ML and BD in Healthcare, Homework #3

1. Answer homework questions below
2. When ready, submit your answers online as instructed

# Answer the questions

**Question 1:** (1 pt)

Given patient arrival, exam begin, and exam complete (end) timestamps, one can find the following

1. Patient wait time
2. Patient wait line size
3. Patient exam length
4. All 3 choices

**Question 2:** (1 pt)

Pushing resource utilization to 100% will usually result in

1. Making wait lines shorter
2. Making wait times longer
3. Making wait times shorter
4. Staying perfectly on time

**Question 3:** (1 pt)

When introducing new technology into healthcare, one often fails because the technology

1. Is not using HL7
2. Is not patient-friendly
3. Does not address real problem
4. Does not run on mobile devices

**Question 4:** (1 pt)

Scheduling optimization analysis discussed in the class was based on

1. GPS
2. WiFi indoor tracking
3. Manual counts of patients in the waiting rooms
4. HL7

**Question 5:** (2 pts)

An exam room in a hospital is breaking for lunch from 12:00 to 13:00. You suspect that you can hire extra staff members to eliminate this break, and you will even make profit by doing so (that is, taking additional patients in the 12:00-13:00 time range will pay the salaries for the extra staff).   
What time stamps should you use to evaluate, in the most realistic way, the financial impact of the break time, and your gain from its elimination?

* Patient arrival and exam begin times
* Patient exam begin and end times
* Patient arrival and exam end times
* Lunch time (12:00-13:00)

# Solve data problems

One of the main challenges in any machine learning problem is to compute good features, which may not be readily available in the original data. Sometimes the most primitive data can discover find non-trivial patterns, which can be used in your models. Let’s try?

File XRays contains very basic data: a log of patients coming for their Xray examinations. Imagine a large facility with several Xray units, where a patient can come to have one (or more) Xray exams done. All you have in this log is unique patient IDs (MRN, “Medical Record Number”), and three timestamps: patient arrival time, Xray exam begin and complete (end) time. Can we learn something more interesting from this trivial data?

*Hints:*

* *To solve the problems below, you will have to sample facility events at different times of day. For consistency, let’s do this every half hour. That is, you should define your time sampling array as numpy.arange(7, 20, 0.5) – that is, going from 7am to 8pm with 0.5-hour increment. I also suggest plotting each of the metrics below as a function of this time.*
* *Unsurprisingly, the data may contain errors: as we know, timestamps are usually entered manually. Do not make any effort to fix them – take the data as is.*

**Question 6:** (6 pts)

As I mentioned, there are several Xray units (exam rooms) in this facility. At any given time, only at most one patient can be in any Xray room. Let’s also assume, that at some times all rooms were used. How many Xray rooms does this facility have?

1. 4
2. 6
3. 8
4. 10
5. 12
6. 14

**Question 7:** (6 pts)

The facility realized that at certain times its waiting area gets crowded, and wants to expand it. The waiting area is occupied by the patients only. Based on the data provided, the waiting area should be able to hold:

1. 5 patients
2. 10 patients
3. 15 patients
4. 20 patients
5. 25 patients
6. 30 patients

(note that waiting area renovation is expensive, and the facility does not want to provide more room than needed)

**Question 8:** (4 pts)

Overall, the peak utilization (total number of Xray exams done during the entire log time) happens during the following time interval:

1. 7:00-9:00
2. 10:00-12:00
3. 13:00-15:00
4. 14:00-16:00