In class and in recent assignments, you have learned how to construct and evaluate basic statistical models that relate neuronal signals to various input stimuli. The underlying idea of statistical modeling is that neurons encode information (environmental or biological) in their firing patterns.

Working in groups, you will construct statistical models using a population of hippocampal cells that encode spiking activity and decode spatial information. This population includes both place (unimodal and multimodal) cells and grid cells. These neurons fire when an animal is in specific locations in its environment. For more information on these cells, please refer to the papers provided.

Part 1: Encoding

In the encoding phase, information is translated by a cell into spiking activity. Your job is to define what information is needed to accurately model these cells.

A model is defined by the conditional intensity function, which is in turn defined by your choice of covariates. Your task is to define at least three different models that encode the activity of all 10 neurons by defining three sets of covariates. Justify your choice of covariates by referencing the data or papers. At least one model should include history-dependence. Show results of your covariates (see papers for reference on how to plot covariates). Which model was the best for each neuron? Validate your models by measuring statistics such as AIC and KS.

By using the same set of covariates to build models of the same neuron, you can compare the different weights that each neuron puts on the covariates. Which covariate was more important for which cell?

You are free to define your own covariates or use covariates that you come across during your literature search. Some examples of possible covariates include:

- Position information (or functions of)
- Short term history
- Long term history
- Ensemble activity
- Velocities
- Etc.

Interspike interval

Part 2: Decoding (bonus!) 20 points

In this part of the project, you will design a decoding strategy that will use your best encoding strategy to estimate the trajectory of the rat using neuronal activity from your encoding model.

In order to accomplish this, you should define a function that uses one of your encoding strategies from part 1 to estimate the position of the rat at each time point based on spiking activity. Your decoder should be trained using the data from `train.mat`. I will then feed your trained model a novel data set, called `test.mat`, to calculate the accuracy of your decoded position compared to the actual position using the script called `decoder_accuracy.m` on the last day of presentations. Assume `test.mat` has the same format as `train.mat` with similar spiking behavior. You must define your decoder function generally enough such that I can input the data from `test.mat` and it will output the estimated (x,y) position.

You may use any decoding strategy that you find in literature. For a place to start, please refer to the paper by Agarwal, et al.

The amount of extra credit points your group receives will depend on the performance of your decoder on the data set `test.mat` for up to 20 points added to this project's grade.

Groups

You will work in groups of 2-3. Be sure to form groups that have a balance of strengths and weakness. Please confirm your groups with the TA by 11/15/2017. You may choose a team name if you so desire.

Data

In the experimental protocol, a Long-Evans rat was freely foraging in an open field arena (~1 m radius) for a period of 36.36 minutes. Custom micro-electrode drives with variable numbers of tetrodes were implanted in the rat's medial entorhinal cortex and dorsal hippocampal CA1 area. Spikes were acquired with a sampling rate of 31.25 kHz and filter settings of 300 Hz-6 kHz. Two infra-red diodes alternating at 60 Hz were attached to the micro-electrode array drive of the animal for position tracking. Spike sorting was accomplished using a custom manual clustering program (Xclust, M.A. Wilson). All procedures were approved by the MIT Institutional Animal Care and Use Committee.

In this selected recording session, m=74 neurons were recorded from both entorhinal cortex (m=8) and dorsal hippocampus (m=66). Out of these 74 neurons, 21 (6 from entorhinal cortex) neurons were discarded as they fire either independently of position (x,y) (i.e., putative interneurons) or they had very low firing rates (<0.175 Hz). Out of remaining m=53 neurons, m=27 (1 from entorhinal cortex, 26 from hippocampus) were unimodal and m=26 multimodal (1 from entorhinal cortex, 25 from hippocampus).

In this smaller data set, you have a total of 10 cells. The first 5 cells in the data set are multimodal cells, and the last 5 are unimodal cells.

In MATLAB File called `train.mat`:

- `xN`, vector containing x coordinates of rat's position in meters
- `yN`, vector containing y coordinates of rat's position in meters
- `spikes_binned`, matrix that contains the spike train data for the 10 cells. The rows contain the binary data of whether a spike occurred at each time point and the columns are the cells.

All data is stored in bins of 1 ms.

Final presentation

You and your group will present your results in the form of a presentation, which will take place on 12/06/2017 through 12/08/2017. Each group will have 12 minutes to present (10 minutes to present plus 2 minutes for questions). The presentation order will be random.

Your final presentation should follow the standard research outline: background, methods, results, discussion, conclusion, references. **Please make sure that your figures are legible to those seated in the back row.** It should also include:

- Plot of the raw data
- Questions of interest
- ISI histograms

- For each model:
    - List your covariates and motivate your model structure using raw data
    - Define your model equation
    - Plot covariate results (see papers) and goodness-of-fit
- Which model was the best? Discuss how you justified your decision with statistics such as AIC.
- Based on your results, what are the cells encoding?
- (Bonus) State your decoder strategy.

Please submit your final presentation, along with your code, to Blackboard by 12/06/2017 @ 9 am.

Grading
You will receive a group score (180 points) and an individual score (20 points) for a total of 200 points. Your group score will be based off your final presentation. The individual score will be based off peer reviews from the other members of your group. Each member of the group is expected to review every other member. Peer reviews are taken on Blackboard, which can be found in the Homework section. Please complete your review(s) by the day of your presentation.

There is also the opportunity for extra credit of up to 20 points on this project, as stated above.

Good luck!