# Grafos dirigidos para identificar la reducción suficiente en el modelo PFC

Expositor: María Eugenia Szretter Noste (Instituto de Cálculo, FCEN, UBA, meszre@dm.uba.ar)

Autor/es: María Eugenia Szretter Noste (Instituto de Cálculo, FCEN, UBA, meszre@dm.uba.ar)

Cook (2007) propone el modelo *Principal Fitted Components* (PFC). Un vector aleatorio $(\mathbf{x}, y)$ con $\mathbf{x} \in \mathbb{R}^p$, $y \in \mathbb{R}$ satisface el *modelo PFC* si existen un tvector t$\boldsymbol{\mu}_0 \in \mathbb{R}^p$, una matriz $\Gamma_0 \in \mathbb{R}^{p\times d}$ tcon rango $(\Gamma_0) = d \leq p$, una matriz $\beta t_0 \in \mathbb{R}^{d\times r}$ con $d \leq r$, una función $\mathbf{f}t : \mathbb{R} \to \mathbb{R}^r$, y una matriz $\Delta_0 \in \mathbb{R}t^{p\times p}$ definida positiva tales que
t

$$t\mathbf{x} = t\boldsymbol{\mu}t_0 + \Gamma_0\beta_0\mathbf{f}(y) + \Delta_0^{1/2}\mathbf{u}, tt \tag{1}$$

donde $\mathbf{u}$ es un vector aleatorio $p-$dimensional independiente de t$y$. Los valores de los parámetros $t\boldsymbol{\mu}t_0, \Gamma_0, \beta_0$ y $\Delta_0$ son desconocidos, pero la tfunción $\mathbf{f}$ es conocida. Al término $\Delta_0^{1/2}\mathbf{u}$ se lo denomina error tdel modelo. t ttt t tEl modelo PFC surge para resolver el problema que informalmente se puede describir del tsiguiente modo: obtener el menor número $d$ de combinaciones lineales que permiten reemplazar tla $\mathbf{x}$ sin perder información sobre $y$. Para aplicar éxitosamente técnicas de regresión no paramétricas, el tamaño de muestra debe crecer en forma exponencial con $p$. tPor esa razón se intenta reducir la cantidad de covariables, de $p$ a $d$. t Formalmente, diremos que $\mathbf{R} : \mathbb{R}^p \to \mathbb{R}^d$ con $d \leq p$ es una reducción suficiente si $y \mid \mathbf{x}$ tiene la misma distribución que $y \mid \mathbf{R}(\mathbf{x})$; o equivalentemente, si $y$ y $\mathbf{x}$ son condicionalmente independientes dado $\mathbf{R}(\mathbf{x})$. Nos interesa encontrar reducciones suficientes, ya que en ese caso, $\mathbf{x}$ puede ser reemplazado por la reducción suficiente $\mathbf{R}(\mathbf{x})$ sin pérdida de información sobre la regresión de $y$ en $\mathbf{x}$.

Cook (2007) y Cook y Forzani (2008) prueban que $\mathbf{R}(\mathbf{x}) = \Gamma_0^T\Delta_0^{-1}\mathbf{x}$ es una reducción suficiente para el modelo (1) cuando la distribución del error es normal.t t tUsando *Directed Acyclic Graphs* (DAG's), que son herramientas para modelar independencia condicional, probamos que bajo el modelo PFC la reducción t$\mathbf{R}(\mathbf{x}) = \Gamma_0^T\Delta_0^{-1}\mathbf{x}$ es suficiente bajo condiciones ttque pueden expresarse a través de la independencia de las proyecciones ortogonales del error en espacios complementarios. Esta condiciones, que incluyen a la normalidad, son más generales. Para comprobarlo, exhibimos dos ejemplos de familias de distribuciones no normales que las satisfacen.
tt