

Conferencia invitada

Hilbert Space Representations of Probability Distributions

por

Dr. Arthur Gretton

Research Scientist, Max Planck Institute for Biological Cybernetics

Abstract:

Many problems in unsupervised learning require the analysis of features of probability distributions. At the most fundamental level, we might wish to determine whether two distributions are the same, based on samples from each - this is known as the two-sample or homogeneity problem. We use kernel methods to address this problem, by mapping probability distributions to elements in a reproducing kernel Hilbert space (RKHS). Given a sufficiently rich RKHS, these representations are unique: thus comparing feature space representations allows us to compare distributions without ambiguity. Applications include testing whether cancer subtypes are distinguishable on the basis of DNA microarray data, and whether low frequency oscillations measured at an electrode in the cortex have a different distribution during a neural spike.

A more difficult problem is to discover whether two random variables drawn from a joint distribution are independent. It turns out that any dependence between pairs of random variables can be encoded in a cross-covariance operator between appropriate RKHS representations of the variables, and we may test independence by looking at a norm of the operator. Finally, we establish that this operator norm is itself a difference in feature means.

Biography:

Arthur Gretton was born in Canberra (Australia) in 1976. He received the Bachelor of Science in Physics in 1996, and the Bachelor of Engineering with Honours in 1998. In 1996 he obtained the Shell Award for top aggregate score for third year units in Physics and Chemistry, in 1998 he was awarded the University Medal in Engineering and the ANU Honours Scholarship, and in he obtained the Australian Postgraduate Award for PhD studies. From 1999 to 2003 he undertook his PhD studies in Engineering at Cambridge University's Engineering Department. Since 2003 he is Research Scientist at the Max Planck Institute for Biological Cybernetics. His main research interest is in using kernel methods to reveal properties of probability distributions.

Día: 26 de junio de 2007

Hora: 16:00

Lugar: Sala 4.3.A.05

Edificio Torres Quevedo

Universidad Carlos III de Madrid

Campus de Leganés

Avda. De la Universidad, 30

28911 Leganés (Madrid)

Organizada y patrocinada por:

- * Programa PRO-MULTIDIS, Comunidad de Madrid

- * Master Oficial Interuniversitario en Multimedia y Comunicaciones

(SRC: dluengo@tsc.uc3m.es)