CSCI-UA 9473,
Introduction to Machine Learning
List of questions for the midterm evaluation

October 2018

Part I: Statistics

1. State Bayes' Theorem and explain how it is used in the Bayesian framework as opposed to the frequentist framework
2. Describe each approach by listing a few key characteristics.
3. Give the expression of the Maximum A Posteriori (MAP) estimator and the Maximum Likelihood estimator (MLE). Explain how each estimator can be related to one of the two approaches.
4. Explain how the MLE can be understood as a particular case of the MAP.

Part II: Linear regression

5. What is a linear regression model (Mathematical formulation)? How to compute the regression coefficients? What does the model look like for a given set of training points?
6. How can we predict the quality of prediction of a linear regression model?
7. What does the mean squared error (MSE) represent and how does it decompose ? What is the influence of model complexity on this decomposition?
8. What is a BLUE estimator. Give an example of such estimator and explain the advantage/disadvantage of this example.
9. Given a linear regression model with correlation among the variables, give three possible approaches to mitigate the influence of correlated variables and explain the result of each approach. Why is it bad to have correlated features ? What is the effect of correlation on the MSE/Bias-Variance decomposition?
10. Represent each of the regularization approaches that can be used to reduce the variance in linear regression.

Part III: Linear classification

11. What is the simplest example of a two classes classification model? Give the mathematical expression of the model and the problem that one has to solve to get the coefficients.
12. Give the two classes of probabilistic classification models and explain where the names come from. Give an example from either of these classes.
13. Give the expression of the logistic regression classifier and explain why it is considered as a generalized linear model.
14. *What is Linear discriminant analysis?

Part IV: Kernels and SVM

15. Explain the difference between parametric and non parametric classifiers. Give one example from each class of models.
16. List the advantages of parametric and non parametric methods. When is each approach most useful?
17. Give the two main frameworks in which Kernels are used.
18. Give two possible approaches at turning feature based models into similarity based models. Using the example of ridge regression, show how the kernel trick can turn a feature based model into a model based on similarity
19. Support Vector Machines have two main characteristics. List those characteristics and illustrate them by drawing an SVM classifier on top of some dataset.
20. *Derive the expression of the separating plane corresponding to an SVM classifier. Based on your calculations and final result, explain why those models are also called sparse vector machines.

Part V: Neural Networks

21. The first model of an artificial neuron was developed by Rosenblatt in 1957/1958. Give the expression of this first model. Why is considered a linear classification model?
22. The perceptron is particularly interesting because of a simple learning rule (how to update the weights) and its associated convergence Theorem (Perceptron convergence theorem). State the theorem and explain how to update the weights in the perceptron algorithm.
23. Recount the main events that led to the development of neural networks.
24. Explain how artificial neurons relate to biological neurons.
25. Give three examples of activation functions.
26. State the Universal Approximation Theorem
27. What is backpropagation used for ? give the steps used in this approach.
28. * Give the mathematical derivation used in backpropagation and illustrate your mathematical derivation with a drawing.