CSCI-UA 9472. Machine Learning
Material for the Midterm

Augustin Cosse

October 2020

1 Material covered

1. You must be able to explain the residual sum of squares criterion, how to use it to learn a linear model on the data. You must be able to derive the gradient iterations on that criterion as well as the closed form solution for the vector of weights (obtained by setting the derivatives to zero).
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 explain the statistical assumptions leading to the residual sum of squares model, the Ridge and the Lasso. You must be able to describe the distributions involved in each regularization approach (Gaussian and Laplace).
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 to use the Residual Sum of Squares criterion to learn a binary classifier
6. 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 one hot encoding)
7. You must be able to explain the distinction between generative and discriminative classifiers and give an example from each family.
8. You must be able to explain and derive the expression for the logistic regression classifier
9. You must be able to explain and derive the expression for the Linear/Gaussian Discriminant Analysis Classifier
10. You must be able to discuss the use of Kernels and motivate this use through gradient descent on large feature vectors.
11. 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)
12. You must be able to discuss the kernel trick and apply it to the Residual sum of squares criterion to derive a formulation that only depends on the similarities and not on the feature vectors.
13. You must be able to explain the notion of Maximal Margin Classifier. You must be able to derive the optimization problem that one has to solve to learn this classifier from the distance of a point to a plane.

14. You must be able to give the final expression of the Maximal Margin Classifier/SVM. In particular, you must be able to use this expression to illustrate the notion of support vectors.

15. You must be able to explain the perceptron model as well as the perceptron learning rule. You must be able to state the associated convergence theorem

16. You must be able to understand and explain how the perceptron can be extended into a neural network to learn non linearly separable datasets

17. You must be able to provide the general expression for a neural network and draw the corresponding diagram

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

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

20. 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

21. You must be able to explain how K means clustering can be extended to fit Gaussian mixture models to a dataset through the EM algorithm.

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

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

24. 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 each other.

25. You must be able to explain how the Principal component analysis problem can be solved through the singular value decomposition of the prototype matrix (or equivalently the eigenvalue decomposition of the empirical covariance matrix). You must be able to explain how to compute the projection onto the PCA plane from those decompositions.

26. You must be able to understand the assumptions behind the independent component analysis model and provide one algorithm for this model (Fast ICA, MLE,..)

27. You must be able to list and describe the main dimension reduction approaches covered in the course notes. This include Multidimensional scaling (classical and metric), Isometric feature mapping (ISOMAP), Locally linear embedding (LLE) as well as Self Organizing map (SOMs).

28. For all the aforementioned approaches, you must be able to provide the pseudo code.