**支持向量机**

# 支持向量机概述

Vapnik等人从20世纪60年代开始研究有限样本情况下的机器学习问题，在20世纪90年代初期，有限样本的机器学习理论逐渐成熟起来，形成了一个较为完善的理论体系——统计学习理论。统计学习理论利用结构风险最小化准则，在寻求经验风险最小化的同时，寻求置信范围的最小化，因而具有很好的学习推广能力。支持向量机是建立在统计学习理论基础上的一种机器学习方法，着重研究在小样本情况下的统计学习问题。

支持向量机(Support Vector Machine，SVM)是由前苏联科学家Corinna Cortes和Vapnik等在20世纪90年代首先提出的，它在解决**小样本、非线性及高维模式识别**中表现出许多特有的优势，并能够推广应用到函数拟合等其他机器学习问题中。

支持向量机是一种经典的监督式机器学习算法，可以用来解决分类问题、回归和异常检测等问题。它使用一种叫做核方法(Kernel Trick)的技术来对数据进行转换，把非线性问题转换成线性问题来求解。

对于分类问题，支持向量机在高维或无限维空间中构造超平面或超平面集对样本进行分割，分割的原则是间隔最大化，最终转化为一个凸二次规划问题来求解。 在机器学习中，使用支持向量机的学习算法的监督学习模型，可以分析数据，识别模式，用于分类和回归分析。直观地，通过与任何类的最近训练数据点具有最大距离的超平面（所谓的功能边界）实现良好的分离，因为通常边缘越大，分类器的泛化误差越低。SVM是一种二分类模型，它的目的是寻找一个超平面来

SVM作为一种处理问题的新方法，已经广泛应用于信息处理，自动控制和通信等许多领域。

基于这些转换，它在各种可能的解中找到最优的边界(Optimal Margin)。

最优分隔超平面(Optimal Separating Hyperplane)

# SVM问题类型

## 分类问题

1. C\_SVC：多类别识别问题



1. NU\_SVC：多类别识别问题



## 回归问题

1. EPSILON\_SVR：回归分析



1. NU\_SVR：回归分析



## 异常检测

1. ONE\_CLASS：两类别识别问题



# 符号表

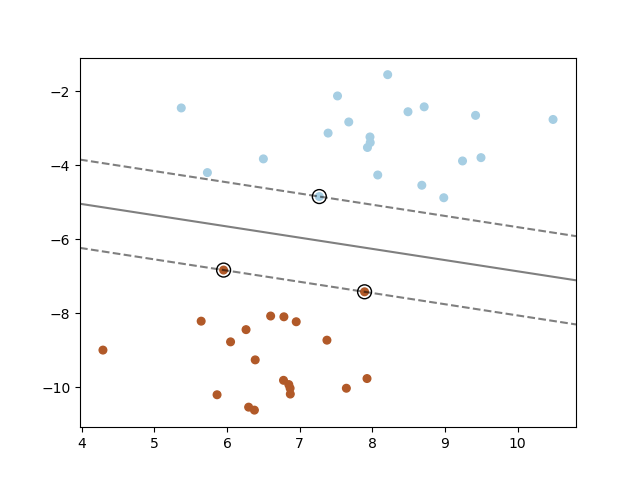
|  |  |
| --- | --- |
| 符号 | 含义 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

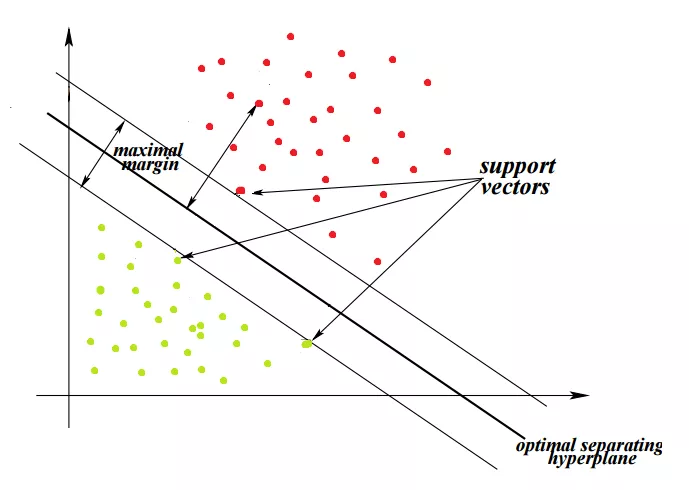
# 支持向量机的推导

## 最大间隔分类器

在分类问题中的训练数据







在二维平面中的直线方程为：



更一般的超平面的方程如下：





代表垂直平面的向量

在一个分类问题中，我们试着找出一个函数

有无数种可能性，必须加一个限制条件，缩小的范围。在SVM的例子中, 必须满足

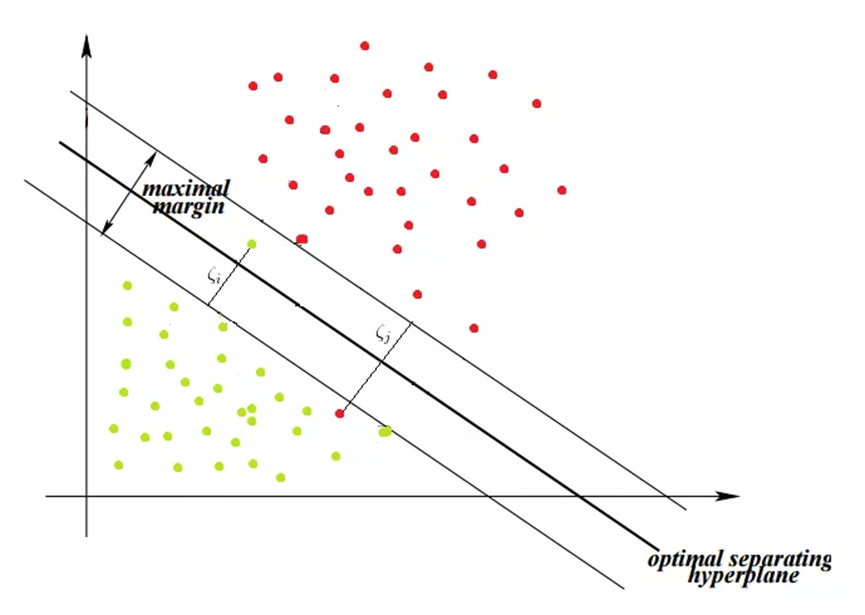
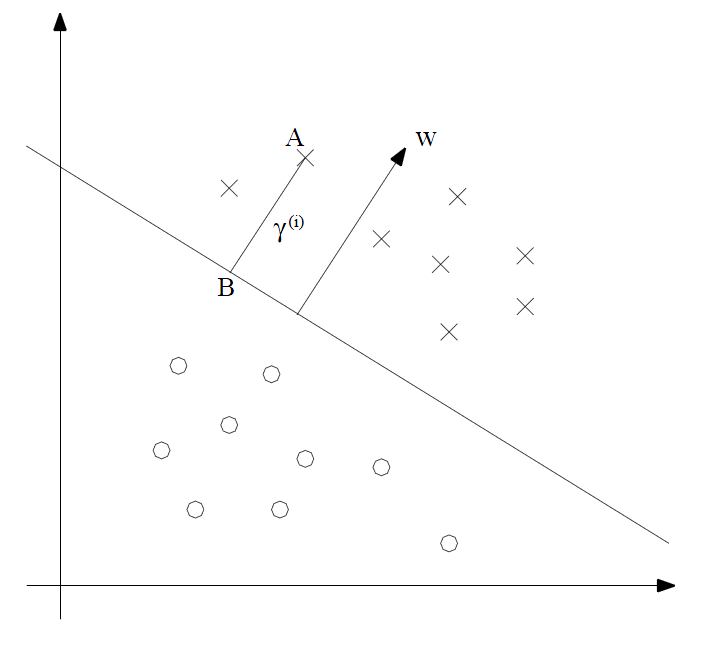
函数间隔

对于单个样本：



对于整个训练集：





几何间隔

对于单个样本：

设A为样本，距离超平面的几何间距为。超平面的法向方向的单位向量为，所以B点坐标为：，又因为B点在超平面上，满足超平面方程，即，整理一下得



几何间隔是归一化权重向量后的函数间隔。

对于整个训练集：



最优间隔分类器

最大几何间距



由几何间距和函数间距之间的关系，得



令



即



## 拉格朗日对偶问题

对于一般的约束优化问题，可以表示为



拉格朗日函数：



定义原问题为：对





定义对偶问题





由性质



如果满足KKT条件，（Karush-Kuhn-Tucker Conditions）



在一定条件下满足



## 最大间隔分类器的求解



改造一下：



拉格朗日函数



一般原问题不好解，并且为了更好地引入核函数，求对偶问题。

对偶问题：先对求最小，再对。



带入拉格朗日函数，恰好消掉



序列最小优化方法求解

由KKT互补条件，最优解一定满足



这意味着仅仅是函数间隔是1的输入点也就是最靠近超平面的点对应的非零，其他点对应的都等于0，这称为解的稀疏性。支持向量包括了重构超平面的所有必要信息，及时移除所有的非支持向量，仍然可以找到相同的最大间隔超平面，这也可以从对偶问题的形式中可以看出，去除非支持向量的行和列，对剩余子矩阵仍有相同的优化问题，因此最优解也是不变的。定理表明，支持向量的数目越小，泛化能力越强。

最优间隔分类器的预测公式



# 支持向量机（Support Vector Machine，SVM）

## 线性可分支持向量机（硬间隔支持向量机）

原问题



当训练数据线性可分是，通过硬间隔最大化，学得一个线性可分支持向量机。

求解的凸二次规划问题



必有解。若，对应的样本为支持向量。



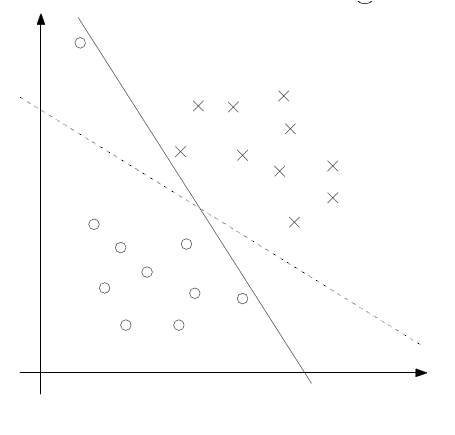
预测公式



支持向量：过两类中离分类线最近的样本且平行于分类线的直线分类间隔最大能最大化模型的泛化能力

## 线性支持向量机（软间隔支持向量机）

当训练数据近似线性可分时，通过软间隔最大化，学得一个线性支持向量机。有时候数据并不是完全线性可分的，或者说数据由于受到噪声的干扰，即使线性可分，但是噪声会使得找到的超平面并不是最优的。对于这个问题我们引入松弛变量(Slack Variable), 允许一些点落到分类间隔之内，但是对于这些点我们进行一些惩罚：



在经验风险和泛化能力之间求得某种平衡，通过引入松弛因子，允许样本错误分类的存在，允许一定程度上违反间隔约束。这类问题就是近似线性可分问题。

当，样本正确分类；

当，样本错误分类，在目标函数中加入惩罚项，惩罚因子；

原问题：



求解的凸二次规划问题

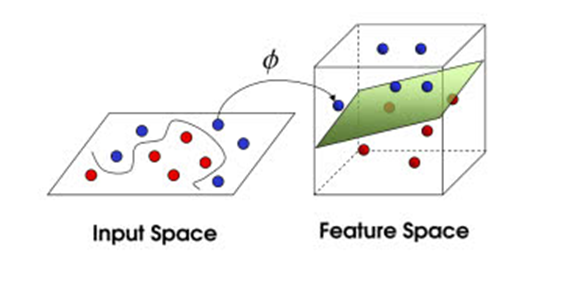


预测公式



## 非线性支持向量机

### C-支持向量机



对于非线性分类问题，一方面引进松弛变量放松约束，另一方面通过特征变换函数将数据样本集中的输入空间映射到空间。当训练数据线性不可分时，通过使用核技巧以及软间隔最大化，学得一个非线性支持向量机。

特征变换函数，将数据样本集中的输入空间映射到高维特征空间，在高维特征空间中构造最优分类面。

原问题：



求解的凸二次规划问题



预测公式



注意到，在对偶问题的凸优化问题和预测公式中，只涉及到高维空间的点积运算，没有单独的的出现，而且特征变换函数映射出来的高维特征空间的维度可能非常高，甚至是无限维的，这导致在高维空间中计算点积消耗资源非常大，甚至是无法计算。如果能够找到一个函数满足：



则高维空间的点积运算转化为用原输入空间的函数计算实现，甚至没有必要知道的具体形式。

根据泛函的相关理论，满足Mercer条件的对称函数都可作为核函数，它就对应于某一变换空间的点积。因此，采用适当的核函数就可以实现某一非线性变换后的线性分类，而计算的复杂度却没有增加。

通过引入核函数，在构造最大间隔超平面中，不是先对输入空间的样本做非线性变换，然后在特征空间求解，而是现在输入空间求点积或者某种距离运算，然后对结果非线性变换。（应该觉得是线性变换）

若对应于特征变换函数的核函数为，则求解的优化问题为



若K是正定核，则必有解



预测公式



SVM分类函数在形式上类似于一个神经网络，中间节点对应于一个支持向量，输出是中间节点（未知向量与支持向量的内积）的线性组合，因此计算复杂度取决于支持向量的个数。

SVM实现的就是一个两层的感知器神经网络，网络的权值和隐藏层节点数由算法自动确定的[2]，不存在局部极小点的问题。

### nu-支持向量机

C-SVM中有两个相互矛盾的目标：最大化间隔和最小化训练误差，参数是正则化参数，控制分类间隔大小和松弛变量大小的平衡。

越小，会使得越过边界的点更容易被忽视，从而使得分类间隔增大。

越大，会使得越过边界的点更难被忽视，从而使得分类间隔减小，更强调最小化训练错误。

本身没有确切的意义，所以出现了nu-SVM。

原问题：



当时，意味着两类点以的间隔被分开

若对应于特征变换函数的核函数为

求解的凸二次规划问题



若K是正定核，则必有解



其中



如果，那么间隔错误数据样本点要么是两个超平面和之间间隔内的点，要么是被求解的最优超平面错分的点。

若，则是间隔错误数据样本的个数所占总数据样本个数的比例上界，还是支持向量的个数所占总数据样本个数的比例下界。

### 最小二乘支持向量机

原问题：



注意到只有等式约束，拉格朗日函数



对求导，得到马鞍点



以为变量，矩阵形式



其中

，

消去，得到



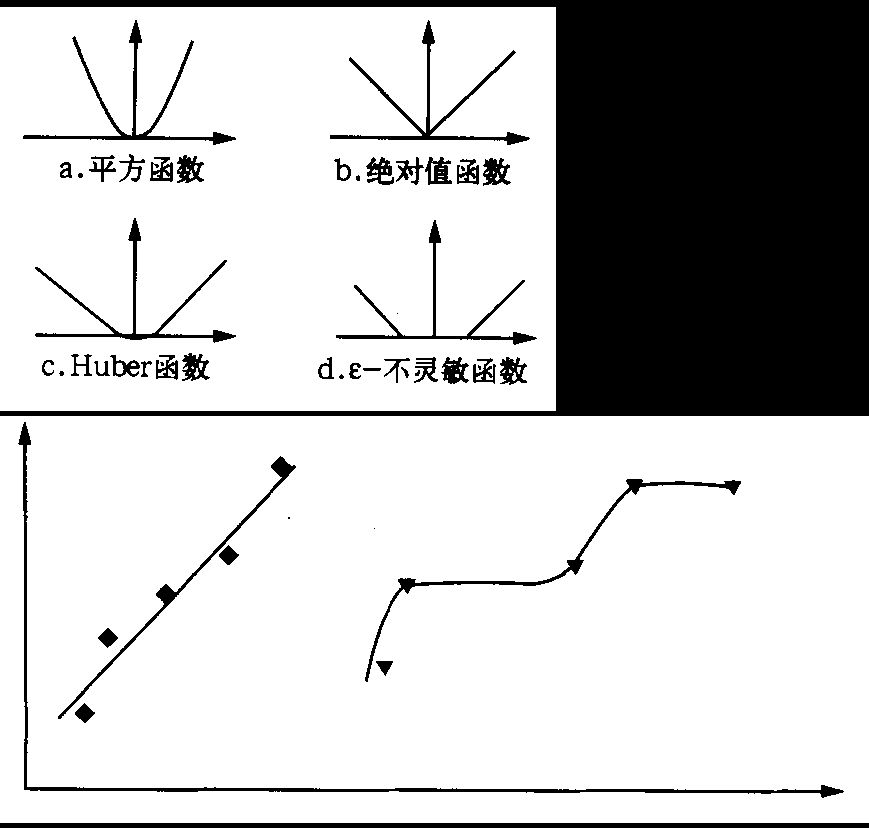
其中



# 支持向量回归（Support Vector Regression，SVR）

Steve1998的称呼

支持向量机与支持向量回归是建立在统计学习理论的VC维理论和结构风险最小化的基础上的，根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折中，以期获得最好的泛化能力。



## 线性支持向量回归

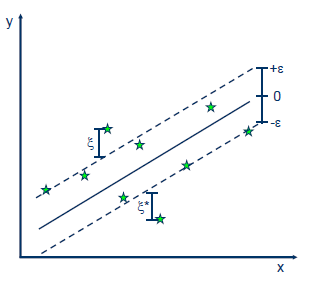
不敏感损失函数

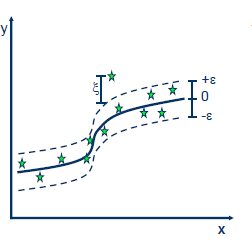
不敏感损失函数，忽略真实值上下某个范围内的误差，它的解以函数的最小化为特征，它可以确保对偶变量的稀疏性，确保全局最小解的存在和可靠泛化界的优化，具有很好的稀疏性[1]，不敏感损失函数区域内没有支持向量（Steve,1998）[1]。二次损失函数不具有稀疏性解。

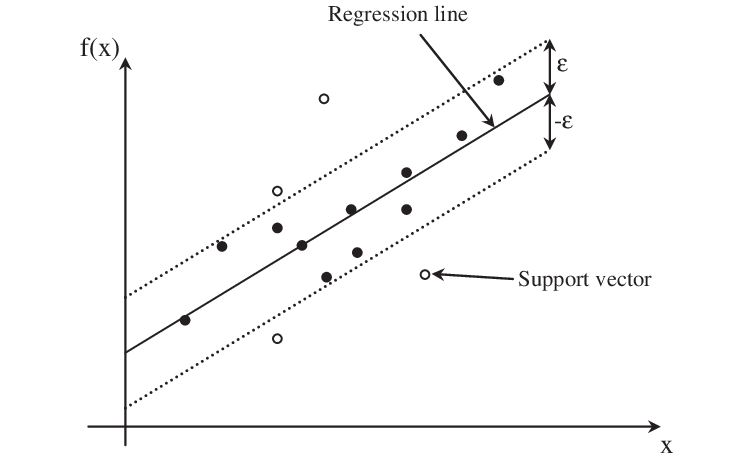


变量度量了训练样本上的误差，越小，回归函数与输出的误差越小，估计精度越高。

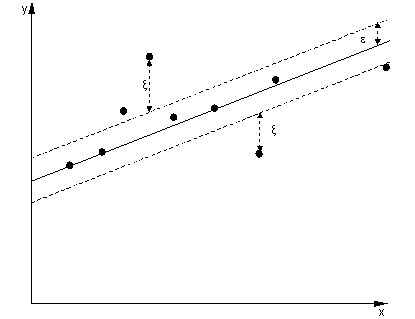
和分类问题一样，回归问题最终也要转化为最优化一个凸函数，并且它的解是稀疏的。

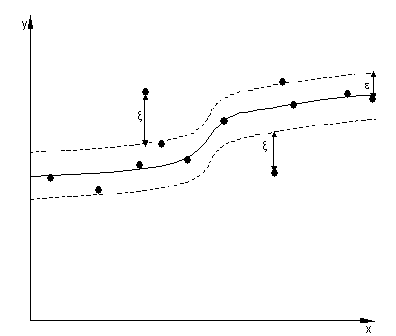




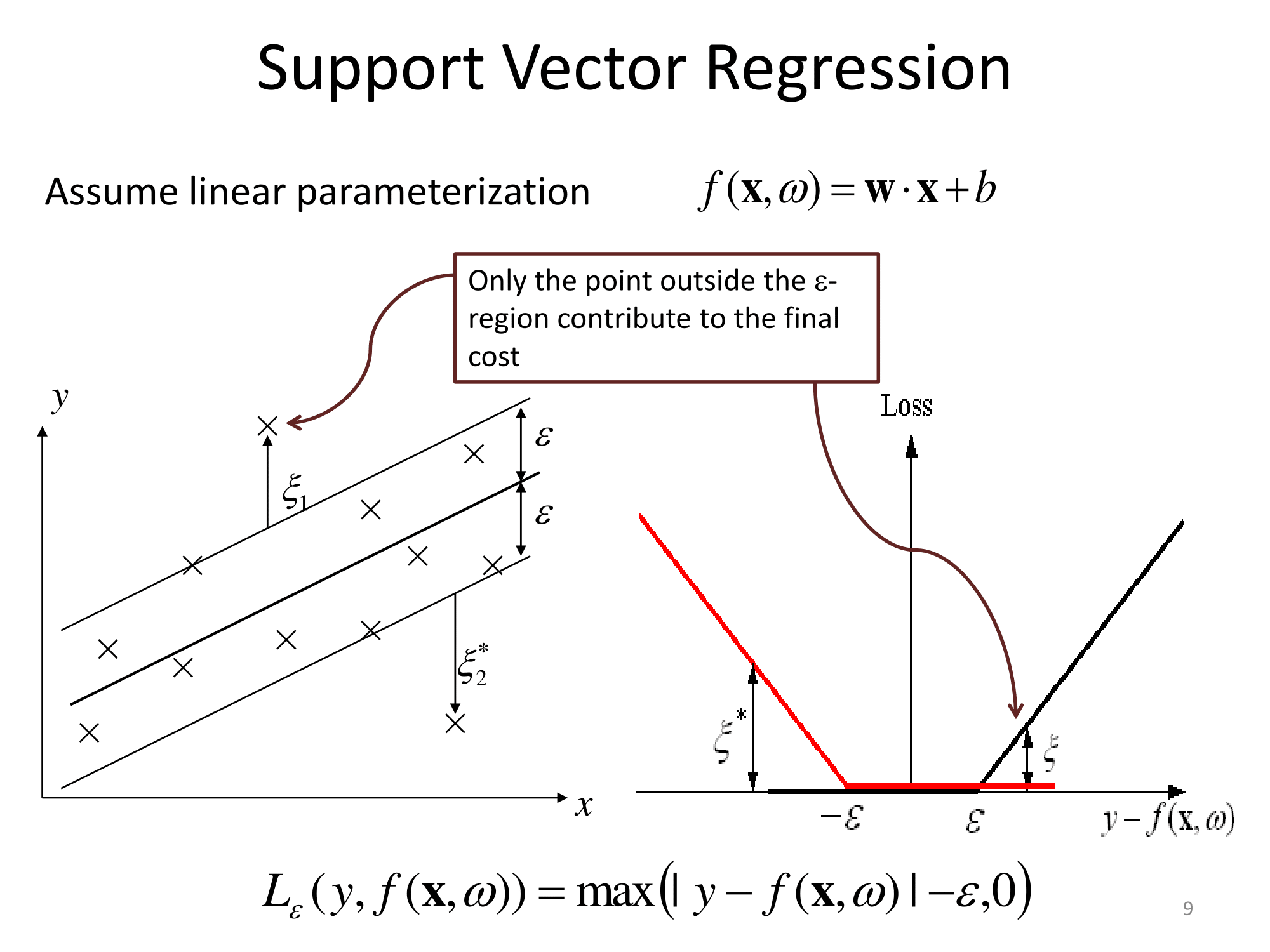


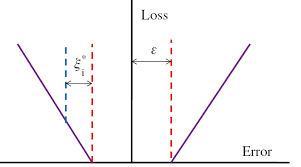
一维线性回归问题的不敏感损失带



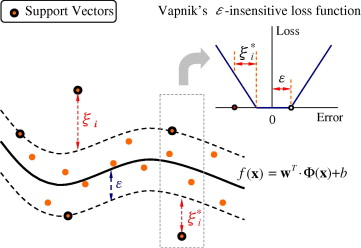


一维非线性回归问题的不敏感损失带









一次不敏感损失函数：

二次不敏感损失函数：

二次不敏感损失函数中，当时，对应着有权重衰减因子的最小二乘回归，又称为岭回归。

设在线性支持向量回归问题中的训练数据



所要拟合的线性函数为



原问题：





第一项使函数更为平坦，从而能够提高泛化能力，第二项是回归的误差。当，误差不计入损失函数，当时，损失误差计入.



对最小化，对最大化。



对偶问题



必有解



由KKT条件，对于，满足



由KKT条件和求导可知



由此可见，不能同时非零，至少有一个是0或者全为0，即满足

当或者时，可能大于，

当或者时，和（两个必须都等于0），即（KKT）

当或者时，或者，即。



预测公式



，即至少其中一个不是0所对应的数据样本是支持向量，应该尽量让支持向量位于预测管道壁上，而非管道外。

## 非线性支持向量回归

### C-支持向量回归

非线性映射函数，将数据样本集中的输入空间映射到高维特征空间（即空间），在高维特征空间中进行线性回归，对应于低维输入空间的非线性回归。其具体实现是通过核函数来实现的。

原问题：



对偶问题



必有解



预测公式



Vapnlik等（1997）之处，满足Mercer条件的人和对称的核函数都对应于一个特征空间的点积[1]。

### nu-支持向量回归

为了解决C-支持向量回归中参数的选取问题，可采用nu-支持向量回归自动选择参数，即把当成变量对待。

原问题



拉格朗日函数



对最小化，对最大化。



对偶问题



预测公式



若，则是误差大于的数据样本个数所占总数据样本个数的比例上界，还是支持向量的个数所占总数据样本个数的比例下界。

### 最小二乘支持向量回归

在回归问题中的训练数据





原问题：



注意到只有等式约束，拉格朗日函数



对求导，得到马鞍点



以为变量，写成向量形式为



矩阵形式



其中

，

消去，得到



其中



带人拉格朗日函数，消去和

求解的凸二次规划问题





由于没有使用不敏感损失函数，最小二乘支持向量机的解不具有稀疏性

**参考文献**

[1] 王定成. 支持向量机建模预测与控制[M]. 北京: 气象出版社, 2009.

[2] 白鹏，张喜斌，张斌，等. 支持向量机理论及工程应用实例[M]. 西安: 西安电子科技大学出版社, 2008.

**校对报告**

当前使用的样式是 [Chinese Physics]

当前文档包含的题录共7条

有0条题录存在必填字段内容缺失的问题

所有题录的数据正常

# SVM核类型: 特征空间的隐式映射

1. 线性核函数SVM::LINEAR :



线性核，主要用于**线性可分**的情况，我们可以看到特征空间到输入空间的维度是一样的，在原始空间中寻找最优线性分类器，具有**参数少，速度快**的优势。对于线性可分数据，其分类效果很理想，因此我们通常首先尝试用线性核函数来做分类，看看效果如何，如果不行再换别的。

1. 多项式核函数SVM::POLY :



多项式核函数可以实现将**低维**的输入空间映射到**高维**的特征空间。多项式核适合于**正交归一化**（向量正交且模为1）数据。属于**全局核函数**，允许相距很远的数据点对核函数的值有影响。参数d越大，映射的维度越高，计算量就会越大。但是多项式核函数的**参数多**，当多项式的阶数d比较高的时候，由于学习复杂性也会过高，易出现“过拟合"现象，核矩阵的元素值将趋于无穷大或者无穷小，计算复杂度会大到无法计算。

1. 高斯径向基核函数SVM::RBF :





径向基函数是指取值仅仅依赖于特定点距离的实值函数，也就是。任意一个满足特性的函数都叫做径向基函数，标准的一般使用欧氏距离，尽管其他距离函数也是可以的。所以另外两个比较常用的核函数，幂指数核，拉普拉斯核也属于径向基核函数。此外不太常用的径向基核还有ANOVA核，二次有理核，多元二次核，逆多元二次核。

高斯径向基函数是一种**局部性强**的核函数，其可以将一个样本映射到一个**更高维**的空间内，该核函数是**应用最广**的一个，无论**大样本还是小样本**都有比较好的性能，而且其相对于多项式核函数**参数要少**，因此大多数情况下在不知道用什么核函数的时候，优先使用高斯核函数。

径向基核函数属于**局部核函数**，当数据点距离中心点变远时，取值会变小。高斯径向基核对数据中存在的噪声有着较好的抗干扰能力，由于其很强的局部性，其**参数决定了函数作用范围**，随着参数的增大而减弱。

幂指数核（Exponential Kernel）



拉普拉斯核（Laplacian Kernel）



二次有理核（Rational Quadratic Kernel）



多元二次核（Multiquadric Kernel）



逆多元二次核（Inverse Multiquadric Kernel）



1. Sigmoid核函数SVM::SIGMOID :



采用Sigmoid函数作为核函数时，支持向量机实现的就是一种多层感知器神经网络。

# 核函数的选择

## 先验知识

一是利用专家的先验知识预先选定核函数

## 交叉验证

二是采用Cross-Validation方法，即在进行核函数选取时，分别试用不同的核函数，归纳误差最小的核函数就是最好的核函数．如针对傅立叶核、RBF核，结合信号处理问题中的函数回归问题，通过仿真实验，对比分析了在相同数据条件下，采用傅立叶核的SVM要比采用RBF核的SVM误差小很多．

## 混合核函数

三是采用由Smits等人提出的混合核函数方法，该方法较之前两者是目前选取核函数的主流方法，也是关于如何构造核函数的又一开创性的工作．将不同的核函数结合起来后会有更好的特性，这是混合核函数方法的基本思想

## 吴恩达的课

曾经给出过一系列的选择核函数的方法，Andrew的说法是：

1.当样本的特征很多时，特征的维数很高，这是往往样本线性可分，可考虑用线性核函数的SVM或LR（如果不考虑核函数，LR和SVM都是线性分类算法，也就是说他们的分类决策面都是线性的）。

2.当样本的数量很多，但特征较少时，可以手动添加一些特征，使样本线性可分，再考虑用线性核函数的SVM或LR。

3.当样特征维度不高时，样本数量也不多时，考虑用高斯核函数（RBF核函数的一种，指数核函数和拉普拉斯核函数也属于RBF核函数）。

Mercer条件



多项式核函数和高斯径向基核函数总是满足Mercer条件，Sigmoid核函数只有在选择特定的参数时才满足核函数

内核参数

Degree:内核函数（POLY）的参数Degree。

Gamma:内核函数（POLY/ RBF/ SIGMOID）的参数。

Coef0:内核函数（POLY/ SIGMOID）的参数Coef0。

Cvalue:SVM类型（C\_SVC/ EPS\_SVR/ NU\_SVR）的参数C。

Nu:SVM类型（NU\_SVC/ ONE\_CLASS/ NU\_SVR）的参数 。Nu的范围是0到1，还有Nu是错分样本所占比例的上界，支持向量所占比列的下界。

P:SVM类型（EPS\_SVR）的参数。

Class\_Weights:C\_SVC中的可选权重，赋给指定的类，乘以C以后变成 。所以这些权重影响不同类别的错误分类惩罚项。权重越大，某一类别的误分类数据的惩罚项就越大。

Term\_Crit:SVM的迭代训练过程的中止条件，解决部分受约束二次最优问题。您可以指定的公差和/或最大迭代次数。

---------------------

作者：Xiake001

来源：CSDN

原文：Https://Blog。Csdn。Net/Xiake001/Article/Details/77509023

版权声明：本文为博主原创文章，转载请附上博文链接！

# SVM的优缺点

支持向量机的优点是：

* 支持向量机算法求解的目标函数的凸优化本质保证了找到最优解的可行性，而且这个最优解是全局最优解；
* 模型简单，具有很好的推广能力
* 对于高维数据和低维数据，SVM都有效, 即使在高维数据下SVM也能高效工作；
* 因为SVM的特性只由支持向量数目确定，而不是整个数据集的大小，因此它具有内存效率；在样本量很少时，也能够抓住数据之间的非线性关系；
* 在特征数量大于样本数量的情况下仍然有效；
* 通用性好， SVM既适用于线性问题也适用于非线性问题，可以为决策功能指定不同的核函数；

支持向量机的缺点包括：

* SVM对缺失数据敏感，对非线性问题没有通用解决方案，必须谨慎选择核函数来处理，计算复杂度高。
* 基于二次规划的优化计算方法，不适合大数据集训练，因为训练时间很长，而且对计算性能要求较高；
* 对于有重叠类别的噪声数据，SVM不是很有效

# 在线支持向量机

在线训练的一个突出特点是支持向量机回归的学习不是一次离线进行的，而是数据逐一加入，不断进行的优化过程，而常规的支持向量机训练是所有的样本都一次性的加入或者批量加入并进行训练[1]。

# 模型选择与参数调节

## 模型选择

1. 交叉验证（Cross Validation）
2. K折交叉验证（k-fold Cross Validation）

把个训练样本随机地分成个互不相交的子集，即，每个子集的大小大致相等。然后进行次训练和测试。即第次选择作为测试集，其他子集作为训练集。算法根据这个训练集求出决策函数后，对测试集进行测试，得到平均测试误差。经过次循环后，便得到，对这些误差求和得到K折交叉验证误差，把该值作为算法的预测精度。一般情况下，取。若取，称为留一法（Leave One Out，LOO），此时的K折交叉验证误差分别称为10-折交叉验证误差和LOO误差。

## 参数调节

惩罚因子C的大小直接影响到误差的大小，但是很难用理论方法确定。参数调节的准则是：检查某特定加权的修正是否确实减小了误差。如果产生了超调，那么C就减小。若连续几步迭代都是降低误差，那么C就应该增大。经验表明，C的增加量最好是一个常数，C的减小量应按集合规律选取。自适应关系如下所示[2]：



在选用高斯径向基核函数进行支持向量回归时，要首先指定三个参数，即目标函数中的正则化参数C，不敏感损失函数中的参数和径向基核函数的宽度。正则化参数C控制着经验风险和VC维的平衡，一般都取一个极大的数来降低误差，以取得最训练样本较好地拟合，参数控制着拟合管道的宽度，也就是误差的边界。

核函数中的控制着支持向量机对输入量变化的敏感程度，过大的会使SVM反应迟钝，不能随着输入的变化迅速调整，过小的对输入过于敏感。

若确定和后，调整C对结果影响不太大，应该取较小的C，避免过大的C引起经验误差和VC维失去平衡，导致VC维急剧增大，泛化能力下降。

惩罚因子C决定了对超出误差的样本的惩罚程度。从结构风险的角度考虑，C值取得越大，问题的求解越倾向于经验风险最小，湖绿对结构复杂程度的考虑，反之则更多地考虑了问题的复杂程度，忽略经验数据的作用，因此C是支持向量回归和泛化能力的平衡参数。当C大于一定的值时，起变化对分析结果产生的影响变小。

# 支持向量机与神经网络

支持向量机的优点

1. 有完备的理论基础。统计学习理论和优化方法
2. 具有全局最优解
3. 采用结构风险最小化，解决了神经网络过学习问题
4. 针对小样本的建模问题。传统的建模方法实质是针对样本无穷的建模方法，实际实现上有困难。

神经网络的缺点

1. 神经网络结构的确定（隐藏层数目和隐藏层节点数目）
2. 局部极小点的问题
3. 过学习的问题（经验风险最小化）

神经网络的优点

1. 非线性映射能力强
2. 适应能力强
3. 不需要对象模型的先验知识
4. 具有学习能力

非线性优化方法：解析法（梯度信息），随机法与枚举法。

预测模型的功能是根据对象的历史信息和未来输入预测其未来输出。

采用径向基核函数的支持向量机与径向基神经网络结构形式上相似，但是径向基神经网络权值的初值和结构的选取决定了训练的速度和网络的收敛特性，支持向量机通过有约束的二次规划确定网络的权值和隐节点的个数，隐节点对应于支持向量。

# 岭回归

二次不敏感损失函数：

二次不敏感损失函数中，当时，对应着有权重衰减因子的最小二乘回归，又称为岭回归。





拉格朗日函数



对最小化，对最大化。



对偶问题



写成向量形式



其中

对求导置零



得



必有解



# Accord支持向量机

高斯内核需要调整适当的σ值。 该问题的不同方法包括使用强力（即使用网格搜索算法）或梯度上升优化。

高斯径向基核函数：





Accord支持向量机采用高斯径向基核函数时，构造函数需要输入的参数是，自动计算。

高斯核函数类的属性

Gaussian. Sigma

Gaussian. SigmaSquared

Gaussian. Gamma

Gaussian.

Gaussian.

Gaussian.

Gaussian.

Gaussian.

Gaussian.

Gaussian.

Gaussian.

Gaussian.

Gaussian.

Gaussian.





