CSCI-UA 9473. Machine Learning
Material for the Midterm

Augustin Cosse

Fall 2022

1 Theory

1. You must be able to provide and explain the residual sum-of-squares loss associated to the linear regression model. You must be able to derive the gradient iterations on that loss as well as the closed form solution for the regression coefficients (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 explain how cross validation can be used for model selection (e.g. in best subset selection).

4. You must be able to explain the statistical assumptions behind the sum-of-squares, Ridge and Lasso objective functions. You must be able to describe the distributions involved in each regularization approach.

5. You must be able to compare the regularization approaches in terms of their respective (feature selection) efficiency and their computational complexity

6. You must be able to explain the notion of bias-variance tradeoff and how it can be used to explain the connection between model complexity and overfitting.

7. You must be able to explain how to use the Residual Sum-of-Squares objective to learn a binary classifier

8. You must be able to explain how the binary classifier can be extended to more than two classes (i.e. one-vs-rest, one-vs-one, multiple discriminant)

9. You must be able to explain and derive the expression of the logistic regression model.

10. You must be able to motivate the binary cross entropy loss (a.k.a log loss) from the maximization of the likelihood for the logistic regression classifier.

11. You must be able to explain the difference between generative and discriminative classifiers. You must be able to provide at least one model from each family.

12. You must be able to give the general expression of the Gaussian Discriminant Analysis (GDA) model and provide the expression of the parameters under the shared covariance assumption (Linear Discriminant Analysis)

13. You must be able to explain the assumption behind the Naive Bayes classifier and give zt least one particular example of such a classifier.

14. You must be able to explain the perceptron model as well as the perceptron learning rule and state the associated convergence theorem
15. You must be able to provide the general expression of a neural network and draw the corresponding diagram
16. You must be able to explain the backpropagation algorithm (list the main steps) and provide the associated equations.

2 Coding

You must be able to provide pseudo-code for:

1. Gradient descent applied to the ℓ_2 loss (i.e. residual sum-of-squares loss) for the linear model, including with additional polynomial features and regularization (LASSO and/or Ridge)
2. The solution of the normal equations for the least squares loss and the Ridge loss.
3. Cross validation (e.g. K-fold cross validation)
4. Linear (binary) classification through gradient descent and its extension to multiple classes through the one-vs-rest, one-vs-one and multiple discriminant approaches
5. Gradient descent for logistic regression for the log-loss.
6. The Perceptron algorithm
7. Forward propagation and backpropagation in a simple neural network (2 hidden layers and a couple of units per layers)