Homework 1

Professor Lydia Y.Chen

CS4215: - Quantitative Performance Evaluation for Computing systems

September 3, 2020

Exercise 1. (15 Points)
The following output was obtained from a learning algorithm that performed a three-factor ANOVA with replication (report how many!) on a factorial experiment.

1. Fill in the blanks in the ANOVA table. You can use bounds on the P-values.
2. How many levels were used for factor B?
3. Based on the levels used for each factor, how many experiments are required if there is significant interaction among factors?
4. What conclusion would you draw about this experiment?

Three-factor ANOVA: y versus, A, B, C

Source	DF	SS	MS	F	P
A	1	1.367	-	-	-
B	-	7.825	7.825	-	-
C	2	1.235	-	-	-
Interaction (AB)	-	0.012	-	-	-
Interaction (AC)	-	0.565	-	-	-
Interaction (BC)	-	-	0.060	-	-
Interaction (ABC)	-	-	-	-	-
Error	60	1899.044	-		
Total	71	1911.418			

Exercise 2. (20 Points)
Consider an image classification task on four UCI datasets, *mushroom*, *bank*, *car evaluation*, and *zoo* (https://archive.ics.uci.edu/ml/datasets.php), using the following three classifiers, K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Multilayer Perceptron (MLP). The classifiers are called A1, A2, and A3 respectively and the datasets B1 through B4 are shown in table below. The goal of this exercise is to see if the classifiers are different from each other in accuracy. You need to train each classifier for each dataset and report the accuracy using the steps below:

1. Find the test classification accuracy of each classifier for each dataset. For each dataset use 80% of the data for training and 20% of the data for testing. Report the accuracy in percentage ($\times 100$).
2. Repeat the experiments for two more times by changing the train and test set for each dataset and fill the table below.
3. Are the classifiers different from each other at 90% level of confidence?
4. What percentage of variation is explained by the learning model-dataset interaction?

Test Accuracy(%)			
	A1	A2	A3
B1	-	-	-
	-	-	-
	-	-	-
B2	-	-	-
	-	-	-
	-	-	-
B3	-	-	-
	-	-	-
	-	-	-
B4	-	-	-
	-	-	-
	-	-	-

Exercise 3. (15 points)
Analyze the 2^{4-1} design shown in Table below:

	C_1D_1	C_1D_2	C_2D_1	C_2D_2
A_1B_1	-	40	15	-
A_1B_2	-	20	10	-
A_2B_1	100	-	-	30
A_2B_2	120	-	-	50

1. Quantify all main effects.
2. Quantify percentages of variation explained.
3. List all confoundings.
4. Can you propose a better design with the same number of experiments.
5. What is the resolution of the design?

Exercise 4. (8 Points)
After spending months carefully building his closed batch data storage system, David comes to see his advisor with the following description and measurements: The MPL for the system

is fixed at 19 jobs. David explains that 90% of jobs find the data they need in the cache, and hence their expected response time is only 1 second. However, 10% end up having to go to the database, where their expected response time is 10 seconds (for just the database). David's advisor asks one question: "How many jobs do you see on average at the database?" When David answers "5" his advisor says he needs to go back to the drawing board. What went wrong?

Exercise 5. (15 Points)
Consider an interactive system with a CPU and two disks. The following measurement data was obtained by observing the system:

- observation interval = 17 minutes
- mean think time = 12 seconds
- number of complete transactions during observation interval = 1,600
- number of completions at CPU = 1,600
- number of fast disk accesses = 32,000
- number of slow disk accesses = 12,000
- CPU busy time = 1,080 seconds
- fast disk busy time = 400 seconds
- slow disk busy time = 600 seconds

1. Give asymptotic bounds on throughput and response time as a function of the number of terminals.
2. Now consider the following modifications to the system:
   (i) Move all files to the fast disk
   (ii) Increase the CPU speed by 50% (with the original disks).

   For each of these two modifications, compute the effects on the original system. Explain in words the effect when the multi-programming level, N, is small and when N is large.

Exercise 6. (12 Points)
Data centers alternate between "working" and "down." There are many reasons why data centers can be down, but for the purpose of this problem we mention only two reasons: (i) a backhoe accidentally dug up some cable, or (ii) a software bug crashed the machines. Suppose that a data center that is working today will be down tomorrow due to backhoe reasons with probability $\frac{1}{3}$, but will be down tomorrow due to a software bug with probability $\frac{1}{6}$. A data center that is down to day due to backhoe reasons will be up tomorrow with probability 1. A data center that is down today due to a software bug will be up tomorrow with probability $\frac{5}{6}$.

1. Draw a DTMC for this problem.
2. Is your DTMC ergodic? Why or why not?
3. What fraction of time is the data center working?
4. What is the expected number of days between backhoe failures?

Exercise 8. (15 points) This is a review exercise and meant to showcase you the potential project topics. To complete this exercise, you need to strictly follow the review template available on the brightspace `https://brightspace.tudelft.nl/d2l/le/content/280410/viewContent/1919112/View`. You need to first select ONE of the following four papers which talk about performance optimization issues of machine learning clusters, consisting of heterogeneous CPU and GPU. The details of this exercise and papers are provided on the `https://brightspace.tudelft.nl/d2l/le/content/280410/viewContent/1919112/View`. Make sure you check them before you start the exercise.

1. AlloX: compute allocation in hybrid clusters
   https://dl.acm.org/doi/abs/10.1145/3342195.3387547
2. HetPipe: Enabling Large DNN Training on (Whimpy) Heterogeneous GPU Clusters
   https://www.usenix.org/system/files/atc20-park.pdf
3. Nexus: A GPU Cluster Engine for Accelerating DNN-Based Video Analysis
   https://homes.cs.washington.edu/ arvind/papers/nexus.pdf
4. NeuOS: A Latency-Predictable Multi-Dimensional Optimization Framework for DNN-driven Autonomous Systems
   https://www.usenix.org/system/files/atc20-bateni.pdf