2.10 核方法

CSDN学院

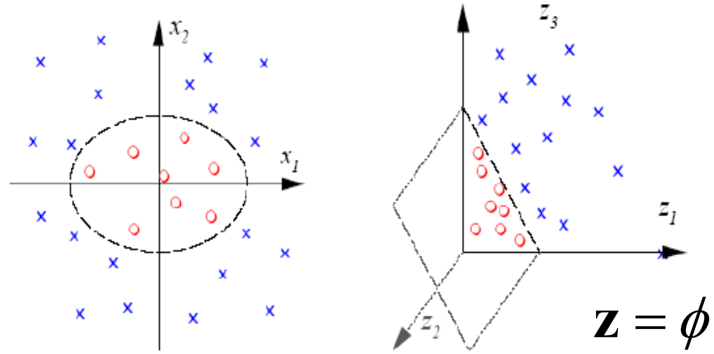
2017年11月

大纲

- SVM基本原理
- 带松弛因子的SVM：C-SVM
- **核方法**
- 支持向量回归
- Scikit learn中的SVM实现
- 案例分析

核方法

- 前面我们假设数据是线性可分的，即存在一个超平面可将训练样本正确分类。但在实际任务中，原始样本空间也许并不存在一个超平面能将训练样本分开。
- 例如：



$$\mathbf{z}=\phi(\mathbf{x})=\left(x_1^2,\sqrt{2}x_1x_2,x_2^2\right)$$

- 对这类问题，我们可以将原始空间映射到一个更高维空间，使得在这个特征空间数据线性可分。

核方法

- 令$\phi(\mathbf{x})$表示将x映射后的特征向量，则在特征空间中划分超平面对应的模型可表示为

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + w_0$$

- 根据之前SVM的推导，

$$\min \left[\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \right]$$

$$\text{subject to} \quad y_i \left(w_0 + \mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}_i) \right) \geq 1 - \xi_i,$$

$$\xi_i \geq 0$$

核方法——对偶

- 将结果代入拉格朗日函数，得到其对偶问题

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left\langle \phi\left(\mathbf{x}_i\right), \phi\left(\mathbf{x}_j\right) \right\rangle$$

- s.t. $\sum_{i=1}^{N} \alpha_i y_i = 0$

$$0 \le \alpha_i \le C$$

- 由于特征空间维数很高（甚至无穷维），直接计算特征空间的点积通常是困难的。
- 解决方案：核函数 $k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left\langle \phi\left(\mathbf{x}_i\right), \phi\left(\mathbf{x}_j\right) \right\rangle$

核方法——对偶

- 利用核函数，可求解得到

$$
\begin{aligned}
f(\mathbf{x}) &= \mathbf{w}^T \phi(\mathbf{x}) + w_0 \\
&= \sum_{j=1}^{N} \alpha_i y_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) + w_0 \\
&= \sum_{j=1}^{N} \alpha_i y_i k\left(\left(\mathbf{x}_i, \mathbf{x}\right)\right) + w_0
\end{aligned}
$$

▶核技巧（Kernel Trick）

- 将问题变为对偶问题：只需计算点积，与特征的维数无关
  - 如在SVM中，最大化下列目标函数

$$Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k\left(\mathbf{x}_i,\mathbf{x}_j\right)$$

- 高维空间中的点积可写成核(*kernel*)的形式
- 选定核函数，无需计算映射 $\phi(x)$ 就可以计算点积
  - 如将SVM核化，目标函数为

$$Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \left\langle\mathbf{x}_i,\mathbf{x}_j\right\rangle$$

  - 预测为

$$f(\mathbf{x})=\sum_{j=1}^{N}\alpha_i y_i k\left(\mathbf{x}_i,\mathbf{x}\right)+w_0$$

▶ 非直接方式构造核函数

- 一种核函数的构造方式是显式地定义一个特征映射 $\phi(\mathbf{x})$，将每个输入x映射到 $\phi(.)$
- 从而得到核函数的间接定义：$k(\mathbf{x},\mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$
  - $\phi(\mathbf{x})$ 称为基函数。

▶ 直接定义核函数

- 另一种可选的方式是直接定义核函数
- 此时需保证函数是有效核。在显式定义特征映射的情况下，核函数为特征空间（可能为无限维）中的内积

核函数

- 令$\mathcal{X}$为输入空间，$k(.,.)$是定义在$\mathcal{X}*\mathcal{X}$上的对称函数，则k是核函数的充要条件是对任意数据 $\mathcal{D}=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$ 核矩阵**K**总是半正定的：

$$\mathbf{K}=\begin{pmatrix} k(\mathbf{x}_1,\mathbf{x}_1) & \cdots & k(\mathbf{x}_1,\mathbf{x}_N) \\ & \vdots & \\ k(\mathbf{x}_N,\mathbf{x}_1) & \cdots & k(\mathbf{x}_N,\mathbf{x}_N) \end{pmatrix}$$

- 对于一个半正定核矩阵，总是能找到一个与之对应的映射ϕ
  - 任何一个核函数都隐式定义了一个再生Hilbert空间

▶ 多项式核

- 多项式核：$k(\mathbf{x},\mathbf{x}')=\left(\gamma\mathbf{x}^T\mathbf{x}'+r\right)^M$
  - 当$M=2,\gamma=1,r=1,$

$$\left(1+\mathbf{x}^T\mathbf{x}'\right)^2=\left(1+x_1x_1'+x_2x_2'\right)^2$$
$$=1+2x_1x_1'+2x_2x_2'+\left(x_1x_1'\right)^2+\left(x_2x_2'\right)^2+2x_1x_1'x_2x_2'$$

$$\phi(\mathbf{x})=\left[1,\sqrt{2}x_1,\sqrt{2}x_2,x_1^2,x_2^2,\sqrt{2}x_1x_2\right]$$

  - 当 $M=1,\gamma=1,r=0,$ 为线性核：

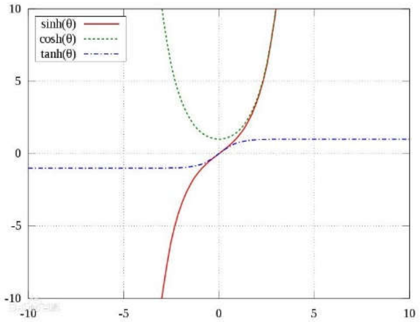
$$\phi(\mathbf{x})=\mathbf{x},$$
$$k(\mathbf{x},\mathbf{x}')=\phi(\mathbf{x})^T\phi(\mathbf{x}')=\mathbf{x}^T\mathbf{x}'$$

Sigmoid核

- 另一个核函数的例子是sigmoid核：

$$k(\mathbf{x}, \mathbf{x}') = \tanh\left(\gamma \mathbf{x}^T \mathbf{x}' + \theta\right)$$

- 其Gram矩阵通常不是p.s.d.
- 但在实际应用中有用到
  - 可能是因为其核展开形式（如SVM）与神经网络十分类似



▶ RBF核

- 平移不变核：$k(\mathbf{x},\mathbf{x}')=\kappa(\mathbf{x}-\mathbf{x}')$
- 径向基函数 (RBFs)

$$k(\mathbf{x},\mathbf{x}')=\kappa(\|\mathbf{x}-\mathbf{x}'\|)=\kappa(r)$$

  - 指数平方(Squared exponential, SE)/高斯核

$$k(\mathbf{x},\mathbf{x}')=\exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}')^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{x}')\right)$$

  - 当Σ为对角阵时，$k(\mathbf{x},\mathbf{x}')=\exp\left(-\frac{1}{2}\sum_{j=1}^{D}\frac{(x_j-x_j')^2}{\sigma_j^2}\right)$
  - 当Σ为球形时，得到各向同性核：$k(\mathbf{x},\mathbf{x}')=\exp\left(-\frac{(\mathbf{x}-\mathbf{x}')^2}{2\sigma^2}\right)$
    - σ称为核函数的带宽

▶ 更多的核

- 核的闭包性质：令 k_1 & k_2 为定义在X×X核，则下列表示也是核：

$$k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$$

$$k(\mathbf{x},\mathbf{z}) = ak_1(\mathbf{x},\mathbf{z})$$

$$k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z})k_2(\mathbf{x},\mathbf{z})$$

$$k(\mathbf{x},\mathbf{z}) = k_1(\phi(\mathbf{x}),\phi(\mathbf{z})),\ \text{ where } \phi : X \to \mathbb{R}^n$$

$$k(\mathbf{x},\mathbf{z}) = \mathbf{x}^T\mathbf{B}\mathbf{z},\ \text{ where } \mathbf{B} \text{ is SPD}$$

$$k(\mathbf{x},\mathbf{z}) = p(k_1(\mathbf{x},\mathbf{z})),\ \text{ where } p \text{ is a polynomial with positive coefficient}$$

$$k(\mathbf{x},z) = \exp(k_1(\mathbf{x},\mathbf{z}))$$

$$k(\mathbf{x},z) = \exp\left(\|\mathbf{x}-\mathbf{z}\|^2/2\right)$$

▶小结：核方法

- 核函数应编码特征相似度的先验信息。
- 核函数的参数需要通过交叉验证选择或通过学习得到 (如 Multiple Kernel Learning).
- 非线性核有时能极大提高分类器的性能。
- 非线性核通常计算繁重（在训练和预测时）。

CSDN
不 止 于 代 码

THANK YOU

AI100