CSCI 635: Introduction to Machine Learning
Homework 2: Basics of Statistical Learning

Due Date: Saturday, October 24, 11:59pm *Late:* Sunday, October 25, 11:59pm

Instructions: The assignment is out of 50 points. You will need your textbook "Deep Learning" by Goodfellow et al. (DL), which contains the required readings for this class, i.e., Chapters 1-3 & 5, up until this point. Submit your assignment through the Dropbox in MyCourses as a single *.pdf file named according to the convention `<your_last_name>-hw2.pdf`. The document should:

1. should be written in Latex and compiled to a PDF. You must integrate directly any plots you generate externally for any relevant questions that require them.
2. in its text, the file should map your answers to the correct problem number. For answers to various questions throughout, please copy the question(s) being asked and write your answer underneath (as well as the problem number).

Grade penalties will be applied (but not limited to) for:

- Not submitting the write-up as instructed above.
- Submitting with incorrect file names.

Chapter #1

- **Question 1:** What is representation learning? Give an example of it – where would it be applied? What are the factors of variation?

- **Question 2:** Explain the relationship between an artificial neural network's architecture, e.g., number of layers, number of nodes etc., to its complexity.

Chapter #2

- **Question 1:** What is Eigenvalue Decomposition? Explain/define (in words) what is: 1) a positive definite matrix, 2) a positive semi-definite matrix, 3) a negative definite matrix, and 4) a negative semi-definite matrix.

- **Question 2:** Give 2 applications where we would use Singular Value Decomposition (SVD).

Chapter #3

- **Question 1:** What is the difference between absolute independence and conditional independence? Explain in words (referring to the equations in the book): what is expected value, variance, and covariance?

- **Question 2:** Briefly explain the different kinds of statistical distributions (make sure to mention/refer to at least 4 different specific types of named distributional models). What is a mixture model and why might it prove useful in modeling some types of data?

- **Question 3:** Suppose that we have three coloured boxes: r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2$, $p(b) = 0.2$, and $p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

- **Question 4:** Generate a set of random points and plot graphs for: 1) the Gaussian distribution, 2) the exponential distribution, 3) the Laplace distribution, and 4) the Dirac distribution. Note that you will need to write code in order to solve this problem, e.g., Python.
  Bonus: For each distribution, check for 3 different values of mean and variance and compare your estimates across the 3 samples (do not forget to seed your code).

- **Question 5:** What is the Kullback-Leibler (KL) Divergence? (describe it **in words**, referring to the equation/formula in the book). Give an example of when KL Divergence would be particularly useful.

- **Question 6:** Write the probability distribution for each of the following two graphs in the diagram below, i.e., graph a and graph b.

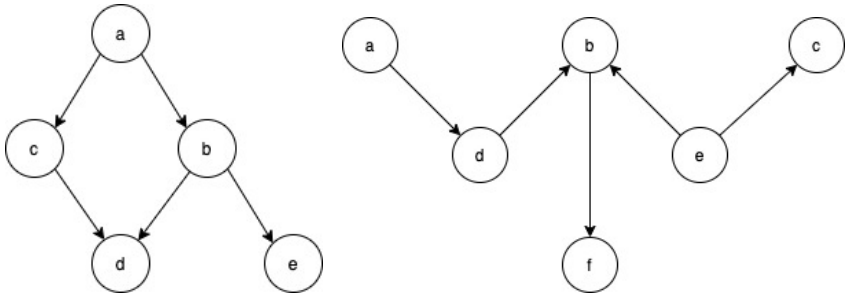

Figure 1: Graph a (left) and graph b (right).

Chapter #5

- **Question 1:** What is the no free lunch theorem (NFLT)? Why is it important for statistical learning in general and how would it affect your choice of a statistical model to apply for any particular problem?

- **Question 2:** When is it is appropriate to use an L1 (Laplacian) regularization term? When is it appropriate to use an L2 (Gaussian) regularization term)? What effect does L1 and L2 regularization have on model weights?

- **Question 3:** Describe the steps to be taken in each case when the bias is high and when the variance is high. How do we balance the trade-off in each case?

- **Question 4:** When would you use the categorical cross entropy loss and the mean squared error, respectively?