CSCI-UA 9473. Machine Learning
Material for the Final

Augustin Cosse

Fall 2022

1 Material covered

1. You must be able to explain how to learn a linear regression model through the residual sum-of-squares (RSS) loss. You must be able to derive the gradient iterations on that loss as well as the closed form solution for the regression coefficients (obtained through the normal equations).

2. You must be able to list and explain the three main regularization approaches (Ridge, Lasso, Best Subset Selection)

3. You must be able to interpret the RSS, Ridge and LASSO minimizers as particular instances of Maximum Likelihood (MLE) and Maximum A Posteriori (MAP) estimators, and describe the corresponding assumptions in terms of probability distributions.

4. You must be able to compare the regularization approaches in terms of their respective efficiency and complexity

5. You must be able to explain how cross validation can be used for model selection (e.g. for best subset selection).

6. You must be able to explain how the MSE can be decomposed as the sum of a bias and a variance contribution and how those contributions evolve with model complexity

7. You must be able to explain how the Residual-Sum-of-Squares (RSS) criterion can be used to learn a binary classifier

8. You must be able to explain how the binary classifier can be extended into a multiclass classifier (i.e. one-vs-rest, one-vs-one, and Multi-class discriminant through one-hot encoding)

9. You must be able to describe and derive the expression for the logistic regression classifier. You must be able to explain how to learn the model.

10. You must be able to describe and derive the expression for the Gaussian Discriminant Analysis classifier/Linear Discriminant Analysis classifier. You must be able to explain how to fit the parameters of the model to a given set of pairs $\left\{\boldsymbol{x}^{(i)}, t^{(i)}\right\}_{i=1}^{N}$.

11. You must be able to describe the Naive Bayes classifier

12. You must be able to explain the difference between generative and discriminative classifiers and give an example from each family.

13. You must be able to describe the perceptron model as well as the perceptron learning rule and the associated convergence theorem

14. You must be able to provide the general (mathematical) expression of a neural network and draw the corresponding diagram

15. You must be able to explicitly construct a simple neural network as a combination of separating planes.

16. You must be able to explain the backpropagation algorithm (list the main steps) and provide the associated equations.

17. You must be able to apply the Kernel trick and motivate its use from the batch gradient iterations on the residual-sum-of-squares loss.

18. You must be able to describe how to learn a kernel classifier through gradient iterations on the Residual-sum-of-squares loss.

19. You must be able to define the notion of Mercer Kernel (i.e. list the two properties that a matrix has to satisfy to be derived from such a kernel). You must be able to determine whether a simple kernel is a valid (i.e. Mercer) kernel.

20. You must be able to describe the notion of Maximal Margin Classifier. You must be able to derive the optimization problem that one has to solve to learn a Maximal Margin Classifier (both the min/max formulation and the constrained formulation) from the distance of a point to a plane.

21. You must be able to give the final expression of the Maximal Margin Classifier/SVM. In particular, you must be able to explain why this expression only depends on the support vectors. You must be able to explain the notion of support vectors.

22. You must be able to list and explain the most important clustering algorithms (K-means, K-medoid, hierarchical clustering, Gaussian Mixture models)

23. You must be able to provide the pseudo code for K-means and K-medoid and list the most popular initialization approaches for those two algorithms

24. You must be able to explain the EM algorithm.

25. You must be able to list and explain the three main agglomerative clustering approaches (single linkage, complete linkage and group average)

26. You must be able to explain how divisive clustering work and give the formulas used to split a cluster into the resulting two subclusters

27. You must be able to describe the Market Basket Analysis problem. You must be able to explain how to find frequent item sets efficiently by means of the A priori Algorithm. You must be able to explain the notion of Association Rules.

28. You must be able to list and discuss the assumptions behind the most important latent variable models (Factor Analysis, Principal Component Analysis and independent component analysis). You must be able to give their mathematical formulation and explain how they relate to one another.

29. You must be able to explain how the Principal component analysis problem can be solved through the eigenvalue decomposition of the empirical covariance matrix. You must be able to explain how to compute the projection onto the PCA space from this decomposition.