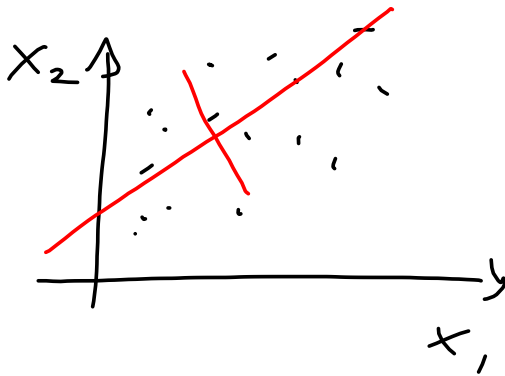


Today: PCR, PLS



Idea of PCR: construct new covariates as linear combinations of the original covariates in a way that:

- try to keep as much information as possible;
- construct orthogonal covariates

Consider a matrix X of data, suppose with mean $\underline{0}$, and covariance Σ
 if $\mu \neq 0$, we can subtract the mean

We want to create a new variable Z
 → linear combination of $X \rightarrow Z = \alpha X$

We want its variance to be largest possible

$$\max_{\alpha} \text{Var}(Z) = \max_{\alpha} \text{Var}(\alpha X) = \max_{\alpha} \alpha^T \text{Var}(X) \alpha = \max_{\alpha} \alpha^T \Sigma \alpha$$

z_1 first principal component by solving in α

$$\alpha_1 = \arg \max_{\alpha} \alpha^T \Sigma \alpha$$

subject to $\|\alpha\| = 1$

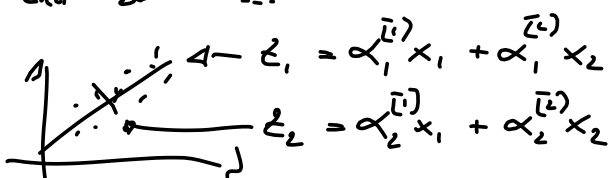
$$z_1 = \alpha_1^T X$$

z_2 second principal component,

$$\alpha_2 = \arg \max_{\alpha} \alpha^T \Sigma \alpha$$

$$\text{subject to } \|\alpha\| = 1 \quad \text{and} \quad \alpha^T \alpha_1 = 0$$

and so on ...



Mathematically, this can be done by the spectral decomposition of the matrix Σ

$$\Sigma = (\alpha_1 \dots \alpha_p) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

eigenvectors eigenvalues

first PC $\rightarrow \alpha_1$
first vector of principal loadings $\rightarrow \alpha_1$

$z_1 = \alpha_1 x$

$\lambda_k, k=1, \dots, p$ is a measure of how much of the original variance is contained in the k -th principal components

- each PC contains $\frac{\lambda_k}{\sum_{i=1}^p \lambda_i}$ of the original variance

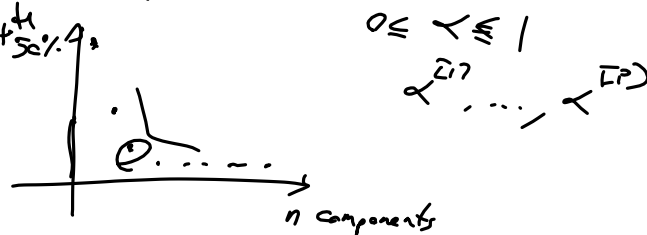
- first q components contain $\frac{\sum_{k=1}^q \lambda_k}{\sum_{i=1}^p \lambda_i}$ of the original variance

We use λ_k to decide how many PCs to have in the model

$\rightarrow$ all those that contain more than 10% of the original variance

$\rightarrow$ keep q components until $\frac{\sum_{k=1}^q \lambda_k}{\sum_{i=1}^p \lambda_i} \geq 90\%$

$\rightarrow$ use the elbow rule



$$PCR: y = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_q z_q + \varepsilon$$

$\rightarrow$ useful to simplify the model

$\rightarrow$ useful in the case of high dimensional data

$\rightarrow$ we solve problems with collinearity

$$\left. \begin{array}{l} \rightarrow \text{useful to simplify the model} \\ \rightarrow \text{useful in the case of high dimensional data} \end{array} \right\} q < p$$

cons:

$\rightarrow$ less interpretable (but in a very few cases)

$\rightarrow$ less portable (all PC are linear combinations of all original variables)

Note:

$\rightarrow$ when $q=p$, then $\hat{f}_{OLS}(x) = \hat{f}_{PCR}(x)$

$\rightarrow \Sigma$ is almost always unknown, so we need to work with $\hat{\Sigma}$

$\rightarrow$ PC are more used as a graphical tool for exploratory analysis



PARTIAL LEAST SQUARES

→ similar idea of PCR

→ PCR ignore y , PLS focuses on the information of X useful to predict y

Basically, instead of $\max_{\alpha} \text{Var}(\alpha X)$ st $\|\alpha\|=1$, $\alpha^T \mathbf{1}_m = 0$

$$\alpha = \arg\max_{\alpha} \text{Cov}(y, \alpha X) \text{Var}(\alpha X)$$

subject to $\|\alpha\|=1$ and

$$\boxed{\begin{array}{l} \alpha^T \alpha_{pcr} = 0 \\ \alpha^T S q_c = 0 \end{array}}$$

In practice, the term $\text{Var}(\alpha X)$ is often dominant, so PCR $\approx$ PLS