

CLUSTERING ESPECTRAL: PROBABILIDAD Y EDPs PARA PROBAR CONSISTENCIA EN UN
PROBLEMA DE MACHINE LEARNING.

Expositor: Pablo Groisman (DM Exactas UBA e IMAS-CONICET, pgroisma@dm.uba.ar)
Autor/es: Pablo Groisman (DM Exactas UBA e IMAS-CONICET, pgroisma@dm.uba.ar)

Spectral clustering es un método para encontrar grupos “no lineales” en conjuntos grandes de datos. Se basa en tener una noción de similaridad para cada par de datos y usar esa similaridad para construir un laplaciano discreto. Luego se calculan los autovalores y autovectores del laplaciano discreto y se utilizan para determinar los grupos. Hay muchos resultados de consistencia (convergencia) para estos métodos pero todos asumen que la cantidad de “vecinos” tiende a infinito junto con el tamaño de la muestra. Mostraremos que esta hipótesis puede ser removida. La técnica consiste en probar un principio de invarianza para el paseo al azar definido por la noción de similaridad y un resultado de convergencia de autovalores (y autovectores) que no precisa la convergencia del laplaciano. Es un trabajo en progreso