MACHINE LEARNING 2023/2024
DETAILED PROGRAM
Prof. F. Vandin
Last Update: December 30th, 2023

This document describes, for each topic, the level of detail required for the material presented during the lectures (see the slides). The level of detail relates to what has been presented during the lectures, so *all details* means that all details presented during the lectures are required, while *main idea* means that only the understanding of the concepts presented is required, while the details (e.g., details of propositions, proofs, and specific formulas) are not required.
Note that the background material (probability, linear algebra) is assumed to be known at the level of detail used during the presentation of the topics below. For some propositions the corresponding proposition in the book is referenced for clarity. The book chapters for [UML] are as in the PDF freely available from the authors at:
https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/copy.html

1. **Learning Model (Chapters 2 and 3 [UML])**

- *All details* presented in class required, including: the formal model with definitions, all propositions' statements, and the proof of Corollary 2.3 [UML]; these include: the definition of PAC learnability with 0-1 loss, sample complexity for PAC learnability of finite hypotheses classes with 0-1 loss, definition of agnostic PAC learnability for general loss functions and all related definitions (empirical and true error for general loss functions, etc.)

2. **Uniform Convergence (Chapter 4 [UML])**

- Definition of ε-representative sample: *all details*
- Lemma 4.2 [UML]: statement and proof *with all details*;
- Definition of uniform convergence property: *main idea* (details of definition not required)
- Corollary 4.4 [UML]: *main idea* (details of statement not required)
- Corollary 4.6 [UML]: statement and proof *with all details*

3. **Bias-Complexity Tradeoff (Chapter 5 [UML])**

- No Free Lunch (NFL) theorem, NFL and prior knowledge: *only main idea*
- Corollary 5.2 [UML]: statement *with all details* (no proof)
- Approximation error, estimation error, complexity and error decomposition: *all details*

4. **VC-dimension (Chapter 6 [UML])**

- Restrictions, shattering, VC-dimension: definitions *in detail*
- Fundamental Theorems of Statistical Learning and bound on generalization: *in detail*, as on the slides (no proof)

5. **Linear Models (Chapter 9 [UML])**

- linear predictors/models: definitions with *all details*
- linear regression: definitions, matrix form, derivation best predictor, use of generalized inverse *in detail* (derivation generalized inverse: not required)
- R^2: definition and interpretation *in detail*
- linear classification: perceptron: definitions and algorithm *in detail*
- proposition on perceptron convergence: *only main idea* (as in slide "Perceptron: Notes")
- VC-dimension of linear models: *in detail*
- Logistic regression: *all details*

6. **Model Selection and Validation (Chapter 11 [UML])**

- validation: *all details*, including statement of bound on generalization error using validation set (Theorem 11.1 [UML], no proof)
- validation for model selection: *all details*, including statement of bound on generalization error using validation set (Theorem 11.2 [UML], no proof)
- model-selection curve, train-validation-test split, k-fold cross validation: *all details*
- what if learning fails: *main idea*

7. **Regularization and Feature Selection (Chapters 13 and 25 [UML])**

- Regularized Loss Minimization: *all details*
- l_1 regularization, LASSO: *all details*
- Tikhonov Regularization , Ridge Regression, derivation of optimal solution for Ridge Regression: *all details*
- subset selection, forward selection, backward selection, without and with validation data: *all details*

8. **GD, SGD, and SVM (Chapters 14 and 15 [UML])**

- gradient descent (GD): *all details*
- stochastic gradient descent (SGD): *all details*
- hard-SVM optimization problem: *all details*
- hard-SVM: equivalent formulation, and quadratic formulation: *main idea*
- definition of support vectors for hard-SVM: *all details*
- soft-SVM optimization problem, hinge loss: *all details*
- SGD for solving soft-SVM (algorithm): *all details*
- Hard-SVM dual formulation: *main idea*
- SVM for regression: *all details* only for optimization problem and support vectors definition

9. **Kernels and SVM (Chapter 16 [UML])**

- Definition of kernel: *all details*
- Kernel trick for SVM: *all details*
- commonly used kernels for SVM: *all details*

10. Neural Networks and Deep Learning (Chapter 20 [UML])

- Neuron, activation function, network architecture, point of view of one node, hypothesis set, matrix notation: *all details*
- General construction of NN for a given Boolean formula: *all details*
- Expressiveness of NNs: *main idea*
- Sample complexity, runtime of learning NNs: *main idea*
- Forward propagation algorithm: *all details*
- SGD and Backpropagation algorithm: *main idea* (pseudocode: only main structure)
- Regularized NNs: *main idea*
- CNNs: differences with standard NNs, convolution properties, convolutional layer, pooling layer, dropout, early stopping, data augmentation, cross entropy loss function: *main idea*

11. Clustering (Chapter 22 [UML])

- Unsupervised learning introduction: *all details*
- Clustering definition and difficulties: *main idea*
- Model for clustering: *all details*
- k-means clustering and Lloyd's algorithm: *all details*
- complexity of Lloyd's algorithm: *all details*
- k-means++: *all details*
- linkage-based clustering: *all details*
- silhouette score: *all details*