



## ΚΥΚΛΟΣ ΣΕΜΙΝΑΡΙΩΝ ΣΤΑΤΙΣΤΙΚΗΣ ΔΕΚΕΜΒΡΙΟΣ 2019

**Konstantinos Perrakis**

*Department of Mathematical Sciences, Durham University*

### **Latent group structure and regularized regression**

ΠΑΡΑΣΚΕΥΗ 20/12/2019, 13:15

**ΑΙΘΟΥΣΑ Τ103**  
**Τροίας 2, Νέο Κτίριο ΟΠΑ**

#### **ΠΕΡΙΛΗΨΗ**

Regression modelling typically assumes homogeneity of the conditional distribution of responses  $Y$  given features  $X$ . For inhomogeneous data, with latent groups having potentially different underlying distributions, the hidden group structure can be crucial for estimation and prediction, and standard regression models may be severely confounded. Worse, in the multivariate setting, the presence of such inhomogeneity can easily pass undetected. To allow for robust and interpretable regression modelling in the heterogeneous data setting we put forward a class of mixture models that couples together both the multivariate marginal on  $X$  and the conditional  $Y | X$  to capture the latent group structure. This joint modelling approach allows for group-specific regression parameters, automatically controlling for the latent confounding that may otherwise pose difficulties, and offers a novel way to deal with suspected distributional shifts in the data. We show how the latent variable model can be regularized to provide scalable solutions with explicit sparsity. Estimation is handled via an expectation-maximization algorithm. We illustrate the key ideas via empirical examples.



**AUEB STATISTICS SEMINAR SERIES  
DECEMBER 2019**

**Konstantinos Perrakis**

*Department of Mathematical Sciences, Durham University*

**Latent group structure and regularized regression**

FRIDAY 20/12/2019, 13:15

**Room T103,  
Troias 2, New AUEB Building**

**ABSTRACT**

Regression modelling typically assumes homogeneity of the conditional distribution of responses  $Y$  given features  $X$ . For inhomogeneous data, with latent groups having potentially different underlying distributions, the hidden group structure can be crucial for estimation and prediction, and standard regression models may be severely confounded. Worse, in the multivariate setting, the presence of such inhomogeneity can easily pass undetected. To allow for robust and interpretable regression modelling in the heterogeneous data setting we put forward a class of mixture models that couples together both the multivariate marginal on  $X$  and the conditional  $Y | X$  to capture the latent group structure. This joint modelling approach allows for group-specific regression parameters, automatically controlling for the latent confounding that may otherwise pose difficulties, and offers a novel way to deal with suspected distributional shifts in the data. We show how the latent variable model can be regularized to provide scalable solutions with explicit sparsity. Estimation is handled via an expectation-maximization algorithm. We illustrate the key ideas via empirical examples.