|  |  |  |  |
| --- | --- | --- | --- |
| **Unit** | **Amphi** | **PC** | **TD** |
| 1 | **Introduction to programming in Python**   * Type in Python: int, float, string, list, tuple, dictionary * Functions, lambda * if * for, while * read and write to file * csv, tsv, special characters | No PC | TD1 – History MCQ |
| 2 | **Class and Instance in Python**   * Class and instance * Attribute and methods * Public vs private * Compare attribute, mutability * Class method, instant method, static method | No PC | TD2 – Polynomial |
| 3 | **Useful libraries in Python**   * numpy * scipy * pandas: read csv * matplotlib * url * BeautifulSoup, xmlparser * sys | No PC | TD3 – Vnexpress [1] – Retrieve article, paragraphs, sentences, words |
| 4 | **Algorithms and complexity**   * Basic operations’ complexity for number and for matrix * Recurrence * Divide and conquest. The master theorem * Monte Carlo method * Short introduction to concurrence programming | PC4 – Algorithms and complexity | TD4 – Noughts and Crosses |
| 5 | **Clustering [1] – K-mean**   * Machine learning, what is it? * Supervised vs unsupervised * Applications * Parameters vs non-parameters * Unsupervised problems – Clustering * K-mean * Introduction to text mining | PC5 (to be announced) | TD5 – Vnexpress [2] – Vocabulary fields |
| 6 | **Regression [1] – Linear Regression**   * Linear Regression problem * Differentiation on matrices * Closed form * Gradient descent * Learning and prediction * Feature scaling, feature selection | PC6 | TD6 – VNIndex |
| 7 | **Regression [2] – From Linear Regression to Polynomial Regression**   * Probabilistic model for Linear Regression. Maximum likelihood * Bayesian view * Other function forms * Polynomial regressions, new features creations * Overfitting * Solve overfitting: Ridge and Lasso | PC7 | TD7 – |
| 8 | **Classification [1] – Logistic Regression**   * Maximum likelihood * Loss function in classification * Logistic regression * Newton’s method * Second order method * Feature selection * Categorical feature | PC8 | TD8 – Vnexpress (2): Social news or Sport news? |
| 9 | **Classification [2] – More models**   * K-nearest neighbors * Naïve Bayesian * LDA * QDA * Terminologies * ROC * Overfitting * Cross validation | PC9 | TD9 – Candidate evaluation in DongHanh Association (1) – Digit recognition |