Project 2 Part 1: Bayesian Inference and Decisions

Due Tuesday March 12, 2024 by 11:59 pm on Canvas

The deliverable for this project is a report detailing your approach and answers to all of the questions in each section.

Please submit a single PDF file that contains your text and figures. All code should be in the *appendix*, unless you want to provide a BRIEF 5 line snippet that contains only the algorithm you are describing within the text. In the latter case, make absolutely sure that this algorithm follows the narrative of the text. For every figure that you obtain, please make sure that the axes are labeled, any relevant legend is shown, and a caption describes what is being plotted.

The preference is that a formal typed report is submitted. However, a scanned written report that is legible will be accepted. Illegible reports or reports that are primarily/entirely code with no text will not be evaluated.

The same guidelines outlined in Project 1 and the Syllabus apply here.

1 Find the source!

You have received a fancy new flying robot that has an excellent sense of smell, all thanks to a brand new sensor. Unfortunately the sensor draws a lot of power, and the robot can only use it a couple of times before needing to take a break. You will use this flying robot and sensor to play a game that tries to uncover a buried object. The field of play is two-dimensional (X, Y) and normalized to $[-1, 1] \times [-1, 1]$. An object is buried somewhere in this field of play, and this object gives off a very strong scent. This scent has spread across the entire field of play.

If the ground was structured normally the strength of the scent would dissipate with distance from the object. However, the ground is not structured normally and instead creates many "pockets" that trap the scent and irregularities that make the scent spread out in some more interesting fashion. Nevertheless the location where the scent is strongest is the location where the object is buried. The robot must search for the location of the object, and learn to not be fooled by the localized pockets of scent!

Your sensor provides you with a single number that indicates the strength of the scent — the greater the number, the stronger the smell. I will start you off with an initial set of sensor readings:

X	Y	Sensor
0.1	0.05	3.39382006
-0.9	0.3	3.2073034
0.2	0.4	3.39965035
0.8	-0.3	3.68810201
-0.6	0.3	2.96941623
0.3	-0.2	2.99495501
0.5	-0.84	3.94274928
-0.5	0.85	2.7968011
-0.01	-0.76	3.34929734
-0.9	-0.9	3.91296165

Table 1: Initial Sensor Data

From this set of data, you can **begin to draw a map indicating how the scent is distributed across the field of play. You can also draw maps or other visualizations of the uncertainty in the strength of the scent or in the location of the object.** A map is a two-dimensional surface — that is a function of (X,Y). Drawing such maps will help you decide where to search next[1] Your task is to develop a search strategy for the robot. Using your sensor requires obeying the following rules:

1. If your last name comes at or before "Mi" in the alphabet, then Prof Gorodetsky is your sensor. Otherwise Liliang Wang is your sensor. You can request information from your sensor three times. This request must be performed over email and must strictly follow the following protocol:
   (a) Subject: [AE567] Sensing Request # <request number, 1, 2 or 3>
   (b) Content: Sensing requirements attached
   (c) Attachment: A single .txt file that contains up to 12 rows and 2 columns specifying the requested sensing location on a 2D map. No comma separated lists, just floating point numbers. The filename must be "requestX.dat" where X is the request number (1,2 or 3).
   (d) The sensor begins to break down after Friday March 8th, no responses are guaranteed for sensor requests submitted after midnight of this day.
2. You will receive a reply that includes an attached .txt file with the sensor values at your requested locations.
3. The combined number of sensing locations amongst all **three** requests **cannot exceed twelve**. For example, you can make three requests each with four sensing locations, you can make one request with all twelve sensing locations, or you can have something else.

[1]For example you can use `contourf` to visualize 2D maps.

1.1 Deliverables

Your deliverable has to document all of the things that you have tried to find the location of the object. I suggest that you use the Gaussian process regression algorithms that we covered in class. The goal of this question is to enable you to understand how probabilistic thinking and the Bayesian point of view is used to make decisions. Here are *some* of the discussion points that are required:

1. What is uncertain in this problem? What are you trying to learn? What quantities can you use to summarize your uncertainty? You must be as specific as possible, you cant just say "my uncertainty is represented as variance". Variance of what? What is the random variable you are talking about?

2. How did you mathematically formulate this problem? Please be as specific as possible, to the extent that your results would be reproducible. For example: what is your search space? How is the decision with regards to the next sensor request made? How is it related to the uncertainty you described in the previous question? In general, what is the algorithm[2] you use to determine where next to request sensor readings. Can you provide pseudocode of the algorithms you considered and why you considered them?

3. Relatedly, list at least three possible candidates for the strategy (algorithms for requesting next points) that you considered prior to making a choice? Which one did you choose? Why? A decision making strategy is an algorithm which takes in available information and outputs a sensor request. Provide details on algorithm performance such as

   (a) How did you use the probabilistic description of the map to inform your decisions?

   (b) How sensitive did you find your predictions to be to the choice of models (kernels, parameters, hyperparameters)? How did this impact your approach?

4. When discussing your regression choice, be clear on the folloiwng points. How did you choose the GP kernel (nonparametric) or basis functions/features (parametric)? How did you choose the parameters? What did you do about any hyperparameters?

5. What is your final prediction about the scent location?

6. How confident are you in your answer? Please provide some probabilistic description. If you would like to provide several answers (or a region of possibilities), then please provide your confidence in each one (or all of them). Provide credible intervals where appropriate.

Support your answers above with relevant figures including, but not limited to,

[2] An algorithm is a sequence of rigorous instructions, typically used to solve a class of specific problems or to perform a computation.

1. Plots of your guess of the map of the scent at each stage of your decision making process (both the prior and after each data set is received), include some characterization of uncertainty.

2. Plots indicating how and why you chose to obtain more data in the locations of your request, e.g., plots of some objective function or criterion you are using in your decision making.

3. Plots indicating the evolution of the probability that the scent is coming from some specific points in the X-Y space over the course of the decision making process.