PhD PROGRAMME IN:

SYSTEMS BIOLOGY IN IMMUNE AND INFECTIOUS DISEASES

Course title:	**DATA SCIENCE FOR BIOMEDICAL MEASUREMENTS**
Aim of the course:	This course aims at further extending participants' knowledge in Statistical Learning
Lecturer:	Dr. Emilia Nunzi
Venue:	to be defined

Course Program:	**Introduction**
	• Introduction and examples
	• Supervised and unsupervisedlearning
	• Differencesbetweenprobability and statistics
	• Statisticsapplied to biomedical data
	• Probabilityconceptsreview
	    ○ Lab: introduction to Excel
	
	Statistical Learning
	• Statistical learning and regression
	• Model accuracy and bias trade-off
	• Classificationproblems
	    ○ Lab: introduction to R
	
	Linear Regression
	• Simple linear regression and confidence interval
	• Hypothesistesting
	• Multiple Linear Regression and Interpreting Regression
	• Coefficients
	• Model Selection and Qualitative Predictors
	• Interactions and Nonlinearity
	    ○ Lab: linear regression and correlation coefficient (R and Excel)
	
	Classification
	• Introduction to Classification
	• LogisticRegression and Maximum Likelihood
	• Multivariate Logistic
	• Regression and Confounding
	• Case-Control Sampling and Multiclass Logistic
	• Regression
	• Linear Discriminant Analysis and Bayes Theorem
	• Univariate Linear

	- Discriminant Analysis
	- Multivariate Linear
	- Discriminant Analysis and ROC Curves
	- QuadraticDiscriminant
	- Analysis and NaiveBayes
	    o Lab: Logistic Regression (R and Excel)
	    o Lab: Linear Discriminant Analysis ®
	    o Lab: K-NearestNeighbors ®
	
	ResamplingMethods
	- Estimating Prediction Error and Validation Set Approach
	- K-fold Cross-Validation
	- Cross-Validation: The Right and Wrong Ways
	- The Bootstrap
	- More on the Bootstrap
	    o Lab: Cross-Validation
	    o Lab: The Bootstrap
	
	Tree-BasedMethods
	- DecisionTrees
	- Pruning a DecisionTree
	- Classification Trees and Comparison with Linear Models
	- Bootstrap Aggregation (Bagging) and Random Forests
	- Boosting and VariableImportance
	    o Lab: DecisionTrees
	    o Lab: Random Forests and Boosting
	
	Unsupervised Learning
	- Unsupervised Learning and Principal Components Analysis
	- Exploring Principal Components Analysis and Proportion of Variance Explained
	- K-means Clustering
	- Hierarchical Clustering
	- Breast Cancer Example of Hierarchical Clustering
	    o Lab: Principal Components Analysis (PCA)
	    o Lab: Hierarchical Clustering

Target audience:	The course is mainly meant for PhD students in Systems Biology in Immune and Infectious Pathologies PhD programme. However, the course is open to all scientists with an interest in the issue.