Multiple Kernel Learning

qhy

2017 年 8 月 22 日

目录

1 Multiple Kernel Learning

1.1 核方法原理

给定样本$\mathcal{D}=\{(\boldsymbol{x_1},y_1),\cdots,(\boldsymbol{x_n},y_n)\}$,求从输入$\boldsymbol{x}$ 到y的映射。

可以先进行一次非线性映射$\Phi:\boldsymbol{x}\rightarrow\phi(\boldsymbol{x})$,把$\boldsymbol{x}$映射到高维空间

然后再在数据新的表示方法$\mathcal{D}'=\{(\boldsymbol{\phi(x_1)},y_1),\cdots,(\boldsymbol{\phi(x_n)},y_n)\}$上考虑原来的机器学习问题。

和函数把非映射和特征空间中的两个向量的内积两步集合起来,使得非线性映射隐式地进行。具体地说,使用核函数来进行高维空间的内积计算,通过对低维空间上的运算来计算高维空间的内积,而不是直接使用高维空间的表示进行内积,避免了维数灾难[1]。

1.2 Mercer条件

Mercer条件:设X是R^n的一个紧子集,$k:X\times X\rightarrow R$是一个连续的对称函数, 如果它在希尔伯特空间上的积分算子满足积分正定条件:

$$\forall f\in L_2(X),\int_{X\times X}k(\boldsymbol{x},\boldsymbol{z})f(\boldsymbol{x})f(\boldsymbol{z})d\boldsymbol{x}d\boldsymbol{z} \tag{1}$$

那么一定存在一个特征空间F和一个映射$\Phi:X\rightarrow F$,使得

$$k(\boldsymbol{x},\boldsymbol{z})=\Phi(\boldsymbol{x})\times\Phi(\boldsymbol{z}) \tag{2}$$

1.3 核函数的性质

- 容许核[2]的正系数线性组合是容许核。
- 容许核的乘积是容许核。
- 函数乘积的积分是容许核。

1.4 多核学习

多核学习:使用多个核的线性或非线性组合来构造新的核,然后使用新的核考虑原来的问题。

以简单地线性组合为例:

新的核函数为:

$$\boldsymbol{K}'=\sum_{i=1}^{n}\beta_i\boldsymbol{K}_i \tag{3}$$

那么对于样本$\boldsymbol{X}$和标签$\boldsymbol{Y}$来说,最小化的问题可以写成以下形式:

$$\min_{\beta,c}E(Y,K'c)+R(K,c) \tag{4}$$

其中,$\boldsymbol{E}$是误差函数,$\boldsymbol{R}$是正则项。

[1]在高维情形下出现的数据样本稀疏,距离计算困难等问题,是所有机器学习方法共同面临的严重障碍,被称为"维数灾难"(curse of dimensionality)

[2]满足Mercer条件的和函数称为容许核

1.4.1 组合方式

多核学习,简单地看,就是多个核的线性和非线性组合(合成出一个新的核),因此,组合的方式是一个重要因素。根据核函数的性质,我们有加和和乘积两个基本方式。

举个例子:

- **pairwise kernels**

$$k((x_{1i},x_{1_j}),(x_{2i},x_{2j}))=k(x_{1i},x_{2i})k(x_{1j},x_{2j})+k(x_{1i,x_{2j}})k(x_{1j},x_{2i}) \tag{5}$$

- **线性组合**

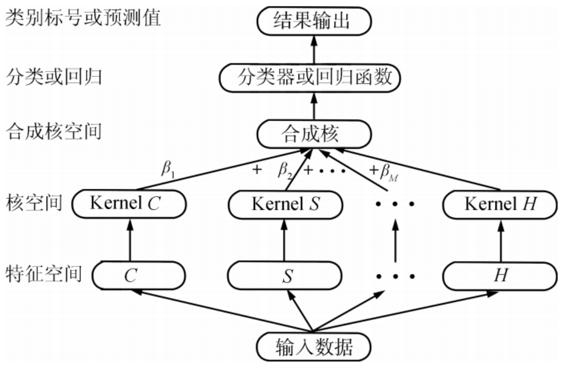

图 1: 多核函数线性组合合成示意图

1.4.2 启发式规则

启发式规则:简单地说,就是给多核增加各种参数,使用学习或者启发式的方法来定义各种参数。

例1:

假设:

- $\boldsymbol{\pi}_m$表示只使用第m个核函数的情况下的精度
- δ表示某个阈值

定义:

$$\beta_m=\frac{\boldsymbol{\pi}_m-\delta}{\sum_{h=1}^{n}(\boldsymbol{\pi}_h-\delta)} \tag{6}$$

例2

定义相似度:

$$A(\boldsymbol{K}_1,\boldsymbol{K}_2)=\frac{<\boldsymbol{K}_1,\boldsymbol{K}_2>}{\sqrt{<\boldsymbol{K}_1,\boldsymbol{K}_1><\boldsymbol{K}_2,\boldsymbol{K}_2>}} \tag{7}$$

则:

$$\beta_m=\frac{A(\boldsymbol{K}_m,\boldsymbol{Y}\boldsymbol{Y}^T)}{\sum_{h=1}^{n}A(\boldsymbol{K}_h,\boldsymbol{Y}\boldsymbol{Y}^T)} \tag{8}$$

为什么是$\boldsymbol{Y}\boldsymbol{Y}^t$

1.4.3 贝叶斯方法

贝叶斯方法:给需要学习的参数一个先验分布

举个例子:

最终预测函数为:

$$f(\boldsymbol{x}) = \sum_{i=0}^{n} \alpha_i \sum_{m=1}^{p} \eta^m \boldsymbol{K}_m(\boldsymbol{x}_i^m, \boldsymbol{x}^m) \tag{9}$$

假设,$\boldsymbol{\eta}$为狄利克雷分布,$\boldsymbol{\alpha}$为高斯分布

1.4.4 Boosting 方法??????

1.5 多尺度核方法

多尺度核方法,使用尺度大的基核函数对平缓变换的样本进行分类,使用尺度小的基和核函数对剧烈变化的样本进行分类。

1.5.1 具有多尺度表示能力的核函数

多尺度核方法的基础就是要找一组能具有多尺度表示能力的核函数,在被广泛使用的核函数中,RBF核

$$k(\boldsymbol{x}, \boldsymbol{z}) = exp(\frac{\|\boldsymbol{x} - \boldsymbol{z}\|^2}{2\sigma^2}) \tag{10}$$

是最受欢迎的。

以此核为例,σ取不同的值表示不同的尺度。σ较大的时,用来对那些平缓变化的样本进行分类,小的时候对剧烈变换的样本进行分类。

k从一定程度上也是描述样本之间的相似度,当σ大的时候,表示相似,因此可以对平缓变化的样本进行分类。

1.5.2 多尺度核的学习方法

考虑两个尺度的核k_1和k_2的分类问题,我们要合成的决策函数为以下形式:

$$f(\boldsymbol{x}) = f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) \tag{11}$$

其中,

$$f_1(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i k_1(\boldsymbol{x}_i, \boldsymbol{x}) + b_1 \tag{12}$$

$$f_2(\boldsymbol{x}) = \sum_{i=1}^{N} \beta_i k_2(\boldsymbol{x}_i, \boldsymbol{x}) + b_1 \tag{13}$$

这里一个尺度的核函数表示一个输出值,最终是值的叠加,而不是前面的核的叠加

这里假设k_1是大尺度的核,k_2是小尺度的核。

首先通过大尺度的单核k_1来构造$f_1(\boldsymbol{x})$来拟合光滑的区域,再在$f_1\boldsymbol{x}$的基础上,使用尺度较小的核k_2构造$f_2\boldsymbol{x}$,这样$f_1(\boldsymbol{x}) + f_2\boldsymbol{x}$比$f_1\boldsymbol{x}$具有更好的拟合性能。