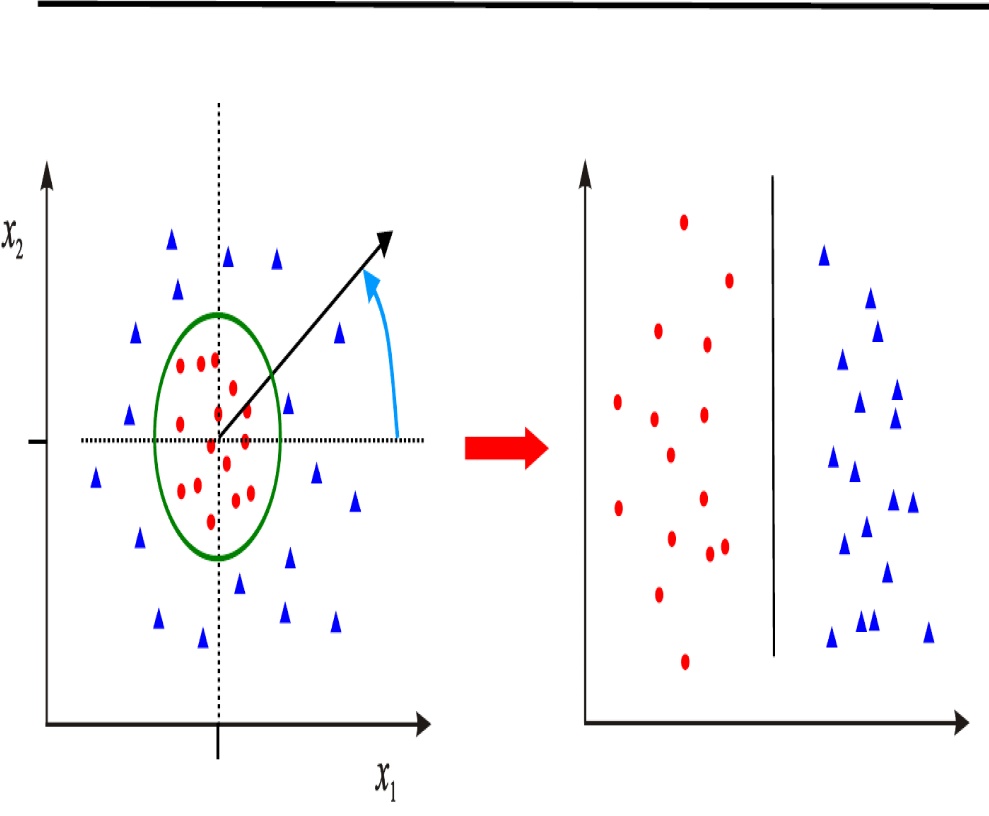
受



•线性分类器不合适

?？

解决方案1:使用极坐标



r

θ

< 0

> 0

θ

0

0

r

•数据在极坐标下是线性可分的

•在原始空间Ã中非线性行为 x1！一个r！Φ: x2→θR2→R2

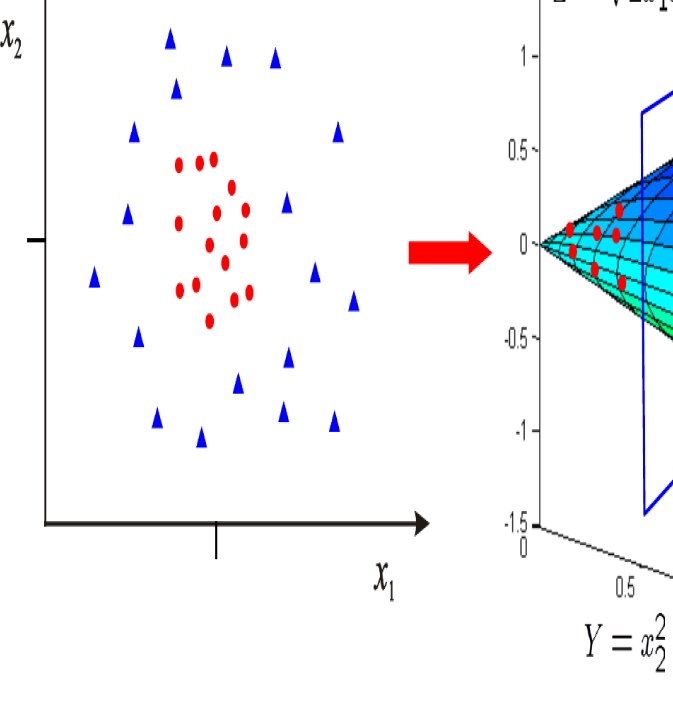
解决方案2:将数据映射到更高维度

IMAGE

一个!⎛x12⎞→⎝⎜√x2 2⎟⎠R2→R3 2x1x2

x1, x2)

Φ:



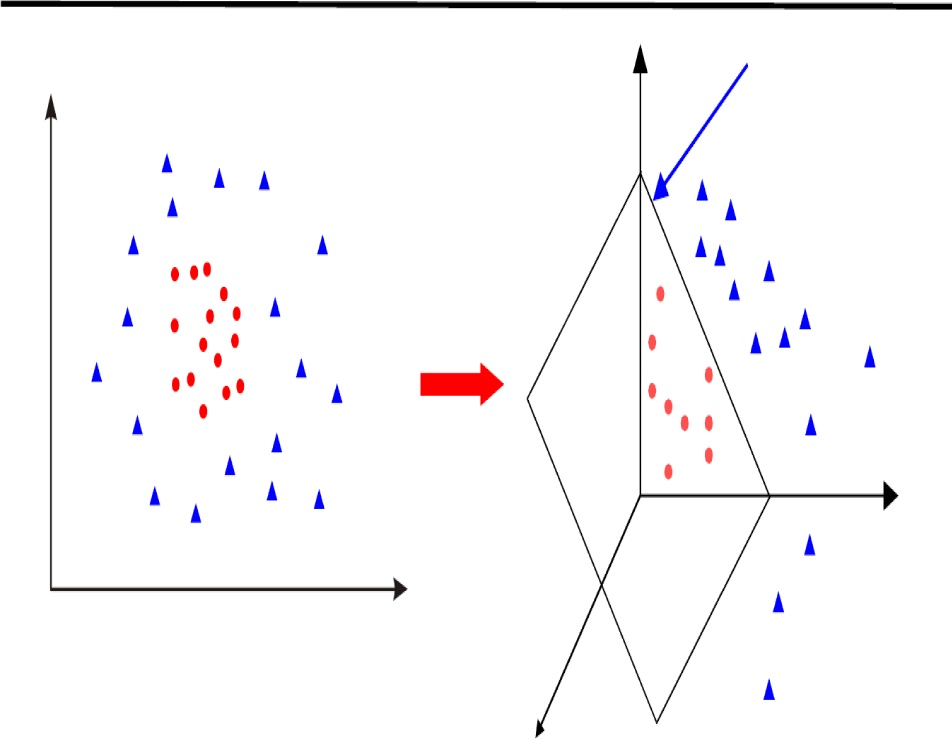
0

0

•数据在3D中线性可分

•这意味着问题仍然可以通过线性分类器解决

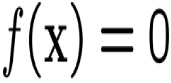
SVM分类器在一个转换的特征空间



Φ: x→Φ(x) Rd→R D

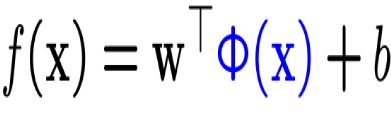
理查德·道金斯

理查德·道金斯



Φ

在w中学习RD的分类器线性:



Φ(x)为特征图

变换特征空间中的原始分类器

IMAGE

分类器，带w∈RD:



学习，对于w∈RD

∈XN wmin∈RD||w||2 + C imax(0,1−yif(xi))

•简单地映射x到Φ(x)，其中数据是可分离的

•求解高维空间RD中的w

•如果D >> D，那么对于w有更多的参数需要学习，这可以避免吗?

变换特征空间中的对偶分类器

IMAGE

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | f (x) =  →f (x) = | | | X N  我  X N  我 | αi y I x I > x + b  αiyi Φ(xi)>Φ(x) + b |
| 学习: | 马克斯  α≥0  →马克斯α我≥0 | X我  X我 | 我α−  我α− | 1 x 2 jk  αj αk y j y k x j > x k  1 x 2 jk  αjαkyjykΦ(xj) >Φ(xk) | |
|  |
|  |
| 受 |  |  |  |  |  |
|  | 对于∀i, 0≤αi≤C | | | | |

分类器:

我

变换特征空间中的对偶分类器

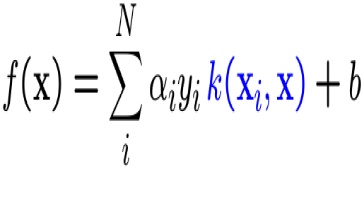
IMAGE

•注意，Φ(x)只成对出现Φ(xj)>Φ(xi)

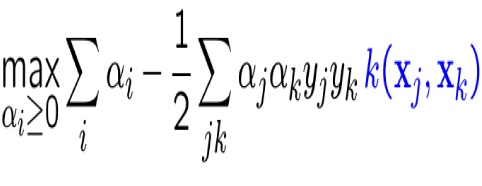
•一旦计算出标量积，只需要学习N维向量α;没有必要在D维空间中学习，因为它是原始的

•写入k(xj, xi) = Φ(xj)>Φ(xi)。这就是所谓的内核

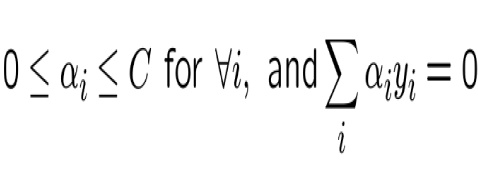
分类器:



学习:



受



特殊的转换

IMAGE

|  |  |  |  |
| --- | --- | --- | --- |
| Φ: | x1  2 x2 !  →⎝√x2⎛⎜⎞  2x1x2⎟⎠R2→R3  Φ(x)>Φ(z) =³x2 1,x2 2，√x´⎜⎝  ⎛z1 2  2x 1 2 z 2  √ 2 2 z1z2  = x2 1 12 21 21 2z2 + x2z2 + 2x x zz | |  |
|  | ⎞⎟⎠ |
|  |
|  |
|  | ＝ | (x1z1 + x2z2) 2 |  |
|  | ＝ | (x > z) 2 |  |
| 内核的技巧 | |  |  |
| •分类器可以学习和应用，而无需明确计算Φ(x)  所需要的就是核k(x, z) = (x>z)2 | | | |
|

一个

x12

•学习的复杂性取决于N(通常是O(n3))而不是D

例子内核

IMAGE

•线性核k(x, x0) = x>x0 0³>´d•多项式核k(x, x) = 1 + x x0对于任何d > 0 -包含所有的多项式项到d度 0³ 0•高斯核k(x, x) = exp−||x−x || 2/2σ 2 for σ > 0 -无限维特征空间

基于高斯核的SVM分类器

IMAGE

N =训练数据的大小

xnf (x) = αiyik(xi, x) + bi

支持向量



体重(可能是

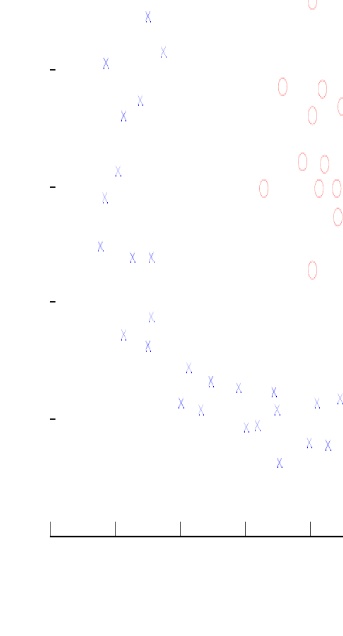
零)高斯核k(x, x0) = exp³−||x−x0||2/2σ2´径向基函数(RBF) SVM f(x) =XN αiyi exp³−||x−xi|| 2/2σ2´+b i

RBF核支持向量机实例

IMAGE

0．6

0．4



0.2功能x

0．4

0．6

0．8

1

功能y

0．2

0

-0.2

-0.4

-0.6 - -0.8

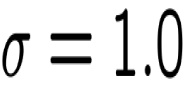
-0.6

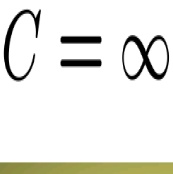
-0.4

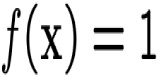
-0.2

0

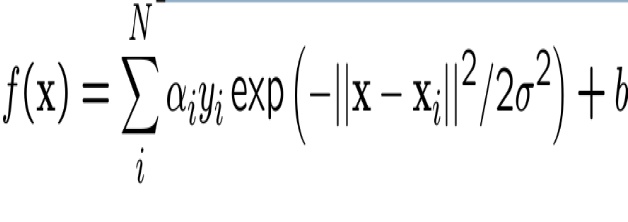
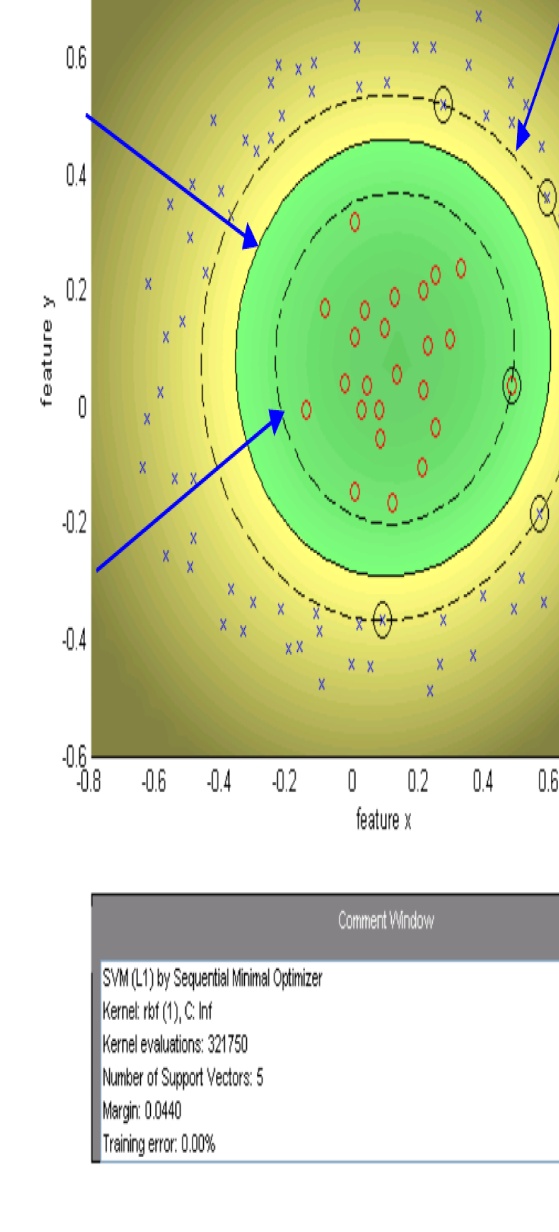
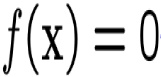
•数据在原始特征空间中不能线性可分

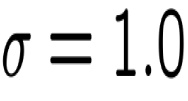


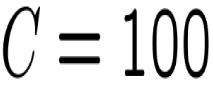


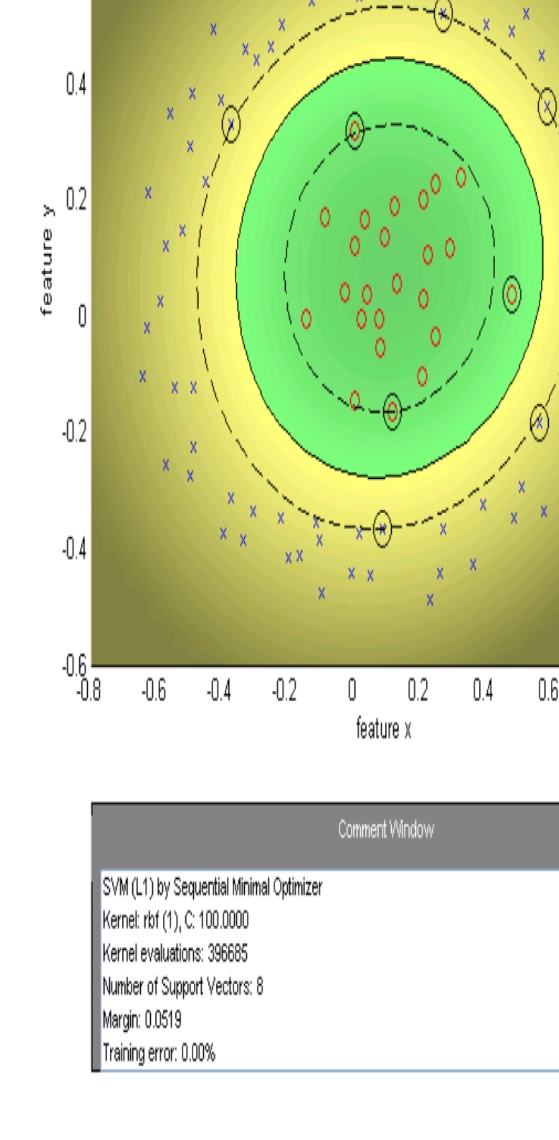


f (x) =−1





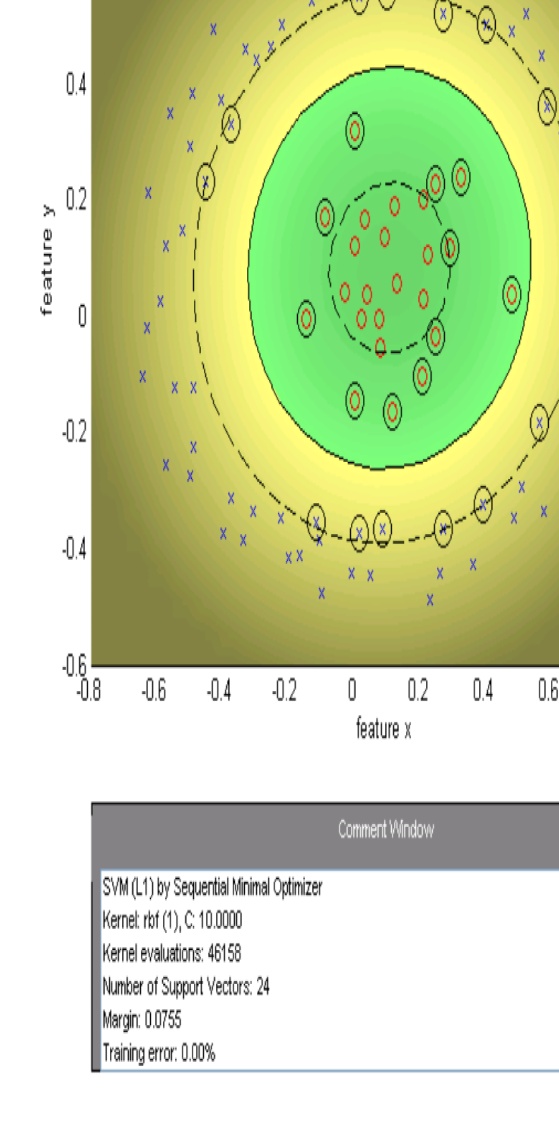
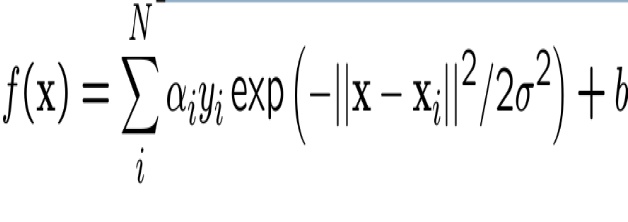


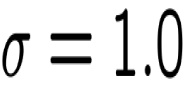


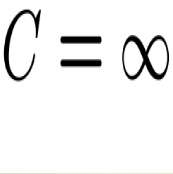
减少C，给予更宽(软)余地

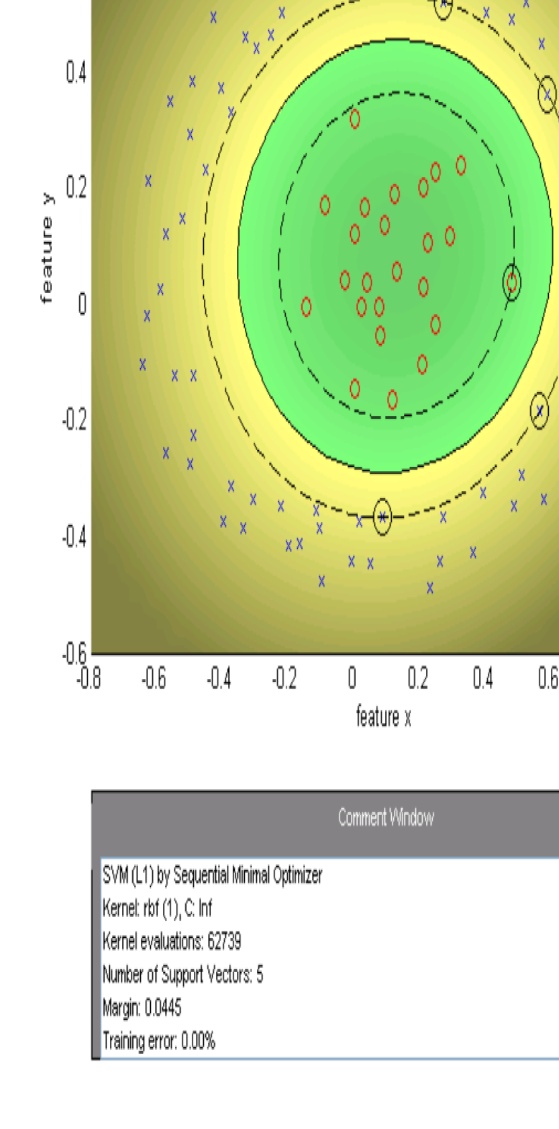
σ= 1.0

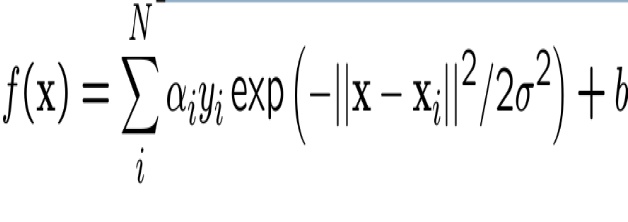
C = 10

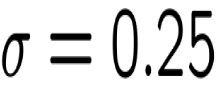


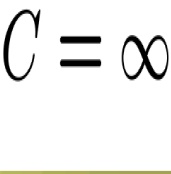


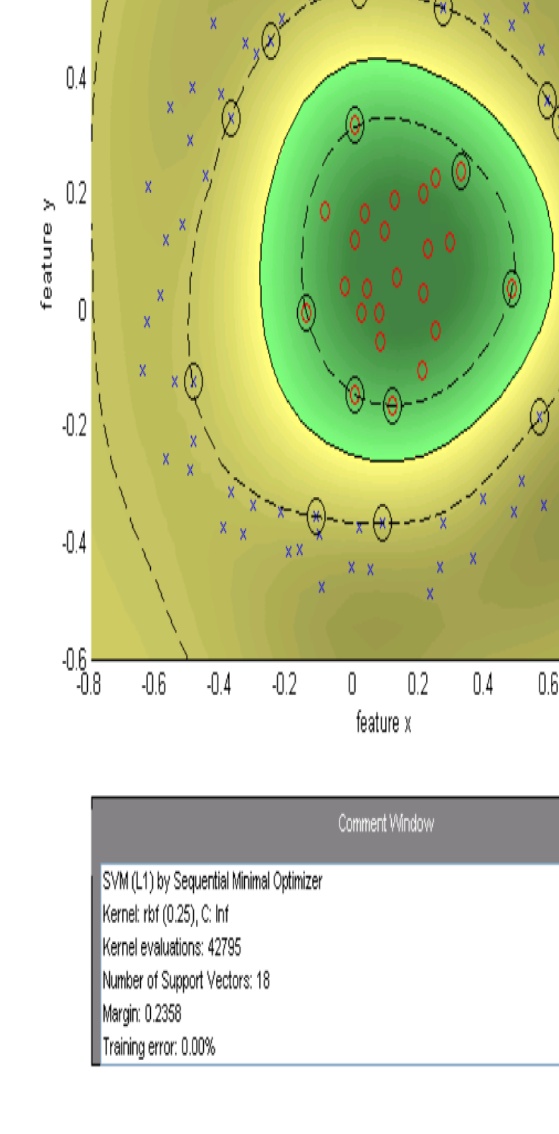








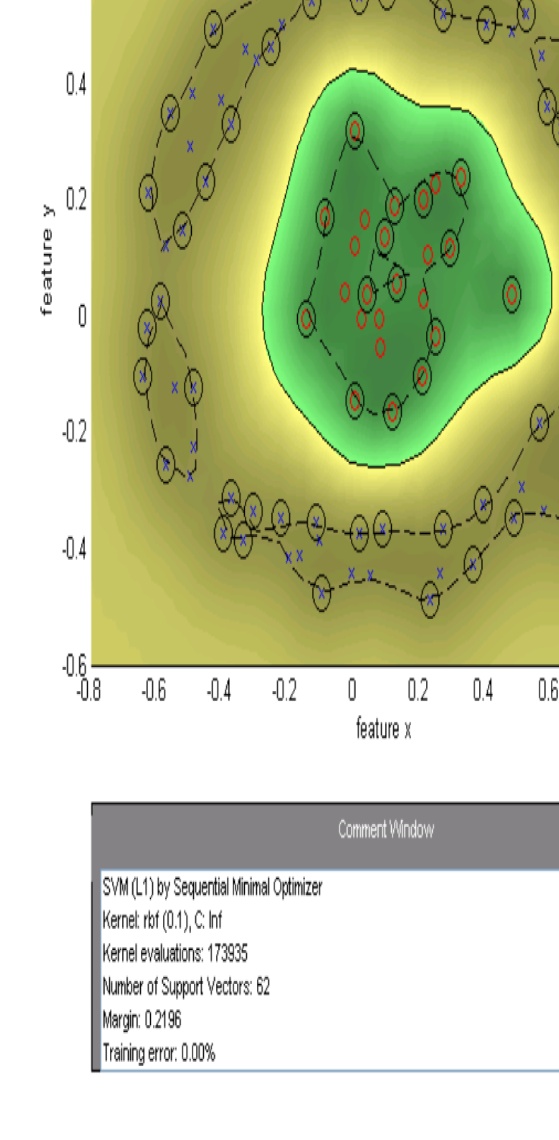
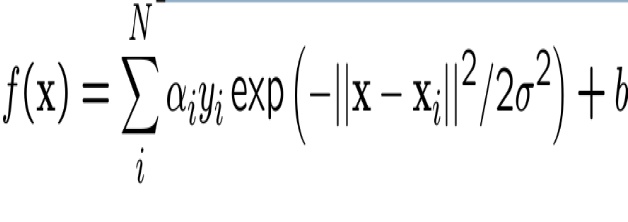




减小，向最近邻分类器移动

σ= 0.1

C =∞



内核技巧-总结

IMAGE

•可以为高维特征空间学习分类器，而无需实际将点映射到高维空间

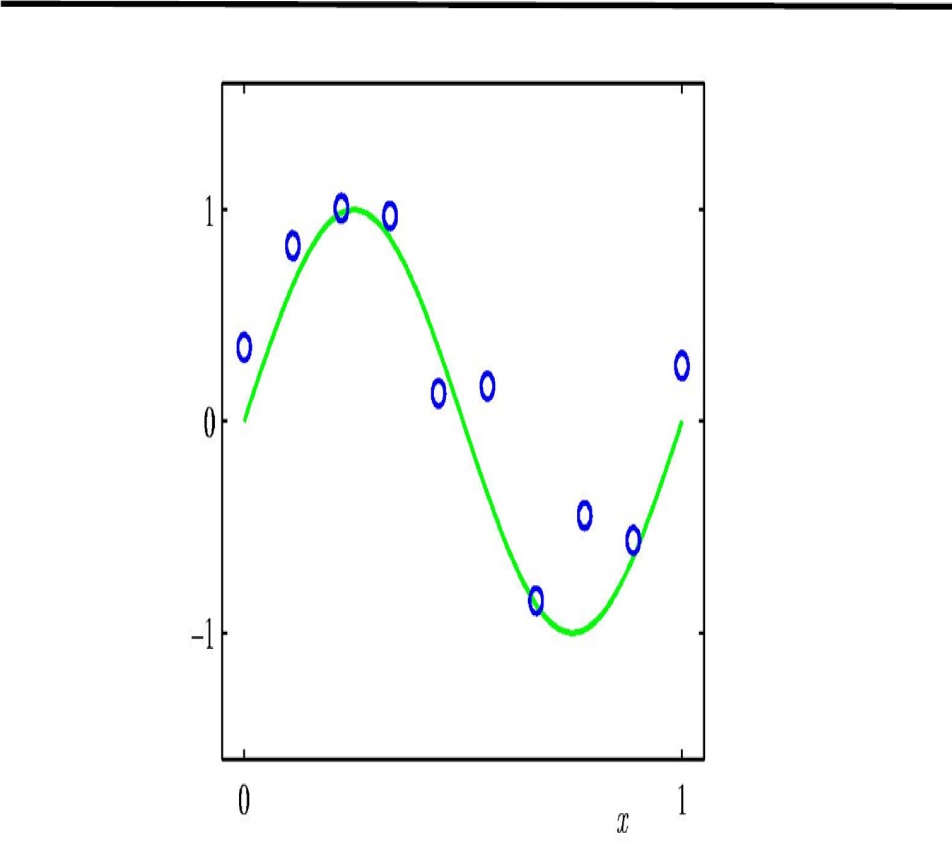
•数据在高维空间中可以线性可分，但在原始特征空间中不能线性可分

•核函数可以用于支持向量机，因为它是双形式的标量积，但也可以用于其他地方——它们与支持向量机的形式无关

内核也适用于非向量的对象，例如。

k(h, h0) = P 00 k min(hk, hk)对于有箱hk, hk的直方图

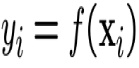
回归



y

•假设我们有N个观察值的训练集((x1, y1)，…， (xN, yN)) with xi∈Rd, yi∈R

•回归问题是估计f(x)从这个数据，以便



通过优化学习

IMAGE

•在分类的情况下，学习一个回归器可以表述为一个优化:

f∈f的最小值

XN l (f(xi)， yi) + λR (f) i=1



损失函数正则化

•有两个选择的损失函数和正则化

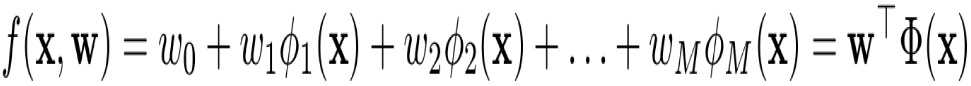
•例如平方损失，SVM“铰链式”损失

•平方正则化器，套索正则化器

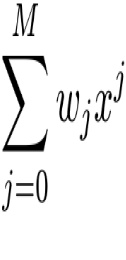
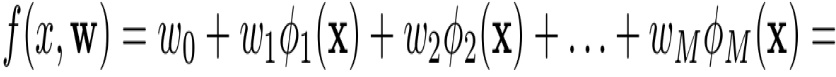
回归函数-非线性基函数的选择

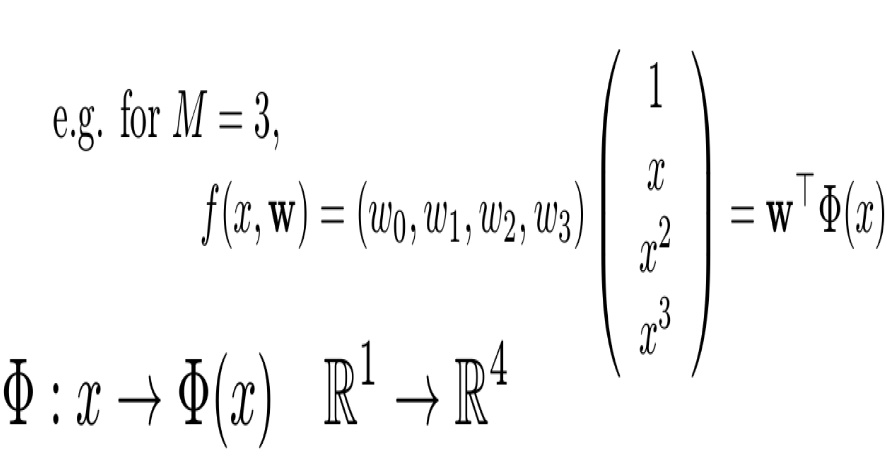
IMAGE

•回归函数y(x, w)是x的非线性函数，但在w中是线性的:

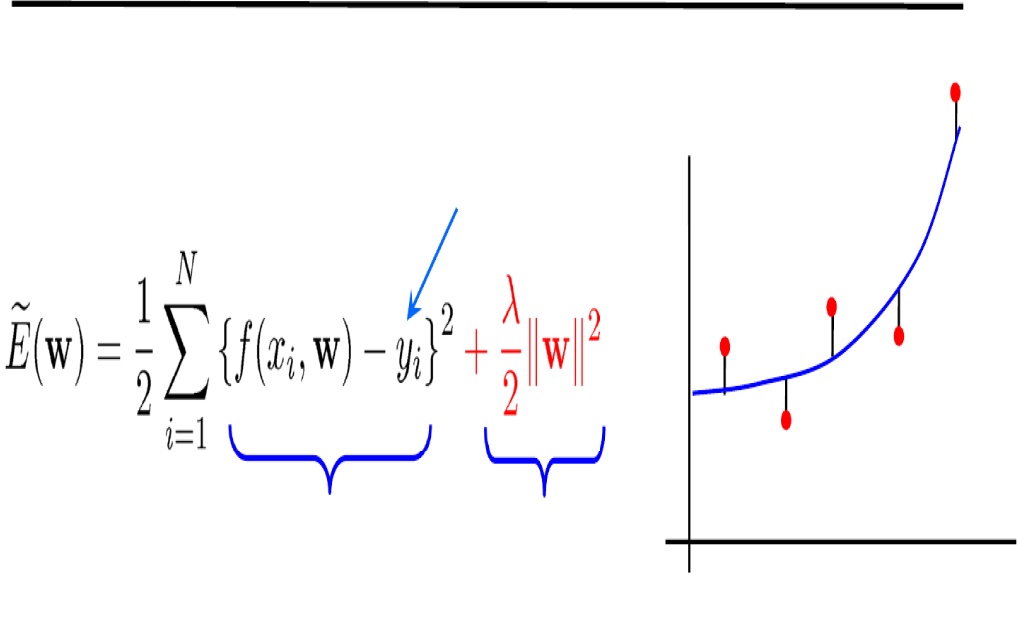


•例如，对于x∈R， φ的多项式回归j (x) = xj ：





最小二乘岭回归



•成本函数-平方损失:  
目标价值

易

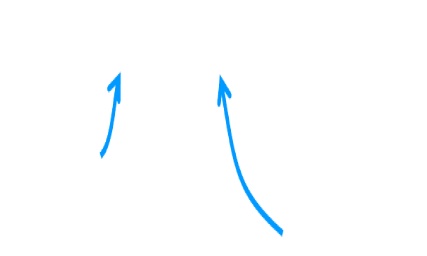
损失函数正则化

西

•x (1D)的回归函数:

F (x, w) = w0 + w1φ1(x) + w2φ2(x) +…+ wM φM (x) = w>Φ(x)

•NB平方损失产生于误差模型的最大似然估计



Yi = y ~ i + ni ni   
∼N(0,σ2)

真正价值

测量值

求权值w

IMAGE

符号:写目标和退化值n个向量=⎜⎜⎜⎜⎛yy2 1⎟⎟⎟⎟⎞=⎜⎜⎜⎜⎛Φ(xΦ(x12)) > > w w⎟⎞⎟⎟⎟= =⎢⎡⎢⎢⎢⎢11。φ1 (x1)…φM (x1)⎥⎥⎥⎥⎤⎜⎜⎜⎜⎛ww01⎞⎟⎟⎟⎟φ1(x2)…φM (x2) y⎜⎜。f⎜⎜。⎝⎟⎟⎠⎝⎟⎟⎠⎢⎣。Φw

．．⎥⎥⎦⎜⎜⎝⎟⎟⎠



yN Φ(xN)>w 1 φ1(xN)…φM (xN) wM

．

．

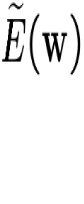
Φ为N × M设计矩阵

例如，对于基底函数为x2的多项式回归

⎡1x1x12⎤=⎢⎢⎢⎢1x2x2⎥⎥⎥⎛⎜⎞⎥⎝w

2 w0⎢⎢..⎥1⎟⎠⎣..⎥⎦w2 1 xN x2

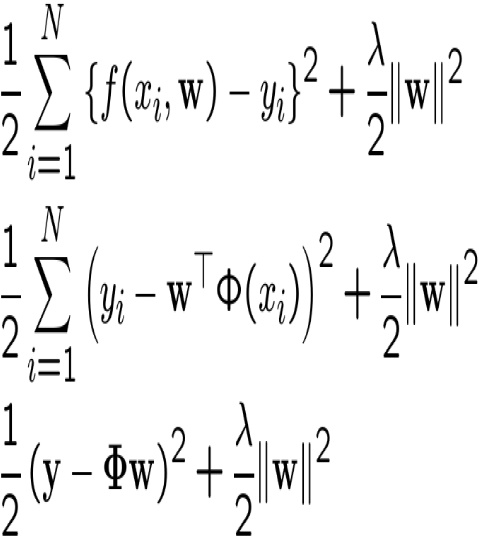
Φw



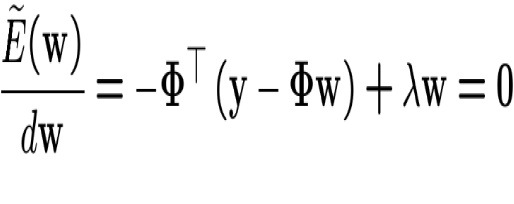
FORMULA

FORMULA

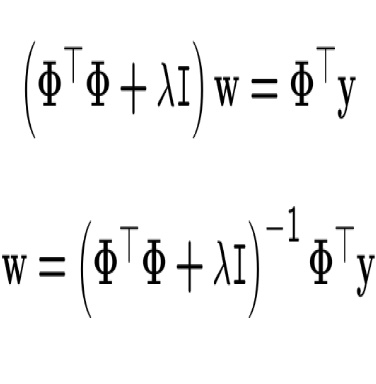
FORMULA



现在，计算导数w, r, t, w为零的最小值



因此



M w基=函数，³Φ> Φ N +数据λpointsI´−1



＝

Φ> y

IMAGE

FORMULA

Mx1 MxM MxN Nx1

•这表明存在唯一解。

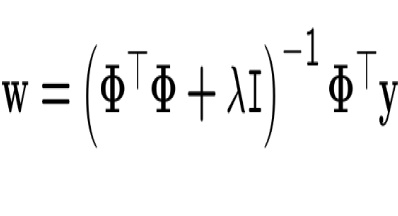
•如果λ= 0(没有正规化),那么w =(Φ>Φ)−1Φ> y =Φ+ y

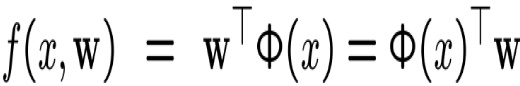
其中Φ+为Φ的伪逆(Matlab中的pinv)

•添加λ i项改善了逆函数的条件，因为如果Φ不是满秩，那么(Φ>Φ + λ i)将是(对于足够大的λ)

•如λ→∞，w→1 λΦ >y→0

•通常，正则化仅应用于w的不均匀部分，即w ~ ~，其中w = (w0, w ~)





= Φ(x)>³Φ> Φ + λI´−1 Φ>y = b(x)>y

输出是训练值{yi}的线性混合值b(x)

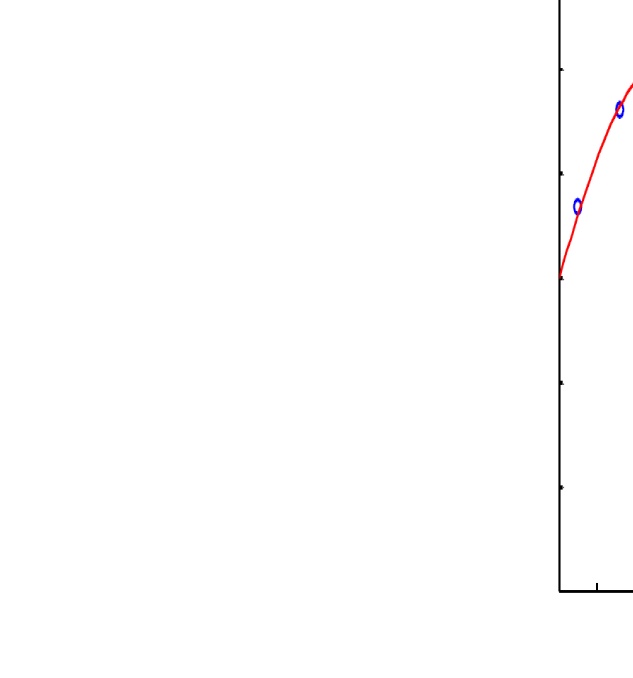
例1:多项式基函数

理想的适合

•红色曲线是真函数(不是多项式)

1．5

样本点



•数据点是y中添加噪声的曲线样本。

•可以选择使用的基函数的度数M，以及正则化的强度

1

0．5

y

0

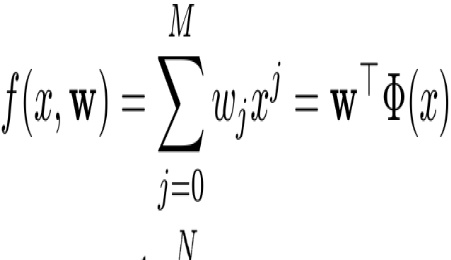
-0.5

－1

-1.5

0

0．1



0．2

0．3

0．4

0.5倍

0．6

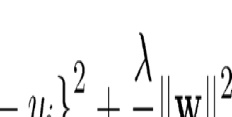
0．7

0．8

0．9

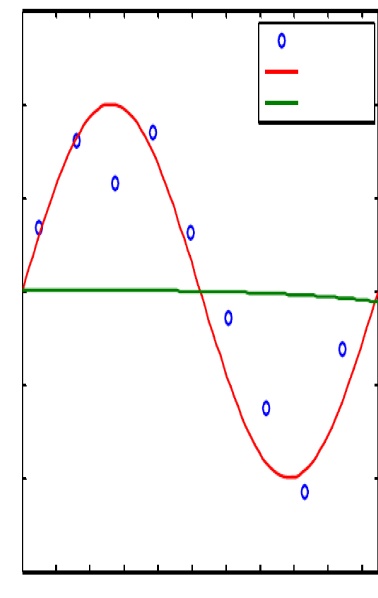
Φ: x→Φ(x) R→RM+1

1



w是一个M+1维向量

N = 9个样本，M = 7



0.5倍

1．5

1

样本pointsIdeal适合

λ= 100

0．5

0

-0.5

－1

-1.5

0

0．1

0．2

0．3

0．4

0．6

0．7

0．8

0．9

1

y

1．5



1

0．5

0

-0.5

－1

-1.5

0

0．1

0．2

0．3

0．4

0.5倍

样本点理想适合lambda = 0.001

y

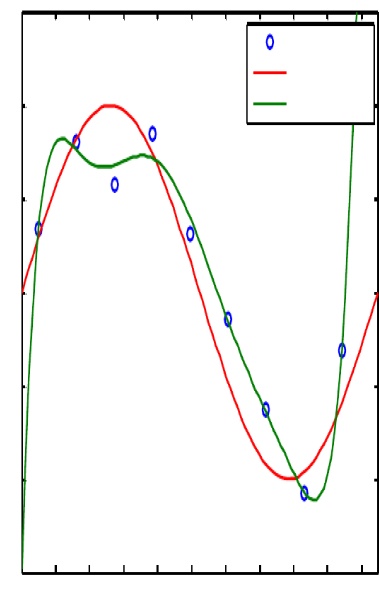
1

0．9

0．8

0．7

0．6



0.5倍

1．5

1

样本pointsIdeal fitlambda = 1e-010

0．5

0

-0.5

－1

-1.5

0

0．1

0．2

0．3

0．4

0．6

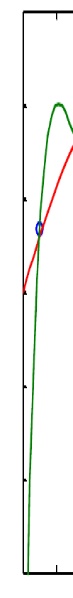
0．7

0．8

0．9

1

y



1．5

1

0．5

0

-0.5

－1

-1.5

0

0．1

0．2

0．3

0．4

0.5倍

样本点理想适合lambda = 1e-015

y

0．6

0．7

0．8

0．9

1

M = 3

M = 5

最小二乘匹配

最小二乘匹配

1．5

1．5

样本点

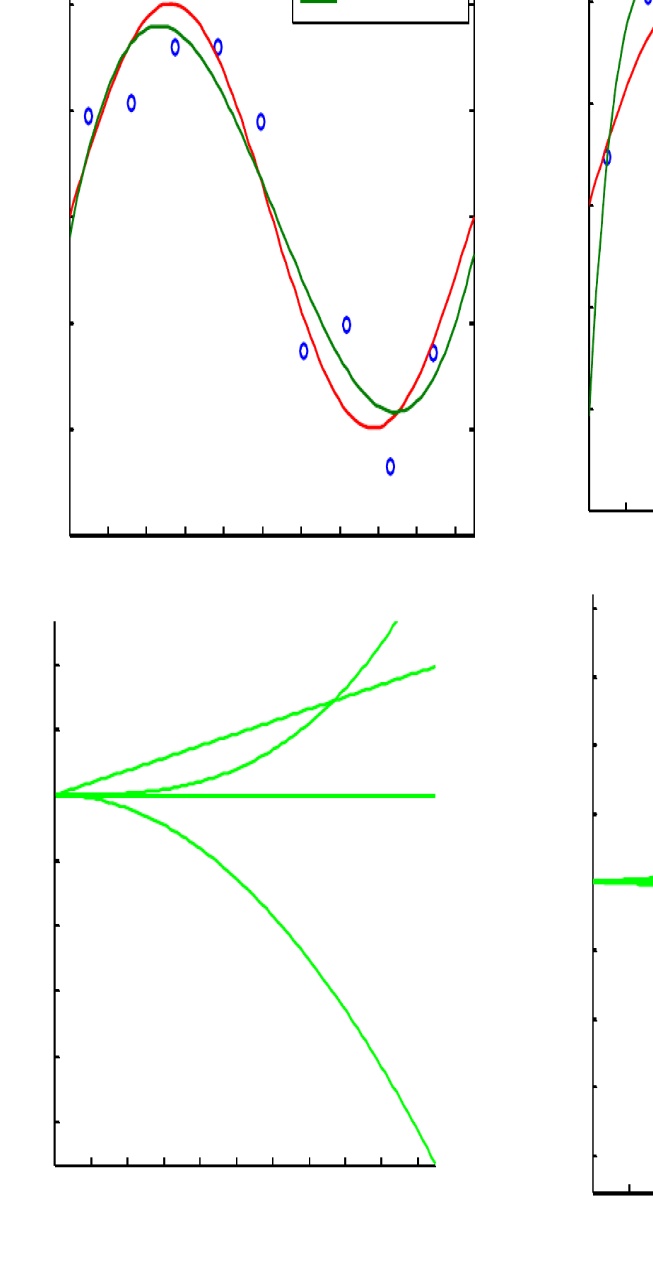
样本点

最小二乘解

1

最小二乘解

1



0．2

0．3

0．2

0．3

0．4

0.5倍

0．6

0．5

0．5

0

y

y

0

-0.5

-0.5

－1

－1

-1.5

15

10

5

0

－5

－10

-15年

－20

-25年

0

多项式基函数

0.3 0.4 0.5 0.6

x

-1.5

0

400

300

200

One hundred.

0

－100

-200年

-300年

-400年

0

0．1

0

0．1

0．2

0．7

0．8

0．9

1

y

y

0．1

0．2

0．3

0．4

0.5倍

0．6

0．7

0．8

0．9

1

0．1

0.4 0.5 0.6 0.7多项式基函数

0．7

0．8

0．8

0．9

0．9

1

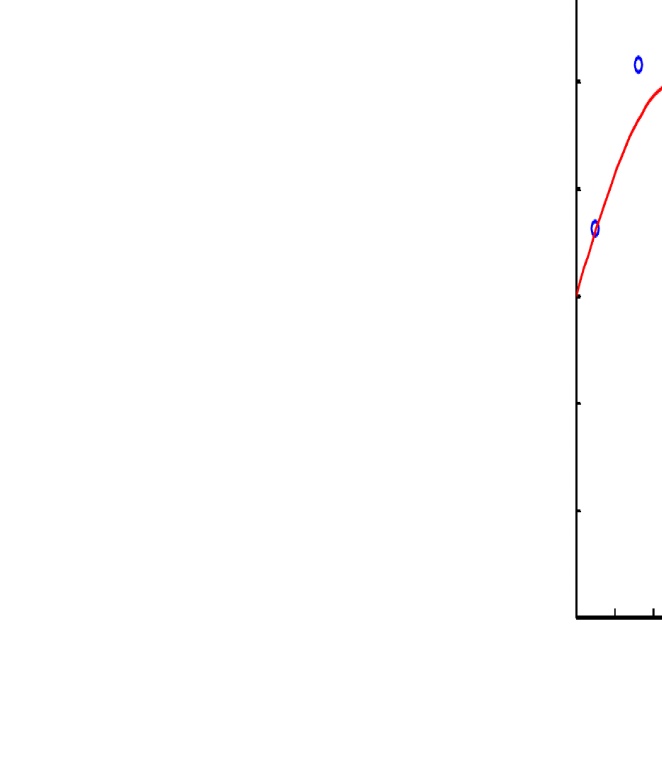
1

例2:高斯基函数

•红色曲线是真正的函数(它是

理想的适合

1．5



1

0．5

•基础函数以培训数据为中心(N点)

•可以选择使用的基函数的尺度和正则化的强度

y

0

-0.5

－1

-1.5

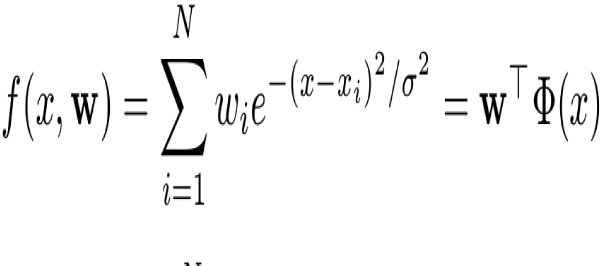
0

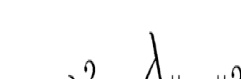
0．1

0．2

不是一个多项式)

•数据点是y中添加噪声的曲线样本。





样本点

0．3

0．4

0.5倍

0．6

0．7

0．8

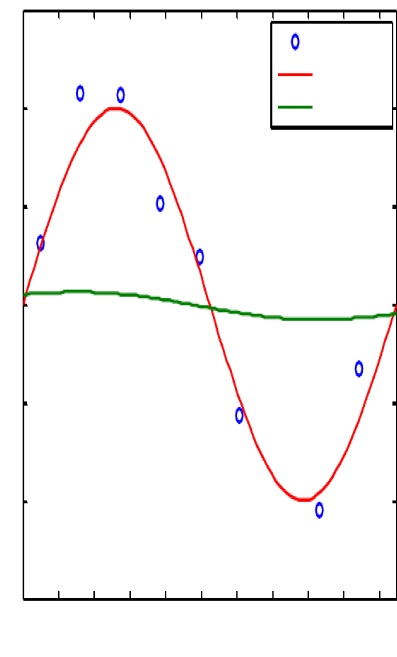
0．9

1

Φ: x→Φ(x) R→RN

w是n个向量

N = 9个样本，∑= 0.334



1．5

1

样本pointsIdeal适合

λ= 100

0．5

0

-0.5

－1

-1.5

0

0．1

0．2

0．3

0．4

0.5倍

0．6

0．7

0．8

0．9

1

y

1．5

1



0．2

0．5

0

-0.5

－1

-1.5

0

0．1

0．3

0．4

0.5倍

0．6

0．7

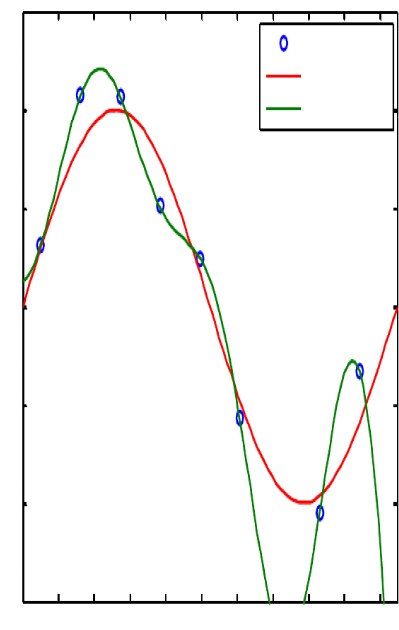
样本点理想适合lambda = 0.001

0．8

0．9

1

y



0.5倍

1．5

1

样本pointsIdeal fitlambda = 1e-010

0．5

0

-0.5

－1

-1.5

0

0．1

0．2

0．3

0．4

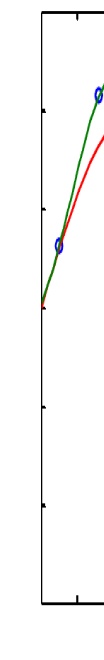
0．6

0．7

0．8

0．9

1



1．5

1

0．5

0

-0.5

y

－1

-1.5 0

0．1

样本点理想适合lambda = 1e-015

y

0．2

0．3

0.4

0.5倍

0．6

0．7

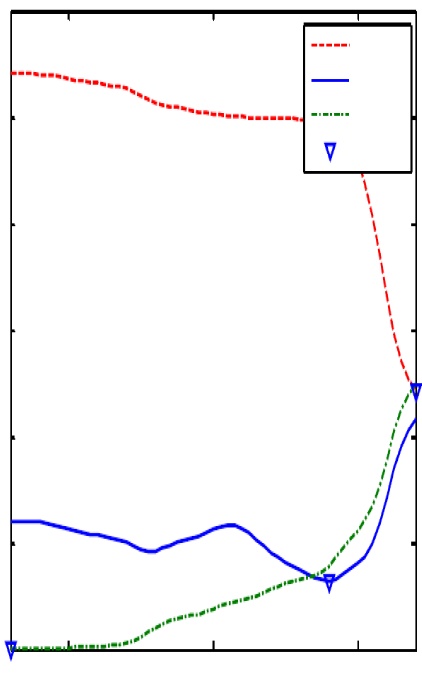
0．8

0．9

1

使用验证集选择lambda

IMAGE



6

5

理想fitValidationTrainingMin错误

4

3.

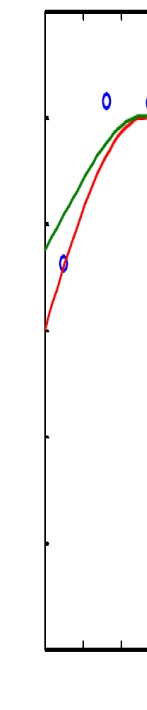
2

1

0



日志



1．5

1

0．5

y

0

-0.5

－1

-1.5

0

0．1

0．2

样本点理想匹配验证集匹配

误差范数

0．3

0.4

0.5倍

0．6

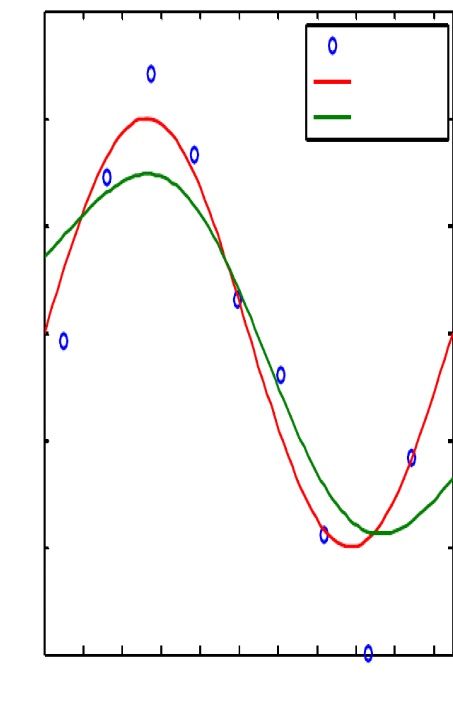
0．7

0．8

0．9

1

FORMULA



1．5

1

样本pointsIdeal fitValidation集适合

0．5

0

y

-0.5

－1

-1.5 0

0．1

0．2

0．7

0．8

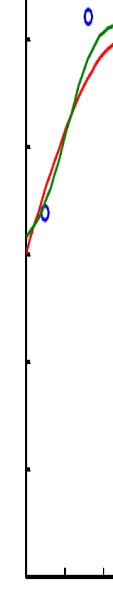
0．9

1

0.3 0.4 0.5 0.6x

y

1．5



1

0．5

0

-0.5

－1

-1.5

0

0．1

0．2

FORMULA

0.4 0.5 0.6 0.7 x

高斯基函数

样本点理想匹配验证集匹配

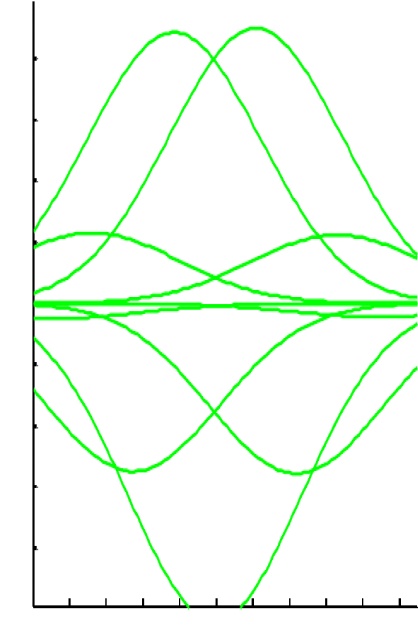
0．3

0．8

0．9

1

高斯基函数



0.5倍

2000

1500

1000

500

0

-500年

-1000年

-1500年

-2000年

0

0．1

0．2

0．3

0．4

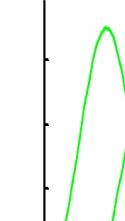
0．6

0．7

0．8

0．9

1



0．8

0．6

0．4

0．2

0

-0.2

-0.4

-0.6

-0.8

y

0

0．1

y

1

0．9

0．8

0．7

0．6

0.5倍

0．4

0．3

0．2

应用:回归面部姿势

IMAGE

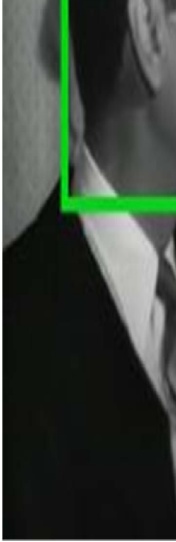
•估计两个面部姿势角度:

•偏航(绕Y轴)

•pitch(围绕X轴)

•计算每个面部区域的HOG特征向量

•学习从HOG向量到两个姿势角度的回归器



总结与对偶问题

IMAGE

到目前为止，我们已经考虑了原始问题，其中XM f(x, w) = wiφi(x) = w>Φ(x) i=1

我们想要w∈RM的解

在支持向量机的情况下，我们也可以考虑对偶问题，其中XN XN w= aiΦ(xi)和f(x, a) = aiΦ(xi)>Φ(x) i=1 i

得到a∈RN的解。

再一次

有一个闭形式的解，

•解涉及到N × N格矩阵k(xi, xj) = Φ(xi)>Φ(xj)，

所以我们可以再次使用核技巧来替换标量积

背景阅读及更多

IMAGE

•Bishop，关于内核和支持向量机的第6和7章

•Hastie等人，第12章

•Bishop，关于回归的第三章

•更多关于网页:

http://www.robots.ox.ac.uk/~az/lectures/ml