

Actividades Formativas IMEIO

Título: Procedimientos inferenciales para datos de alta dimensión
Responsable de la actividad: Leandro Pardo Llorente y Nirian Martin
Profesores involucrados: Leandro Pardo y Nirian Martin
Número de horas totales: 15
Lugar: Facultad de Matemáticas
Fechas: Noviembre 26, 29 y 30 y Diciembre 1, 2 (3 horas al día)

Resumen:

En las últimas tres décadas, debido a la evolución tan espectacular de las computadoras y la tecnología de recogida de datos, se ha producido un cambio sustancial en la metodología de las aplicaciones prácticas de análisis de datos, de modo que ha sido más habitual encontrar situaciones en las que se dispone de un número menor de unidades experimentales en comparación con la dimensión de las variables en estudio. Estas situaciones se presentan en áreas tan dispares como: Genética y Bioinformática; Quimiometría; Neuroimagen, Econometría Genómica Microarrays, Clasificación de documentos, Astronomía, Ciencias Atmosféricas, etc La estadística para datos de alta dimensión (high-dimensional data) hace referencia a los procedimientos estadísticos diseñados cuando el número de parámetros p es mucho mayor que el tamaño de la muestra n , es decir $p \gg n$. Una monografía de referencia imprescindible es Bühlmann y van de Geer (2011).

El objetivo del curso es presentar, para el modelo de regresión línea múltiple, los procedimientos de regularización haciendo un especial hincapié en el método de regularización todo LASSO sin dejar de lado sus generalizaciones. Para cada uno de estos de regularización se abordará la problemática de la "predicción de la superficie de regresión", la "estimación del vector de parámetros" así como "selección de variables" o "estimación del soporte del verdadero valor del parámetro".

Se complementarán los estudios teóricos con estudios prácticos con el manejo de las bibliotecas en R para abordar diversos problemas: MASS, lqa, lars, elasticnet, parcor, relaxo, ncvreg, etc. En concreto la biblioteca, glmnet, desarrollado bajo el título "Lasso and Elastic Net Regularized lineal models" realiza procedimientos extremadamente eficientes para la obtención de los estimadores RIDGE y LASSO y que fue desarrollado por Friedman y otros (2010).

Bühlmann, P. y van de Geer, S. (2011). Statistics for High-dimensional Data: Methods, Theory and Applications. Springer

Friedman, J., R. Tibshirani, and T. Hastie (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33, 1.

¿Aceptarías que el curso se pudiera emitir por videoconferencia restringido a algunos alumnos del doctorado que no pudieran asistir presencialmente? No