ELEN0062: Introduction to machine learning

Instructions for the 2025-2026 oral exam

Procedure:

- Two questions/topics from the list below are (randomly) assigned to you.
- You prepare these questions on paper for no more than one hour (**closed book**).
- You are interviewed for about 20 minutes based on your written response.

List of questions

1. **Machine learning protocols**: supervised and unsupervised learning, reinforcement learning, batch vs online learning, semi-supervised and transductive learning. Give an example of one practical problem and of one method for each protocol.

2. **Model evaluation criteria**: main evaluation criteria in classification (error rate, sensitivity/specificity, ROC curves…) and in regression (squared error, correlation...), evaluation during prediction versus evaluation during training.

3. **Model evaluation methods:** motivation, main approaches (resubstitution, test set, cross-validation, leave-one-out, bootstrap), advantages and drawbacks of each method, model selection by cross validation (small and large datasets), selection bias.

4. **Bias-variance tradeoff**: bias-variance decomposition (derivation and interpretation of the different terms, link with over- and under-fitting), parameters that influence bias and variance, bias and variance estimation.

5. **Variance reduction techniques:** techniques to reduce variance in the context of the different learning algorithms covered in the course (regularization, complexity control, ensemble methods, feature selection), discussion and comparison of the different learning algorithms from the point of view of their bias and variance.

6. **Classification and regression trees**: hypothesis space, growing and pruning algorithms, impurity measures, extensions to regression and continuous attributes, interpretability.

7. **Linear regression**: problem formulation, analytical solution, introduction of a regularization term (motivation and modified solution), linear regression with kernels.

8. **K-Nearest Neighbors methods:** 1-NN and k-NN algorithms, refinements, relations with tree-based, linear, and kernel methods.

9. **From single neuron models to multilayer perceptrons**: single neuron models (batch vs online learning, hard and soft threshold unit, theoretical properties), multilayer perceptrons (definition, representation capacity, general principles of the training algorithm).

10. **Neural networks:** multilayer perceptrons (definition, representation capacity, back-propagation, techniques to avoid overfitting), other models (RBF, ~~convolutional and recurrent networks: general principles only~~).

11. **Support vector machines**: main notions (classification margin, optimization problem, support vectors, soft margin)**,** general idea of the kernel trick.

12. **Kernel methods:** kernel definition, kernel trick, examples of kernels and of kernel methods. Discuss the interest of kernel methods in general. Kernel interpretation of trees and linear classification.

13. **Ensemble methods:** motivation and algorithms (bagging, random forests, boosting, stacking…), ambiguity decomposition.

14. **Feature selection:** motivation, filter, embedded, and wrapper techniques (give examples of approaches in each family and discuss their main advantages and drawbacks), selection bias.

15. **Clustering methods:** problem definition and motivation, hierarchical clustering, k-means, determination of the number of clusters.

16. **Dimensionality reduction**: motivation, principal component analysis (optimisation problem, analytical solution, interpretation), extensions and other methods (general ideas only).

Notes:
- While the examination will be mainly driven by the two selected topics, questions can also be asked on all the course material, including the projects.
- For all methods, you should be able provide their main drawbacks and strengths (with respect to other algorithms), assess the effect of their main hyper-parameters on bias and variance, and discuss their computational complexity.