深入浅出核函数

原创 石头 机器学习算法那些事 2018-11-13

前言

支持向量机是最重要的机器学习算法之一，支持向量机的一个重要特点是通过核函数进行非线性分类。本文深度剖析了核函数的含义，并从该角度去理解线性回归和非线性分类的问题。

目录

线性回归的核函数表示

我们先通过构建最优线性回归模型来引出核函数的表达式。

已知N个样本数据$\{x_i, t_i\}$，$i = 1,2,\cdots,N$，构建最优线性回归模型。

损失函数:

$$J(\vec{w}) = \frac{1}{2}\sum_{n=1}^{N}\{\vec{w}^T\overrightarrow{\phi(x_n)} - t_n\}^2 + \frac{\lambda}{2}\vec{w}^T\vec{w} \qquad (1.1)$$

假设$\overrightarrow{\phi(x_n)}$是$M$维的基函数列向量,$\overrightarrow{\phi(x_n)} = (\phi_0(x_n), \phi_1(x_n), \cdots, \phi_{M-1}(x_n))^T$

令$\frac{\partial J(\vec{w})}{\partial \vec{w}} = 0$，得:

$$\vec{w} = -\frac{1}{\lambda}\sum_{n=1}^{N}\{\vec{w}^T\overrightarrow{\phi(x_n)} - t_n\}\overrightarrow{\phi(x_n)} = \sum_{n=1}^{N} a_n\overrightarrow{\phi(x_n)} = \Phi^T\vec{a} \qquad (1.2)$$

其中:

$$\vec{a} = (a_1, a_1, \cdots a_n)^T, a_n = -\frac{1}{\lambda}\{\vec{w}^T\overrightarrow{\phi(x_n)} - t_n\}$$

$$\Phi = \begin{pmatrix} \phi_0(x_1), \phi_1(x_1), \ldots, \phi_{M-1}(x_1) \\ \phi_0(x_2), \phi_1(x_2), \ldots, \phi_{M-1}(x_2) \\ \vdots \qquad \vdots \qquad \vdots \\ \phi_0(x_N), \phi_1(x_N), \ldots, \phi_{M-1}(x_N) \end{pmatrix}$$

将(1.2)式代入(1.1)式得:

$$J(\vec{a}) = \frac{1}{2}\vec{a}^T KK\vec{a} - \vec{a}K\vec{t} + \frac{1}{2}\vec{t}^T\vec{t} + \frac{\lambda}{2}\vec{a}^T K\vec{a} \qquad (1.3)$$

其中K为格拉姆矩阵，$K = \Phi^T\Phi$

$$K_{nm} = \overrightarrow{\phi(x_n)}^T\overrightarrow{\phi(x_m)}$$

令$\frac{\partial J(\vec{a})}{\partial \vec{a}} = 0$,得

$\vec{a} = (K + \lambda I_N)^{-1}\vec{t}$，其中,$I_N$为单位向量

因此，对于新的输入变量x，目标变量为:

$$y(x) = \vec{w}^T\overrightarrow{\phi(x)} = \vec{a}^T\Phi\overrightarrow{\phi(x)} = \overrightarrow{k(x)}^T(K + \lambda I_N)^{-1}\vec{t} \qquad (1.4)$$

其中,$\overrightarrow{k(x)} = (k_1(x), k_2(x), \ldots, k_N(x))^T$,

$$k_n(x) = k(x_n, x) \qquad (1.5)$$

$$k(x_n, x) = \overrightarrow{\phi(x_n)}^T\overrightarrow{\phi(x)} \qquad (1.6)$$

因此,1.5式就是核函数的表达式,1.4式就是线性回归模型的核函数对偶表示

总结:

（1）1.5式就是传说中的核函数方程，核函数方程的展开式有很多种情况，这一节的核函数方程展开式是基函数向量的内积。

（2）1.4式是线性回归的核函数对偶形式，对偶形式都会包含下面这项：

$$\sum_{n=1}^{N} k(x_n, x)$$

核函数含义解析

我们用1.5式和1.6式来解析核函数的含义。

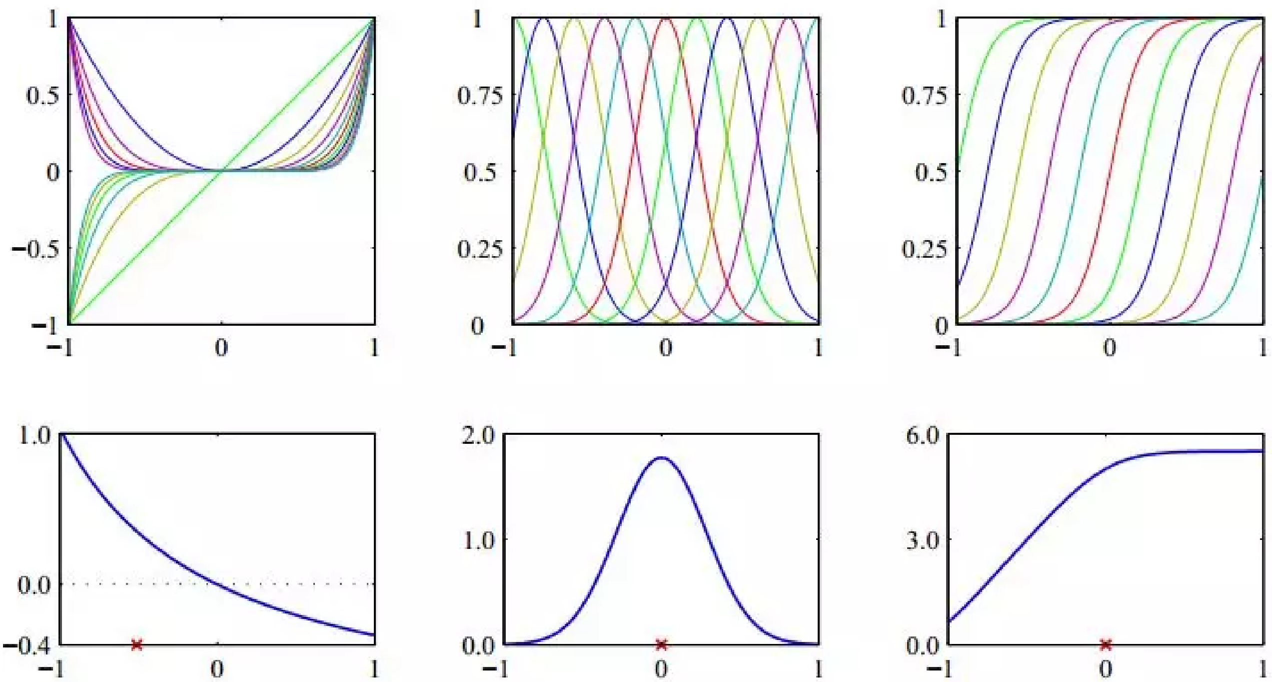
1.5式和1.6式结合，得：

$$k(x,x') = \overrightarrow{\phi(x)}^T \overrightarrow{\phi(x')} = \sum_{i=0}^{M-1} \phi_i(x)\phi_i(x'),\quad 其中\,\phi_i(x)为基函数$$

上式表示，先对输入变量映射为特征空间 $\overrightarrow{\phi(x)}$ ，然后在该特征空间下内积。

现在，我们考虑三种常用的基函数，分别为多项式基函数，高斯基函数，S型基函数。

令输入变量x'为常数，x'的值用红色叉号表示，可以画出x与k(x,x')的关系曲线：



图的第一行为基函数曲线，第二行为x与核函数k(x,x')的曲线。

由上图可知，核函数k(x,x')取得最大值的必要条件是x=x'，核函数值的大小可以表示两个输入变量x和x'的相似度。举例说明这一含义：

> 如上图第二列，归一化的正态分布在x=0具有最大概率，当两个输入变量值都是0时，核函数具有最大值，它们的相似度最高。

如上图第三列，S型函数值表示P(y=1|x)的概率，当两个输入变量都是1时，核函数具有最大值，它们的相似度最高。

因此，核函数的含义可以用相似度来解释，若两个输入变量的特征空间 $\overrightarrow{\phi(x)}, \overrightarrow{\phi(x')}$ 都较大，则两个输入变量在该核函数下具有较高的相似度。

核函数含义理解线性回归

为了阅读方便，1.4式预测目标表达式：

$$y(x) = \vec{w}^T \overrightarrow{\phi(x)} = \vec{a}^T \Phi \overrightarrow{\phi(x)} = \overrightarrow{k(x)}^T (K + \lambda I_N)^{-1} \vec{t} \qquad (1.4)$$

为了更好的定性分析问题，令$(K + \lambda I_N)^{-1}$为常数C

1.4式等价为：

$$y(x) = C * (k_1(x), k_2(x), ..., k_N(x)) * (t_1, t_2, ... t_N)^T$$

$$\Rightarrow y(x) = C * \sum_{n=1}^{N} k(x_n, x) * t_n \qquad (3.1)$$

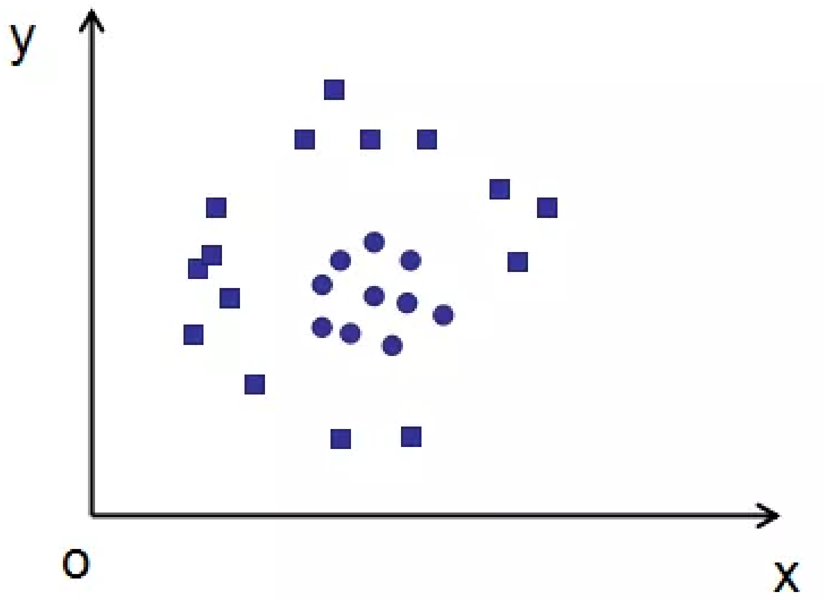
其中$k(x_n, x)$表示核函数，核函数越大，那么相似度越高。

结论

线性回归可以理解成训练目标值的权值相加，权值与核函数成正比，若输入特征与某一训练样本的特征相似度越高，相应的核函数越大，则对应的权值就越大，该训练样本的目标变量对预测目标变量的影响亦越大。

核函数含义理解非线性分类

解决如下图的二分类问题，不同形状表示不同的类。



（1）显示映射法

特征映射是沟通线性分类和非线性分类的桥梁

首先对特征进行非线性映射

$\vec{\phi}=(\phi_0,\phi_1,\phi_2,\phi_3,\phi_4)^T$，即特征由二维映射成五维

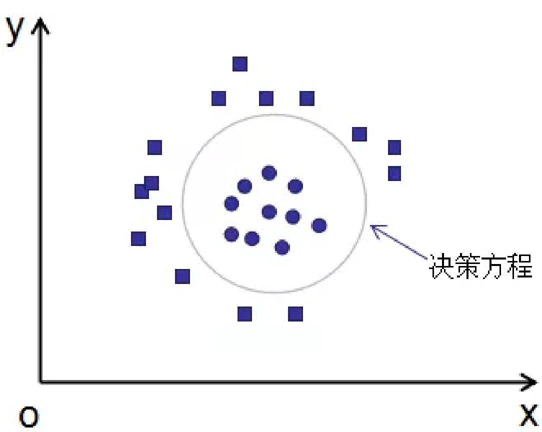
$\Rightarrow \vec{\phi}=(x^2,-2x,y^2,-2y,1)^T$

决策函数：$f(x)=\sum_{i=0}^{4} w_i\phi_i=\vec{w}^T*\vec{\phi}$

令$\vec{w}=(1,1,1,1,1)^T$

$f(x)=(x-1)^2+(y-1)^2$

分类效果图如下：



（2）核函数法

我们知道核函数的含义是相似度，我们考虑用高斯核函数进行分类，高斯核函数用样本的相对距离来表示相似度，若相对距离越小，则相似度越大，反之相似度越小。

高斯核函数的表达式：

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

其中，输入变量x和z的相似度是用x和z的相对距离来衡量

首先用核函数将低维映射成高维空间，然后用线性支持向量机的方法进行分类（后续文章会详细讲支持向量机算法）。

请用核函数法

第二节用显示映射函数的方法说明了核函数的含义，这一节用该方法求决策函数，但是这种方法难点在于定义显示映射函数，而核函数法是隐式的映射特征空间且核函数法计算k(x,x')比较容易，因此推荐核函数法作特征的多维映射。

核函数的应用范围

这一节用一句话来概括：凡是有特征内积出现的式子都可以用核函数来代替。

如下图：

$$K(x_i,x_j)\Rightarrow x_i\bullet x_j$$

左边为核函数，右边为内积。

总结

本文详细的解释了核函数的含义——相似度，然后从核函数的角度去理解线性回归和非线性分类，核函数是隐式的特征空间映射，不需要定义特征映射函数。特征内积出现的式子都可以用核函数来代替，因此，核函数在支持向量机的非线性分类具有不可替代的作用。

参考

李航 《统计学习方法》

Christopher M.Bishop <<Pattern Reconition and Machine Learning>>

推荐阅读文章

[深入理解线性回归算法（一）]

[线性分类模型（一）：浅谈判别模型分析]

-END-
