河北师范大学软件学院
Software College of Hebei Normal University

第二部分 非线性分类

Kernel-LDA

Kernel Fisher's Discriminant Analysis (KFDA)

--监督式非线性特征提取(降维)

张朝晖

2018-2019学年 2018年10月

1.分类模块

产生式分类模型 A.贝叶斯分类模型

判别式分类模型

线性分类模型
- B. Fisher判别分类
- C. 感知器分类模型
- D. 大间隔分类模型（线性*SVM*）

非线性分类模型
- E. 核*SVM*（非线性*SVM*）
- F. 核Fisher判别分类
- G. 神经网络（*bp*, *autoencoder*, *som*）

其它分类模型
- *H*.KNN分类模型
- *I*.决策树分类模型
- *J*.*Logistic*回归
- *K*.*Softmax*回归

2.聚类模块
- *L*.K-均值聚类
- *M*.高斯混合聚类
- *N*.DBSCAN聚类
- *O*.层次聚类

3.回归模块
- *P*.KNN回归
- *Q*.回归树
- *R*.最小二乘线性回归
- *S*.岭回归
- *T*.LASSO回归

4.集成学习
- *U*.*Bagging*
- *V*.随机森林
- *W*.*Boosting*
  - （*AdaBoost*, *GBDT*, *XGBoost*, *lightGBM*, *CatBoost*）

5.特征工程
- *X*.主成分分析*(PCA)*
- *Y*.kernel-PCA,ISOMAP,LLE
- *Z*.SOM

6.评价模块 混淆矩阵（及其相关指标）、ROC曲线、交叉验证

核*Fisher*判别是基于**核技巧**，将原始空间的*Fisher***线性判别**向**非线性问题**的推广

$\Rightarrow$ **非线性*Fisher*判别**

1. 回顾*Fisher*线性判别

基本思想 将d维原始空间样本投影到某方向直线上(1维空间)；在该方向投影线上，能最大限度区分各类数据点,最易分类.

$$\begin{cases} \text{类间样本尽可能远离} \\ \text{类内样本尽可能聚集} \end{cases}$$

线性分类器设计 $\Rightarrow$ 寻找最佳投影直线方向

*Fisher*准则函数

原始d维空间样本直接投影到1维空间

$$\begin{cases} \textbf{类间样本尽量分开} \Rightarrow \textbf{投影后类间离散度} \uparrow \\ \textbf{类内样本尽量聚集} \Rightarrow \textbf{投影后总类内离散度} \downarrow \end{cases}$$

$\Rightarrow$

*Fisher*准则

$$\max_{w} J_F(w) = \frac{w^T S_B w}{w^T S_w w}$$

*Fisher*准则函数 $J_F(w)$

$$w^* = \arg\max_{w} J_F(w)$$

若S_w可逆，则**最优投影方向** $w^* = S_w^{-1}(m_1 - m_2)$

2.非线性*Fisher*判别--核*Fisher*判别(KFDA)

$$
\begin{cases}
\boxed{\begin{array}{l}\textbf{原始空间}\boldsymbol{R^d}\\ \text{输入空间}\end{array}} \xrightarrow[\boldsymbol{x}\in\boldsymbol{R^d}\to\Phi(\boldsymbol{x})\in\boldsymbol{F}]{\text{非线性映射}\Phi(\cdot)} \boxed{\begin{array}{l}\textbf{高维空间}\boldsymbol{F}\\ \text{特征空间}\end{array}} \xrightarrow{\boldsymbol{FLDA}} \boxed{\begin{array}{l}\boldsymbol{Fisher}\\ \textbf{投影空间}\end{array}} \\
\textbf{高维空间的}\boldsymbol{Fisher}\textbf{线性判别}
\end{cases}
$$

(1)将训练样本非线性变换

$$\boldsymbol{x} \in \boldsymbol{R}^d \to \Phi(\boldsymbol{x}) \in \boldsymbol{F}$$

(2)在变换空间$\boldsymbol{F}$中的***Fisher*线性判别**

*Fisher*线性判别准则 $\boldsymbol{J}(\boldsymbol{w}) = \dfrac{\boldsymbol{w}^T \boldsymbol{S}_B^{\Phi} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_W^{\Phi} \boldsymbol{w}}$

$$
\text{其中}\begin{cases}
\boldsymbol{w} -- \boldsymbol{F}\text{空间的权值向量}, \boldsymbol{w} \in \boldsymbol{F} \\
\boldsymbol{S}_B^{\Phi} -- \text{训练样本集在}\boldsymbol{F}\text{空间的类间散度矩阵} \\
\boldsymbol{S}_W^{\Phi} -- \text{训练样本集在}\boldsymbol{F}\text{空间总类内散度矩阵}
\end{cases}
$$

$$
\text{若}\begin{cases}
\text{第}\boldsymbol{i}\text{类训练样本数量}\boldsymbol{l_i} \quad \sum_{\boldsymbol{i}} \boldsymbol{l_i} = \boldsymbol{l} \\
\text{第}\boldsymbol{i}\text{类训练样本在}\boldsymbol{F}\text{空间的均值}\boldsymbol{m_i^{\Phi}} \\
\boldsymbol{m_i^{\Phi}} = \frac{1}{\boldsymbol{l_i}} \sum_{j=1}^{\boldsymbol{l_i}} \Phi\left(\boldsymbol{x_j^i}\right) \qquad \boldsymbol{i} = 1, 2 \\
\text{各类先验概率相等：} \boldsymbol{P}\left(\omega_1\right) = \boldsymbol{P}\left(\omega_2\right) = 0.5
\end{cases}
$$

$$
\text{则}\begin{cases}
\boldsymbol{S_B^{\Phi}} = \left(\boldsymbol{m_1^{\Phi}} - \boldsymbol{m_2^{\Phi}}\right)\left(\boldsymbol{m_1^{\Phi}} - \boldsymbol{m_2^{\Phi}}\right)^{\boldsymbol{T}} \\
\boldsymbol{S_W^{\Phi}} = \sum_{\boldsymbol{i}=1}^{2} \sum_{j=1}^{\boldsymbol{l_i}} \left(\Phi\left(\boldsymbol{x_j^i}\right) - \boldsymbol{m_i^{\Phi}}\right)\left(\Phi\left(\boldsymbol{x_j^i}\right) - \boldsymbol{m_i^{\Phi}}\right)^{\boldsymbol{T}}
\end{cases}
$$

(3)转化为**核函数**的求解形式

再生核理论认为，鉴别向量$w \in F$处于F空间所有训练样本所张成的子空间中，即：

$$\boldsymbol{w} = \sum_{j=1}^{l} \alpha_j \Phi\left(\boldsymbol{x}_j\right)$$

原始空间任意观测x在$\boldsymbol{Fisher}$判别方向的投影为：

$$\boldsymbol{w}^T \Phi\left(\boldsymbol{x}\right) = \left(\boldsymbol{w}, \Phi\left(\boldsymbol{x}\right)\right) = \left(\sum_{j=1}^{l} \alpha_j \Phi\left(\boldsymbol{x}_j\right), \ \Phi\left(\boldsymbol{x}\right)\right)$$

$$= \sum_{j=1}^{l} \alpha_j \left(\Phi\left(\boldsymbol{x}_j\right), \Phi\left(\boldsymbol{x}\right)\right) = \sum_{j=1}^{l} \alpha_j \boldsymbol{k}\left(\boldsymbol{x}_j, \boldsymbol{x}\right)$$

$$
\boldsymbol{w}^T\boldsymbol{m}_i^{\Phi}=\boldsymbol{w}^T\frac{1}{\boldsymbol{l}_i}\sum_{k=1}^{l_i}\Phi\left(\boldsymbol{x}_k^i\right)=\left(\sum_{j=1}^{l}\alpha_j\Phi\left(\boldsymbol{x}_j\right)\right)^T\frac{1}{\boldsymbol{l}_i}\sum_{k=1}^{l_i}\Phi\left(\boldsymbol{x}_k^i\right)
$$

$$
=\frac{1}{\boldsymbol{l}_i}\sum_{j=1}^{l}\sum_{k=1}^{l_i}\alpha_j\left(\Phi\left(\boldsymbol{x}_j\right),\Phi\left(\boldsymbol{x}_k^i\right)\right)=\frac{1}{\boldsymbol{l}_i}\sum_{j=1}^{l}\sum_{k=1}^{l_i}\alpha_j\boldsymbol{k}\left(\boldsymbol{x}_j,\boldsymbol{x}_k^i\right)
$$

$$
=\sum_{j=1}^{l}\alpha_j\left[\frac{1}{\boldsymbol{l}_i}\sum_{k=1}^{l_i}\boldsymbol{k}\left(\boldsymbol{x}_j,\boldsymbol{x}_k^i\right)\right]=\sum_{j=1}^{l}\alpha_j\left(M_i\right)_j=\boldsymbol{\alpha}^T\boldsymbol{M}_i \qquad \boldsymbol{i}=1,2
$$

其中

$$
\begin{cases}
\text{核函数}\,\boldsymbol{k}\left(\boldsymbol{x}_j,\boldsymbol{x}_k^i\right)=\left(\Phi\left(\boldsymbol{x}_j\right),\Phi\left(\boldsymbol{x}_k^i\right)\right) \\
\left(M_i\right)_j=\dfrac{1}{\boldsymbol{l}_i}\displaystyle\sum_{k=1}^{l_i}\boldsymbol{k}\left(\boldsymbol{x}_j,\boldsymbol{x}_k^i\right) \\
\boldsymbol{M}_i=\left[\left(M_i\right)_1 \quad \cdots \quad \left(M_i\right)_l\right]^T \qquad \boldsymbol{l}=\boldsymbol{l}_1+\boldsymbol{l}_2 \\
\boldsymbol{\alpha}=\left[\alpha_1 \quad \cdots \quad \alpha_l\right]^T
\end{cases}
$$

规定 $M = \left(M_1 - M_2\right)\left(M_1 - M_2\right)^T$

由前页知：$w^T m_i^\Phi = \alpha^T M_i \qquad i = 1, 2$

投影空间类间离散度

$$w^T S_B^\Phi w = w^T \left(m_1^\Phi - m_2^\Phi\right)\left(m_1^\Phi - m_2^\Phi\right)^T w$$

$$= \alpha^T \left(M_1 - M_2\right)\left(M_1 - M_2\right)^T \alpha = \alpha^T M \alpha$$

投影空间类内离散度

$$w^T S_W^\Phi w = \sum_{i=1}^{2} \sum_{j=1}^{l_i} w^T \left(\Phi\left(x_j^i\right) - m_i^\Phi\right)\left(\Phi\left(x_j^i\right) - m_i^\Phi\right)^T w$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{l_i} \left(w^T \Phi\left(x_j^i\right) - w^T m_i^\Phi\right)\left(w^T \Phi\left(x_j^i\right) - w^T m_i^\Phi\right)^T$$

$$= \alpha^T N \alpha$$

$$\boldsymbol{w}^T \boldsymbol{S}_B^{\Phi} \boldsymbol{w} = \boldsymbol{\alpha}^T \boldsymbol{M} \boldsymbol{\alpha}$$

$$\boldsymbol{w}^T \boldsymbol{S}_W^{\Phi} \boldsymbol{w} = \boldsymbol{\alpha}^T \boldsymbol{N} \boldsymbol{\alpha}$$

$$\begin{cases} \boldsymbol{N} = \sum_{\boldsymbol{i}=1,2} \boldsymbol{K}_{\boldsymbol{i}} \left(\boldsymbol{I} - 1_{\boldsymbol{l}_i} \right) \boldsymbol{K}_{\boldsymbol{i}}^{\boldsymbol{T}} \qquad \boldsymbol{l}\text{行} \times \boldsymbol{l}\text{列} \\ \boldsymbol{K}_{\boldsymbol{i}} -- \boldsymbol{l}\text{行} \times \boldsymbol{l}_{\boldsymbol{i}}\text{列矩阵} \\ \left(\boldsymbol{K}_{\boldsymbol{i}} \right)_{\boldsymbol{nm}} = \boldsymbol{k}\left(\boldsymbol{x}_{\boldsymbol{n}}, \boldsymbol{x}_{\boldsymbol{m}}^{\boldsymbol{i}} \right) -- \boldsymbol{K}_{\boldsymbol{i}}\text{矩阵的第}\boldsymbol{n}\text{行第}\boldsymbol{m}\text{列元素} \\ \boldsymbol{I} -- \text{单位矩阵，} \qquad \boldsymbol{l}_{\boldsymbol{i}}\text{行} \times \boldsymbol{l}_{\boldsymbol{i}}\text{列} \\ 1_{\boldsymbol{l}_i} -- \text{所有元素均为}\dfrac{1}{\boldsymbol{l}_{\boldsymbol{i}}}\text{的矩阵} \end{cases}$$

变换空间**Fisher**线性判别的目标函数

$$J(\boldsymbol{\alpha})=\frac{\boldsymbol{\alpha}^T\boldsymbol{M}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T\boldsymbol{N}\boldsymbol{\alpha}}$$

最大化$J(\boldsymbol{\alpha})$，得最优解方向为 $\boldsymbol{\alpha}^*\propto\boldsymbol{N}^{-1}(\boldsymbol{M}_1-\boldsymbol{M}_2)$

原始空间任意观测$\boldsymbol{x}$在**Fisher**判别方向的投影为：

$$y=(\boldsymbol{w}^*)^T\Phi(\boldsymbol{x})=(\boldsymbol{w}^*,\Phi(\boldsymbol{x}))=\left(\sum_{j=1}^{l}\alpha^*{}_j\Phi(\boldsymbol{x}_j)\right)^T\Phi(\boldsymbol{x})$$

$$=\sum_{j=1}^{l}\alpha^*{}_j(\Phi(\boldsymbol{x}_j),\Phi(\boldsymbol{x}))=\sum_{j=1}^{l}\alpha^*{}_j\boldsymbol{k}(\boldsymbol{x}_j,\boldsymbol{x})$$

(4)决策

STEP 1. 估计各类训练样本点在***Fisher***判别方向投影中心

$$\boldsymbol{f}_i = \boldsymbol{w}^{*T}\boldsymbol{m}_i^{\Phi} = \frac{1}{\boldsymbol{l}_i}\sum_{j=1}^{\boldsymbol{l}} \alpha^*{}_j \sum_{\boldsymbol{k}=1}^{\boldsymbol{l}_i} \boldsymbol{k}\left(\boldsymbol{x}_j, \boldsymbol{x}_{\boldsymbol{k}}^{\boldsymbol{i}}\right) = \boldsymbol{\alpha}^{*T}\boldsymbol{M}_i \quad \boldsymbol{i} = 1, 2$$

STEP 2.确定原始空间任意观测$\boldsymbol{x}$在***Fisher***判别方向投影

$$y = \boldsymbol{w}^{*T}\Phi\left(\boldsymbol{x}\right) = \sum_{j=1}^{\boldsymbol{l}} \alpha^*{}_j \boldsymbol{k}\left(\boldsymbol{x}_j, \boldsymbol{x}\right)$$

STEP 3.决策

$$\begin{cases} \text{若}\left|y - \boldsymbol{f}_1\right| < \left|y - \boldsymbol{f}_2\right|, \text{则决策}\boldsymbol{x} \in \omega_1\text{类} \\ \text{若}\left|y - \boldsymbol{f}_1\right| > \left|y - \boldsymbol{f}_2\right|, \text{则决策}\boldsymbol{x} \in \omega_2\text{类} \end{cases}$$

3.基本核*Fisher*判别方法的改进

由最大化$\dfrac{\boldsymbol{\alpha}^T \boldsymbol{M} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \boldsymbol{N} \boldsymbol{\alpha}}$，得最优解方向为

$$\boldsymbol{\alpha} \propto \boldsymbol{N}^{-1}\left(\boldsymbol{M}_1 - \boldsymbol{M}_2\right)$$

若$\boldsymbol{N}$非正定，则上述求解为病态

改进：

引入新矩阵 $\boldsymbol{N}_\mu = \boldsymbol{N} + \mu \boldsymbol{I}$，使矩阵$\boldsymbol{N}_\mu$正定

其中μ为正常数

思考

核LDA的基本思想、实现步骤？