

# Regresión sobre componentes principales (Principal Component Regression, PCR)

Antonio Sala Piqueras

Identificación de sistemas multivariables

Dept. Ing. de Sistemas y Automática (DISA)  
Universitat Politècnica de València (UPV)

Video-presentación disponible en:

<http://personales.upv.es/asala/YT/V/pcr.html>



UNIVERSITAT  
POLITÀCNICA  
DE VALÈNCIA

# Presentación

## Motivación:

Al predecir  $y$  a partir de  $x$ , puede haber correlación significativa entre los propios componentes de  $x$ . Esto podría causar problemas.

## Objetivos:

Comprender los problemas por la colinealidad en regresores, y la solución de regularización basada en sus componentes principales.

## Contenidos:

Componentes principales en regresión. Fórmulas del PCR. Regularización y simplificación de modelos. Conclusiones.



## Componentes principales en regresión

Consideremos unos datos formados por  $N$  muestras  $(y, x)$  de  $y_{1 \times m_y}$  datos a precedir a partir de  $x_{1 \times m_x}$ , datos de entrada al modelo.

Matrices de datos  $X_{N \times m_x}$ ,  $Y_{N \times m_y}$ .

- El objetivo de regresión es minimizar  $E(\|x\theta - y\|)$ . Predictor óptimo:

$$\hat{y} = x\theta^{LS}, \quad \theta_{m_x \times m_y}^{[LS]} = \text{pinv}(X)Y = \Sigma_x^{-1}\Sigma_{xy}$$

Si el modelo verdadero es  $y = x\theta + \epsilon$ , siendo  $\epsilon \sim N(0, \sigma^2)$ , cada col. de  $\theta^{[LS]}$  tiene una varianza de parámetros estimados  $(X^T X)^{-1}\sigma^2 \approx \frac{1}{N-1} \cdot \Sigma_x^{-1}\sigma^2$ .

- Es posible que **los datos  $x$  tengan correlación entre sí**, con lo que haya componentes “pequeños” de  $\Sigma_x$ , que produzcan errores “importantes” en los parámetros estimados, salvo que  $N$  sea muy grande.

# 1. Componentes principales de la entrada $x$

Supongamos que  $X = USV^T$ , su varianza  $\Sigma_x = \frac{1}{N-1}X^T X = \frac{1}{N-1}VS^2V^T$ .

Hagamos cambio de variable  $\tilde{x} = xV$ .

Entonces,  $\tilde{X} = US$ ,  $\Sigma_{\tilde{x}} = \frac{1}{N-1}\tilde{X}^T \tilde{X} = \frac{1}{N-1}S^2 = V^T \Sigma_x V$ ,  $\tilde{x}$  son los componentes principales de la entrada, no correlados.

Covarianza entre  $y$  y  $\tilde{x}$ :  $\Sigma_{\tilde{x}y} = E(\tilde{x}^T y) = \frac{1}{N-1}V^T X^T Y = V^T \Sigma_{xy}$ .



## 2.- PCR: Principal Component Regression

Mejor predictor (mínimos cuadrados) de  $y$  dado  $\tilde{x}$ , es:  $\theta^{[PCR]} = \Sigma_{\tilde{x}}^{-1} \Sigma_{\tilde{x}y}$ .

Como  $\Sigma_{\tilde{x}}$  es diagonal  $\Sigma_{\tilde{x}} = \frac{1}{N-1} S^2 = \text{diag}(\sigma_1^2, \dots, \sigma_{m_x}^2)$ , denotando  $t_i$  la fila  $i$  de  $\Sigma_{\tilde{x}y}$ :

$$\theta^{[PCR]} = \begin{bmatrix} \sigma_1^{-2} t_1 \\ \sigma_2^{-2} t_2 \\ \dots \\ \sigma_{m_x}^{-2} t_{m_x} \end{bmatrix}$$

\*El predictor coincide, con cambio de variable, con el de mínimos cuadrados:

$$\hat{Y} = \tilde{X}\theta = (XV) \cdot \Sigma_{\tilde{x}}^{-1} \Sigma_{\tilde{x}y} = (XV) \cdot (V^T \Sigma_x^{-1} V) \cdot (V^T \Sigma_{xy}) = X \Sigma_x^{-1} \Sigma_{xy}$$

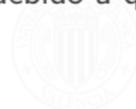
Esto es, por ahora no hemos hecho "nada" aparte de un cambio de variables  $\tilde{X} = XV$ .

## PCR: modelos reducidos

**Elección de un número reducido de componentes:** Si seleccionamos sólo los primeros “ $p$ ”,  $p \leq m_x$ , componentes principales para la regresión, entonces tenemos el mejor predictor dados esos  $p$  componentes:

$$\theta_p^{[PCR]} = \begin{bmatrix} \sigma_1^{-2} t_1 \\ \sigma_2^{-2} t_2 \\ \dots \\ \sigma_p^{-2} t_p \end{bmatrix}$$

\*En efecto, al ser los componentes de  $x$  no correlados, el coeficiente del predictor dado un componente tiene el mismo valor se disponga o no de información de otros, debido a que  $\Sigma_{\tilde{x}}$  es diagonal en la fórmula  $\theta^{[PCR]} = \Sigma_{\tilde{x}}^{-1} \Sigma_{\tilde{x}y}$ .



## Usos del PCR

**Regularización:** Escoger para regresión sólo aquellos componentes con  $\sigma_p$  por encima de un umbral mínimo. Alternativamente, cuyo intervalo de confianza 95% o 99% no incluya el cero.

**Simplificación de modelos:** Llevar eso al extremo, y plantear la regresión sólo con muy pocos componentes...  $p = 1$  o  $p = 2$  en muchos casos...

la varianza teórica del error de predicción usando “p” componentes es:

$$\Sigma_e = \Sigma_y - \Sigma_{y\tilde{x}} \Sigma_{\tilde{x}}^{-1} \Sigma_{y\tilde{x}}^T = \Sigma_y - \frac{1}{\sigma_1^2} t_1 t_1^T - \dots - \frac{1}{\sigma_p^2} t_p t_p^T$$

**Validación:** Usando, si es posible, un conjunto de datos separado (para comprobar el también el efecto de posibles pocas muestras), graficar la suma del error de predicción cuadrático en función del número de componentes elegido.

# Discusión

**Inconvenientes del PCR:** Los componentes principales de  $x$  indican las direcciones de gran varianza de  $x$ , y están directamente relacionados con la “regularización”, pero de cara a obtener un modelo que explique “mucho” con pocos componentes...

- los componentes principales dependen del **escalado**. El número de componentes PCR necesarios para explicar un determinado porcentaje de varianza puede depender de si una variable está en voltios o milivoltios.
- formalmente, lo que relaciona  $x$  e  $y$  es la **covarianza** o la **correlación**.

\*Existen otras técnicas basadas en el SVD de la covarianza que consiguen explicar más porcentaje de la varianza de “ $y$ ” con el mismo número de componentes. **[PLS, CCA]**

## Conclusiones

- Dados datos de entrada  $x$  para predecir  $y$ , el estimador de mínimos cuadrados es una matriz de coeficientes.
  - Si los datos de entrada  $x$  tienen correlación entre sí elevada, puede dar lugar a estimados de parámetros del modelo inciertos, sensibilidad a ruidos de medida elevada.
  - Seleccionando sólo los componentes principales de la entrada  $x$  con varianza alta, podemos hacer regularización... y, quizás, simplificación de la estructura del modelo.
- ▶ Tiene inconvenientes (1: sensibilidad a escalado, 2: los componentes principales de  $x$  no tienen relación con la **covarianza** o **correlación** entre  $x$  e  $y$ ).