

Selección de variables y reducción de la dimensión en modelos de regresión no-paramétrica de gran dimensión

Karine Bertin^{*1}, Guillaume Lecue²

¹*Universidad de Valparaíso, CIMFAV*

²*Université de Provence, Francia*

Resumen

Consideremos un modelo de regresión no-paramétrico

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \dots, n$$

donde f es la función de regresión a estimar y los ε_i son ruidos gaussianos independientes. La estimación de la función de regresión cuando f de regularidad β no puede ser mas rápida que la velocidad $n^{-2\beta/(2\beta+d)}$ donde d es la dimensión de la variable explicativa X . En muchos ejemplos concretos, la dimensión d es muy grande y puede depender del numero de observaciones n y en este caso esta velocidad es muy lenta (lo que se llama la "maldad de la dimensión"). En algunas situaciones, la función f depende solamente de un numero d^* reducido de coordenadas de X . En esta charla, construimos dos métodos. El primero selecciona con gran probabilidad estas d^* coordenadas. En el segundo, usando las coordenadas seleccionadas, utilizamos un estimador por polinomios locales para estimar la función de regresión a la velocidad $n^{-2\beta/(2\beta+d^*)}$. Esta velocidad es mucho mas rapida que la velocidad anterior. Este resultado es obtenido usando un metodo de penalización en norma l_1 en este enfoque no-parametrico (LASSO method).

Palabras Clave: Regresión no-parametrica; LASSO; Reducción dimensión.

^{*}karine.bertin@uv.cl