Subject: **Machine Learning**

Summary: Machine Learning is the analysis of data to automatically learn how to identify and label categories or patterns of interest, for problem-solving and decision-making. This topic focuses on supervised machine learning pattern classification and regression. Unsupervised methods are also covered, but not reinforcement learning.

Keywords

- Parametric and non-parametric supervised machine learning techniques including (deep) neural networks (e.g. fully connected, CNN) recurrent neural networks (e.g. LSTM, GRU, attention), support vector machines, (gradient boosted), decision trees/forests, hidden Markov models, K-nearest-neighbour classifiers, Bayesian Models.
- Data pre-processing, feature selection, missing value imputation, outlier detection
- Component analysis and dimensionality reduction: PCA, ICA, Fisher discriminant analysis, t-SNE, Uniform Manifold Approximation and Projection (UMAP).
- Classification and regression performance metrics and estimation of prediction accuracy, including confusion tables, sensitivity, specificity, precision, F1 score, ROC curves, precision-recall curves, resampling & bootstrapping techniques, hypothesis testing, etc.
- Unsupervised learning and clustering, including K-means, fuzzy K-means, hierarchical clustering, and determining the number of clusters.
- Experiment Design issues, including classifier complexity, generalization, curse of dimensionality, overfitting, class imbalance, sources of experimental bias, and bias-variance tradeoff.