



AR2Tech

Geostatistical technologies and services

Machine learning and data analysis for geomodelling

This 4 day short course aims at providing the essential skills to understand and apply modern tools in data analysis to problems commonly found in geomodelling applications. The course is based on the python scikit-learn package. It covers dimensionality reduction, clustering, regression, classification and data visualization. The attendees should be familiar with the basic of programming (variables, conditional statement, loop), a basic primer for the python language will be provided. The last day is reserved for a workshop where the attendees will apply their new skills on their own data set under the supervision of the instructor. The emphasis of the course is empowering the attendees to use these tools on their own spatial applications with the help of the ar2gems software.

Day 1:

- Introduction to Python. Using the numpy and matplotlib libraries. Input and output of data.
- Dimensionality reduction and factorisation: Principal component analysis (PCA), Kernel PCA, Independent components analysis, multidimensional scaling

Day 2:

- Introduction to clustering and unsupervised classification.
- K-means, Hierarchical clustering.
- Integrating spatial context into clustering

Day 3:

- Introduction to supervised classification and regression:
- Linear and Quadratic discriminant analysis, Support vector machine, Nearest neighbor, Decision trees

Day 4:

- Bring your own data.
- Apply machine learning techniques to each attendee data set.