(February 24, 2015)

Sample Questions for Midterm Exam
CSCI-B555

(DO NOT DISTRIBUTE)

Problem 1. Miscellaneous

1.1. (2 points) Briefly state what we mean by "learning" in Machine Learning.

1.2. (2 points) What is the main difference between supervised and unsupervised learning?

1.3 (2 points) Find groups of synonyms among the following words used in machine learning:

example, attribute, feature, data point, target, input, weight, label, parameter, pattern

1.4. (2 points) Explain standard measures of accuracy for classification and regression? What is the range of their acceptable values? Address specifically binary classification.

1.5 (2 points) What is the purpose of splitting the data set into training, validation, and test. What is each set used for?

Problem 2. Elements of Probability Theory

2.1 (2 points) Briefly discuss the main reason(s) why probability theory is useful at modeling uncertainty.

2.2 (3 points) Let A, B, and C be some elements of the event space $\mathcal{F}$. State the conditions of mutual independence between the three events.

2.3 (3 points) Let Ω be any abstract space and some event space be defined as $\mathcal{F} = \{\Omega, \emptyset\}$.

a) (2 points) Define at least one probability measure P for this space.
b) (1 point) Is P unique?

2.4 (2 points) State the axioms of probability.

Problem 6. Linear Regression

6.1 (3 points) Among the $k = 20$ features in a regression data set D, the 3rd feature can be expressed as a linear combination of the first two, e.g. $f_3 = 2f_1 + 3f_2$. You attempt to use OLS linear regression on such data. What will happen?

6.2 (2 points) Provide a geometric explanation to what is minimized in ordinary-least-squares linear regression?

6.3 (3 points) The solution to the OLS linear regression method is $\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. Matrix $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is called a projection matrix. What does it project and where?

6.4 (2 points) What is Lasso regression?

Problem 7. Maximum-Likelihood Principles

7.1. (10 points) You are given a data set D = {0.1, 0.4, 0.2, 0.6} of numbers sampled independently from an exponential distribution with

$$p(x|\lambda) = \lambda \cdot e^{-\lambda x}$$

a) (4 points) Calculate the log-likelihood function that D was generated from an exponential distribution with parameter $\lambda = 10$, i.e. derive $\log P(D|\lambda = 10)$.
b) (4 points) Derive gradient descent algorithm for calculating the optimal λ that results in maximization of the likelihood.
c) (2 points) Could one derive the closed-form solution for optimal λ.