Main questions:

1. **Model evaluation methods and criteria in supervised learning**

   Model evaluation methods: motivation, resubstitution, cross validation, bootstrap
   Model selection by cross validation: small and large datasets, selection bias
   Main evaluation criteria in classification (error rate, sensibility/specificity, ROC curves...) and in regression (squared error, correlation…), evaluation during prediction versus evaluation during training.

2. **Supervised learning with decision and regression trees**

   Tree growing and pruning algorithms.
   Ensemble methods (bagging, boosting, random forests).
   Kernel corresponding to a single tree or to an ensemble of trees.
   Motivate the different algorithms and discuss their computational complexity.

3. **Least squares regression**

   Linear regression (with a regularization term), multilayer perceptrons for regression, linear regression with kernels.
   Clearly formulate the optimization problem corresponding to each algorithm (objective function to minimize, parameters to be learned, optimality conditions, discussion of pathological cases if any).
   Motivate the different algorithms and discuss their computational complexity.

4. **k-Nearest Neighbors methods and kernel methods**

   1-NN and K-NN algorithms.
   Kernel interpretation of trees and linear classification.
   What are the main advantages and drawbacks of the K-NN method with respect to other supervised learning methods?

5. **Support vector machines and kernel methods**

   Main notions: classification margin, optimization problem, support vectors, soft margin.
   Kernel definition, kernel trick, examples of kernels, kernel methods.
   Discuss the main advantages and drawbacks of support vector machines and the interest of kernel methods in general.

6. **Bias-variance tradeoff and ensemble methods**

   Bias-variance decomposition, over- and under-fitting, parameters that influence bias and variance, bias and variance estimation, variance reduction techniques.
   Give examples of variance reduction techniques in the context of the different learning algorithms covered in the course, discuss and compare the different learning algorithms from the point of view of their bias and variance.
   Ensemble methods: motivation and algorithms (bagging, random forests, boosting, stacking), ambiguity decomposition.

7. **Unsupervised learning**

   Motivation
   Clustering methods: problem definition, k-means and hierarchical clustering, determination of the number of clusters.
   Dimensionality reduction: principal component analysis, extensions.

Secondary 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 used during the test stage in classification (error rate, sensibility, specificity, ROC curves...) and in regression (squared error, correlation...).

3. Error sources (residual error, bias, variance): intuitive discussion about the influence of the choice of input attributes (number, type, combination), of the model complexity, of the learning sample size...

4. Model evaluation methods (resubstitution, test set, cross-validation, leave-one-out, bootstrap): explain the principles and the advantages and drawbacks of these methods.

5. Approach to limit or decrease overfitting (relation between under/over-fitting and bias/variance; complexity control, feature selection, ensemble methods).

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

7. Clustering methods: problem formulation, hierarchical clustering, k-means.

8. Feature selection: motivation, filter, embedded, and wrapper techniques, selection bias.