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

March 2019

1 Statistics

1. Describe the difference between the Bayesian and the Frequentist frameworks in parameter estimation. Illustrate the difference between those two frameworks using Bayes' Theorem. Give an estimator from each framework.

2. We consider data points $\{\boldsymbol{x}_i\}_{i=1}^N$ that follow a Gaussian distribution with mean μ and variance σ^2. Give the expression of the MLE for μ, for σ.

3. We consider data points $\{\boldsymbol{x}_i\}_{i=1}^N$ that follow a Gaussian distribution with mean μ and variance σ^2. Now we assume a Gaussian prior on μ (mean λ and known unit variance). What problem do we have to solve to get the MAP?

2 Regression

1. What problem do you solve when you want to find the regression coefficient of a line approximating the distribution $\{\boldsymbol{x}_\mu, t_\mu\}_{\mu=1}^N$ for points $\boldsymbol{x} \in \mathbb{R}^D$

2. For a general set of points $\{\boldsymbol{x}_i\}_{i=1}^N$ with associated target labels t_i, derive the (general) expression of the regression coefficients.

3. In linear regression, what happens when the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ is badly conditioned. Explain for what type of data this can occur.

4. Give three possible approaches at regularizing the regression problem. Give the mathematical expressions of those regularized models. Finally, given these models, explain how one can solve for the parameters.

5. Explain the notion of bias/variance tradeoff and its connection to model complexity (see the 2018 notes).

6. Explain the notion of BLUE estimator (again see the 2018 notes).

3 Classification

1. What is the simplest way to define a linear classifier for a two classes dataset. How is the class prediction $\hat{y}$ defined?

2. How can one extend the binary classifier to multiple classes. List three possible approaches.

3. Explain the concept of Maximal Margin Classifier. Give the mathematical expression of the classifier. Using this expression, explain the two key characteristics of that classifier

$\boldsymbol{x}$	y
$(0,0)$	0
$(1,0)$	1
$(0,1)$	1
$(1,1)$	0

Table 1: XOR dataset

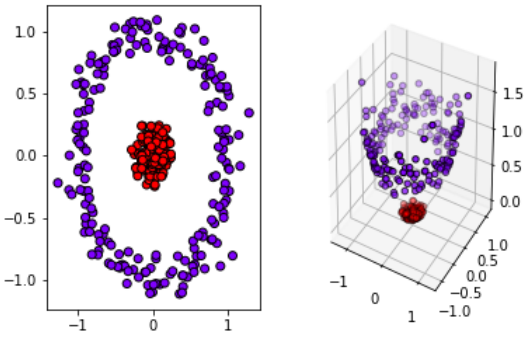

Figure 1: Concentric dataset 1

4. Maximal margin classifiers can be obtained by solving an optimization problem. Give the formulation of this problem and explain the meaning of the objective function that one optimizes to find the classifier, based on the notion of minimum distance between a point and a plane.

5. List the two main families of probabilistic classifiers and explain what is the probabilistic model learned by each family. Give an example from each family.

4 Kernels

1. In what framework is it particularly interesting to rely on Kernels rather than on the feature vectors. Give an example.

2. Explain how to modify the linear regression model to obtain a kernel regression model. Give the expression of the Kernel regression model.

3. Consider the XOR problem shown in table 1. Explain how this problem can be solved by using an appropriate kernel.

4. Same question for the concentric two classes dataset of Fig. 1. Explain how one can solve this problem. What kernel do you use? What is the expression of the classifier?

5 Neural Networks

1. Explain Rosenblatt's Perceptron model. Give the mathematical expression of the model. Give the perceptron learning rule. What can we say regarding the convergence of this learning rule?

2. Give the mathematical expression of a neural network and illustrate your model with a Diagram.

3. The power of neural networks rely on a result known as *Universal approximation Theorem*. Explain this result.

4. How do we train a neural network? Explain Backpropagation on a simple two hidden layer neural network (you can fix the number of neurons in each layer to some small constant of your choice if it helps you.)