偏最小二乘回归分析

1.算法步骤

设 p 个因变量集 $Y=[\,y_1,\dots,y_p\,]^T$ 与 m 个自变量集 $X=[\,x_1,\dots,x_m\,]^T$ 均为标准化变量，自变量组和因变量组的 n 次标准化观测数据矩阵分别记为 $A_0,\ B_0$ ，
1）求矩阵 $A_0^TB_0B_0^TA_0$ 最大特征值所对应的特征向量 ω_1 ，求得成分得分向量 $\hat{t}_1=A_0\omega_1$ ，和残差矩阵 $A_1=A_0-\hat{t}_1\alpha_1^T$ ，$B_1=B_0-\hat{t}_1\beta_1^T$ ，其中 $\alpha_1=A_0^T\hat{t}_1\ /\ ||\hat{t}_1||^2$ ，$\beta_1=B_0^T\hat{t}_1\ /\ ||\hat{t}_1||^2$ ；
2）求矩阵 $A_1^TB_0B_0^TA_1$ 最大特征值所对应的特征向量 ω_2 ，求得成分得分向量 $\hat{t}_2=A_1\omega_2$ ，和残差矩阵 $A_2=A_1-\hat{t}_2\alpha_2^T$ ，$B_2=B_1-\hat{t}_2\beta_2^T$ ，其中 $\alpha_2=A_1^T\hat{t}_2\ /\ ||\hat{t}_2||^2$ ，$\beta_2=B_1^T\hat{t}_2\ /\ ||\hat{t}_2||^2$ ；

$$\vdots$$

r ）求矩阵 $A_{r-1}^TB_0B_0^TA_{r-1}$ 最大特征值所对应的特征向量 ω_r ，求得成分得分向量 $\hat{t}_r=A_{r-1}\omega_r$，求得 $\alpha_r=A_{r-1}^T\hat{t}_r\ /\ ||\hat{t}_r||^2$，$\beta_r=A_{r-1}^T\hat{t}_r\ /\ ||\hat{t}_r||^2$ 。

将 $t_k=\omega_{k1}^*x_1+\omega_{k2}^*x_2+\cdots+\omega_{km}^*x_m\ (k=1,2,\dots,r)$ ，代入 $Y=t_1\beta_1+t_2\beta_2+\cdots+t_r\beta_r$ ，即得到 p 个因变量的偏最小二乘回归方程

$$y_j=c_{j1}x_1+c_{j2}x_2+\cdots+c_{jm}x_m,\quad j=1,2,\dots,p,$$

其中 $\omega_k^*=[\,\omega_{k1}^*,\ \omega_{k2}^*,\ \dots,\ \omega_{km}^*\,]$ 满足

$$\hat{t}_k=A_0\omega_k^*\ ,\ \omega_k^*=\prod_{j=1}^{k-1}(I-\omega_j\alpha_j^T)\omega_k.$$

2.Python调用

```
from sklearn.cross_decomposition import PLSRegression

# n_components为PLS成分数量，a,b分别为自变量集和因变量集
md = PLSRegression(n_coponents=2).fit(a, b)
xzh = md.x_loading_  # x主成分回归系数
yzh = md.y_loading_  # y主成分回归系数
beta = md.coef_  # 标准化y关于x的回归系数
```