Stony Brook University
CSE512 – Machine Learning – Fall 18
Homework 1, Due: Sep 4 at midnight 11:59pm
Version 1 - last updated 28 Aug 2018

This homework contains two questions. The last question requires programming. The maximum number of points is 100 points. For this homework, you should review some material on probability. There are many review notes and tutorials on the Internet and you are allowed to use Google to find them.

1 Question 1 – Probability (60 points)

Consider a positive integer N and let X_1 and X_2 be independent, discrete random variables uniformly distributed from 1 to N. Let $X = \max(X_1, X_2) - X_1$. Compute:

1. *(20 points)* The expectation: $E(X)$.
2. *(20 points)* The variance: $Var(X)$.
3. *(20 points)* The covariance: $Cov(X, X_1)$.

2 Question 2 – Programming (40 points)

Consider the random variables X_1, X_2, X defined as in Question 1. Write a Matlab (or equivalent Python) function with the below signature: `[E, V, C, H] = question2(N, M)` The inputs to the above function are positive integers N and M. M is the number of sample pairs of (X_1, X_2). The outputs of the function are four scalars E, V, C, H for $E(X)$, $Var(X)$, $Cov(X, X_1)$, and the entropy $H(X)$. Plot E, V, C, H as the function of $M = 100, 200, 300, \cdots, 9900, 10000$ for $N = 100$. Produce four separate plots for E, V, C, H. You are allowed to use built-in functions of Matlab such as `mean, var, cov`. Other programming languages might have similar built-in functions, and you can use them if you want.

3 What to submit

You will need to submit both your code and your answers to questions on Blackboard. Put the answer file and your code in a folder named: SUBID_FirstName_LastName (e.g., *10947XXXX_lionel_messi*). Zip this folder and submit the zip file on Blackboard. Your submission must be a zip file, i.e, SUBID_FirstName_LastName.zip. The answer file should be named: answers.pdf. The first page of the answers.pdf should be the filled cover page at the end of this homework. The remaining of the answer file should contain:

1. Answers (and derivations) to Questions 1.1, 1.2, and 1.3
2. Four plots for E, V, C, H as the functions of M.

You can use Latex if you wish, but it is not compulsory.

4 Cheating warnings

Don't cheat. You must do the homework yourself, otherwise you won't learn. You cannot ask and discuss with students from previous years. You cannot look up the solution online.

Cover page for answers.pdf
CSE512 Fall 2018 - Machine Learning - Homework 1

Your Name: Sriram Reddy Kalluri

Solar ID: 111878857

NetID email address: skalluri@cs.stonybrook.edu & sriram.kalluri@stonybrook.edu

Names of people whom you discussed the homework with: