Modelos Lineales para Clasificación

Modelos lineales con regularización L_1

Juan Zamora O.

INSTITUTO DE ESTADÍSTICA

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

Alta dimensionalidad en modelos lineales

Anteriormente, discutimos modelos lineales para clasificación que usaban un número moderado de variables explicativas basados en Máxima Verosimilitud (MV)

En algunos problemas, este número puede ser realmente elevado, por ejemplo en genómica pueden llegar los varios miles.

MV no funciona correctamente en estos contextos.

En problemas con alta dimensionalidad, la reducción de la cantidad de predictores es un asunto relevante.

Existen diversos métodos para la selección de subconjuntos de predictores. Por ejemplo, *Stepwise variable selection*.

Este tipo de métodos puede variar notoriamente su resultado en presencia incluso de pequeños cambios en los datos

Este tipo de métodos exhibe una alta varianza en el error de predicción.

Los métodos de regularización permiten encoger los estimadores hacia el cero e incluso realizar selección de variables vía optimización de log-verosimilitud penalizada

Aspectos relevantes al usar regularización

Estimadores MV fallan en la existencia de solución única para los coeficientes. Regularización garantiza solución única

El modelo regularizado es capaz de producir predicciones comparables o mejores a aquellos basados en MV

Sparsity favorece la generación de modelos parsimoniosos que contengan los efectos más fuertes. Esto a su vez favorece la interpretación.

Regularización de modelos lineales

Métodos de regularización derivados de estimadores de Máxima Verosimilitud (MV) se basan en la *Log-verosimilitud penalizada* l_{P}

$$l_{\mathsf{P}}(\boldsymbol{\beta}) = \sum_{i=1}^{N} l_i(\boldsymbol{\beta}) - \frac{\lambda}{2} J(\boldsymbol{\beta})$$

Donde $l_i(\boldsymbol{\beta})$ representa la contribución del i-ésimo ejemplo a la log-verosimilitud, λ es un parámetro que indica qué tanto regularizar y J es un funcional que penaliza el tamaño de los parámetros

Asintóticamente, métodos de regularización que *encogen* los estimadores reducen la varianza[1].

[1] "On ordinary ridge regression in generalized linear models", Segerstedt, 1992.

Regularización con norma L_1

Al maximizar l_{P} se buscan estimadores para los coeficientes cercanos a la solución por MV, pero con parámetros regularizados.

Por ejemplo, la penalización Lasso utiliza $J(\boldsymbol{\beta}) = \sum_{j=1}^{m} |\beta_j|$. Este esquema favorece la generación de modelos *sparse* con menor cantidad de coeficientes no nulos.

Este tipo de métodos es especialmente útil cuando existen predictores correlacionados.

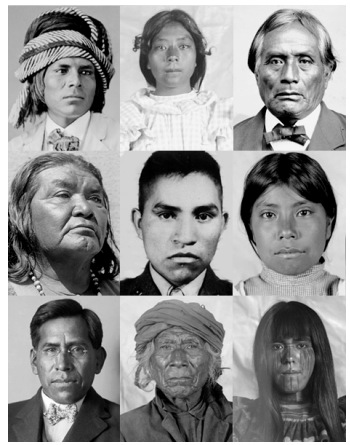
La selección del parámetro λ se hace generalmente estimado el error de predicción.

Regularización con norma L_1

La maximización de la log-verosimilitud regularizada con norma L_1 para los coeficientes, $||\boldsymbol{\beta}||$, producirá un *encogimiento* de algunos, pero también la anulación de otros.

La anulación de algunos coeficientes β_j^* implica que algunas covariables son quitadas del modelo.

El conjunto de datos de diabetes en los indios Pima tiene como tarea la predicción de la aparición de diabetes en un periodo de 5 años usando detalles médicos como covariables.



Extraída desde Wikipedia.