CSCI-UA 9473. Machine Learning
Material for the Midterm

Augustin Cosse

Summer 2022

1 Theory

1. You must be able to explain the residual sum of squares criterion associated to the linear regression model. You must be able to derive the gradient iterations on that criterion as well as the closed form solution for the regression coefficients (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, Ridge and 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 the bias variance decomposition of the Mean Squared Error (MSE)
6. You must be able to explain how to use the Residual Sum of Squares criterion to learn a binary classifier
7. 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, multiple discriminant through one-hot encoding)
8. You must be able to explain and derive the expression for the logistic regression classifier
9. You must be able to explain the perceptron model as well as the perceptron learning rule and the associated convergence theorem
10. You must be able to provide the general expression for a neural network and draw the corresponding diagram

2 Coding

You must be able to provide pseudo-code for:

1. Gradient descent applied to linear regression with the ℓ_2 loss, including with additional polynomial features and regularization (LASSO and Ridge)

2. The solving 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 multiclass frameworks through the one vs rest, one vs one and multiple discriminant approaches

5. Gradient descent for Logistic regression within the MLE framework (i.e for the log-loss or binary cross entropy loss)

6. The Perceptron algorithm