



UNIVERSIDAD COMPLUTENSE
MADRID

AVISO DE CONFERENCIA

Data Mining con Árboles de Decisión para Bioinformática

Dr. Jorge Martín Arevalillo, UNED

Facultad de Informática
Sala de Grados • 19 de junio de 2013 • 11:00
entrada libre hasta completar el aforo

resumen:

La clasificación de observaciones multivariantes es un problema común a muchas disciplinas. Los procedimientos clásicos para abordarlo, que forman parte del contenido de cualquier curso introductorio de Estadística Multivariante, se remontan al Análisis Discriminante de Fisher y sus conclusiones son fiables siempre que sean ciertas las hipótesis estadísticas en que se apoyan. En la práctica, sobre todo al abordar un problema de minería de datos, el conjunto de variables predictoras es grande y variado, la validación de estas hipótesis es una tarea tediosa y su cumplimiento se antoja imposible. Como alternativa, gracias a la creciente capacidad de almacenamiento y cálculo de los ordenadores, se han desarrollado procedimientos de clasificación cuyos resultados y conclusiones no descansan en hipótesis estadísticas y modelos paramétricos preestablecidos; entre ellos se encuentran los árboles de decisión. El objeto de la charla es presentar la manera en que los algoritmos CHAID y CART resuelven el problema de clasificación mediante árboles de decisión, mostrando sus similitudes y diferencias. También se hará énfasis en las ventajas de las técnicas de clasificación con árboles frente a los métodos clásicos a la hora de abordar la modelización de los datos en un problema de Data Mining.

Sobre Jorge Martín:

Profesor Contratado Doctor, miembro del departamento de Estadística, Investigación Operativa y Cálculo Numérico de la UNED. Doctor por la UNED desde el año 2003. Ha colaborado dando asesoramiento estadístico y de data mining en proyectos para compañías como BBVA Continental, McDonald's España, Yell Publicidad, Mutua Madrileña Automovilista, entre otras. También ha trabajado como investigador para Salford Systems, empresa con sede en San Diego (USA) que desarrolla el algoritmo CART. Instructor y formador en instituciones públicas y privadas como Boston Consulting Group, Fundación CIFF y Fundación UNED. Actualmente imparte docencia en la asignatura Estadística del grado de informática de la UNED y desarrolla labores de investigación tanto en el campo de la estadística teórica y como en el del análisis de datos y data mining, con especial énfasis en el reconocimiento de patrones en datos de alta dimensión. Sus trabajos más recientes han sido publicados en revistas de investigación como BMC Bioinformatics, Journal of Multivariate Analysis o Metrika.