Empirical Bayes and the Laplace approximation
Final Projects

Final Project Suggestions

1 Instructions

Partners: You may work in groups of one or two (or three if you have a good rationale). Multiple groups may work on the same topic, but the questions asked and efforts undertaken should be independent.

Topics: Listed below are possible project topics. By no means is this a complete list, and some suggested ideas have considerable overlap so can be combined. Similarly, the references provided are only a starting point. One way to proceed is to find a research article on your topic of interest, which will give you some perspective and suggest open questions and problems. Feel free to propose another topic if you would like to work on something else, e.g. topics related to your thesis work or motivated by other interests, if relevant to this class.

Types of projects: The project can involve i) researching and applying existing statistical modeling techniques to new data sets, ii) developing new statistical models or algorithms and applying them to existing data sets, or iii) providing a new analysis of existing results or extending existing analyses and analytical results.

1. Temporal vs. rate coding in retina (or any other dataset you’d like to try)
   Compare decoding under an LNP model and a Poisson GLM with spike history. How much (if any) additional information can you recover about the stimulus when you incorporate spike timing information?
   Some GLM tutorial code to get started: https://github.com/pillowlab/GLMspiketraintutorial

2. Null spaces of motor preparatory activity - implement the analyses carried out in Kaufman et al 2014 (discussed in class on 9/22). The analysis in the paper involved performing PCA to reduce neural activity to 6 dimensions and measured EMG signals to 3 dimensions, then performed least squares regression to relate neural activity to muscle activity. What if we consider more dimensions of these signals? Perform an analysis of the dimensionality of the signals in the two areas (using PCA) and examine how regression performance varies as we include more or fewer dimensions. Does the story about null spaces hold up?
   Ref: [1].
3 Project Presentations and Writeup

Formulate your project question: Write and submit a concise 1-paragraph statement of the problem together with your proposed approach. Project question due on Monday, Nov 16th (but earlier is even better).

Oral Presentation: Each group will make a 10-15 minute presentation of the final project to the class. The aim of this presentation is to explain the general motivation behind the project, including the relevant background material and methods. These presentations will take place during reading week. (Exact date to be determined).

Written summary: Each group should submit a project report with a brief description of the motivation, background, set-up, results, conclusions, and a discussion on what was learned and what the limitations of the approach were. Length: 3-5 pages.
Due date: Tuesday, Dec 8 (Dean’s date).
last week's quiz

P(y)

1/2

2

y

1/2

1

P(x,y)

P(y)

0

1/4

1/4

1/2

0

0

1/2

0

0

1 2 3

x

Bonus: what is MI between x and y?

\[ H(y) - H(y|x) = 1 \text{ bit} \]
Marginal likelihood (bonus question)

Gaussian “fun facts”

1) Linear transformations: \( x \sim N(\mu, \Sigma) \)

\[ \implies Ax \sim N(A\mu, A\Sigma A^\top) \]

2) Sums: \( x_1 \sim N(\mu_1, \Sigma_1) \quad x_2 \sim N(\mu_2, \Sigma_2) \)

\[ \implies x_1 + x_2 \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2) \]

marginal likelihood \( P(Y|X) \)

\( \tilde{w} \sim \mathcal{N}(0, C) \) (prior)

\[ \vec{Y} = X\tilde{w} + \tilde{\varepsilon}, \quad \tilde{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I) \] (encoding model)

\( X\tilde{w} \sim \) ?
Marginal likelihood (bonus question)

Gaussian “fun facts”

1) Linear transformations: \( x \sim N(\mu, \Sigma) \)
   \[ \implies Ax \sim N(A\mu, A\Sigma A^\top) \]

2) Sums: \( x_1 \sim N(\mu_1, \Sigma_1) \quad x_2 \sim N(\mu_2, \Sigma_2) \)
   \[ \implies x_1 + x_2 \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2) \]

marginal likelihood

\( \tilde{w} \sim \mathcal{N}(0, C) \)  \hspace{1cm} \text{(prior)}

\( \tilde{Y} = X\tilde{w} + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathcal{N}(0, \sigma^2 I) \)  \hspace{1cm} \text{(encoding model)}

\( X\tilde{w} \sim \mathcal{N}(0, XCX^\top) \)

\( X\tilde{w} + \tilde{\epsilon} \sim ? \)
evidence / marginal likelihood (bonus Q)

Gaussian “fun facts”

1) Linear transformations: \( x \sim N(\mu, \Sigma) \)

\[ \implies Ax \sim N(A\mu, A\Sigma A^\top) \]

2) Sums: \( x_1 \sim N(\mu_1, \Sigma_1) \quad x_2 \sim N(\mu_2, \Sigma_2) \)

\[ \implies x_1 + x_2 \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2) \]

marginal likelihood

\( \tilde{w} \sim \mathcal{N}(0, C) \) (prior)

\( \tilde{Y} = X\tilde{w} + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathcal{N}(0, \sigma^2 I) \) (encoding model)

\[ X\tilde{w} \sim \mathcal{N}(0, XX^\top) \]

\[ X\tilde{w} + \tilde{\epsilon} \sim \mathcal{N}(0, XX^\top + \sigma^2 I) \]
evidence / marginal likelihood (bonus Q)

Gaussian “fun facts”

1) Linear transformations: \[ x \sim N(\mu, \Sigma) \]
\[ \implies A x \sim N(A\mu, A\Sigma A^\top) \]

2) Sums: \[ x_1 \sim N(\mu_1, \Sigma_1) \quad x_2 \sim N(\mu_2, \Sigma_2) \]
\[ \implies x_1 + x_2 \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2) \]

marginal likelihood

\[ \tilde{w} \sim N(0, C) \quad \text{(prior)} \]
\[ \tilde{Y} = X\tilde{w} + \tilde{\varepsilon}, \quad \tilde{\varepsilon} \sim N(0, \sigma^2 I) \quad \text{(encoding model)} \]

marginal likelihood

\[ X\tilde{w} \sim N(0, XCX^\top) \]
\[ Y|X \sim N(0, XCX^\top + \sigma^2 I) \]
evidence / marginal likelihood (bonus Q)

marginal likelihood

\( Y | X \sim \mathcal{N}(0, XCXX^\top + \sigma^2 I) \)

\[
P(Y | X, \lambda) = \frac{1}{|2\pi M|^{\frac{1}{2}}} e^{-\frac{1}{2} \bar{Y}^\top M^{-1} \bar{Y}}
\]

where \( M = XCXX^\top + \sigma^2 I \)

prior

ridge \( C = \frac{1}{\lambda} I \)

smoothing \( C = \frac{1}{\lambda} L \)

Reminder: this is the denominator to Bayes’ rule:

\[
P(\bar{w} | X, Y, \lambda) = \frac{P(Y | X, \bar{w})P(\bar{w} | \lambda)}{P(Y | X, \lambda)}
\]

\[
\int P(Y | X, \bar{w})P(\bar{w} | \lambda) d\bar{w}
\]
evidence / marginal likelihood (bonus Q)

marginal likelihood

\[ Y \mid X \sim \mathcal{N}(0, XCX^\top + \sigma^2 I) \]

\[
P(Y \mid X, \lambda) = \frac{1}{|2\pi M|^{1/2}} e^{-\frac{1}{2} \tilde{Y}^\top M^{-1} \tilde{Y}}
\]

where \( M = XCX^\top + \sigma^2 I \)

prior

ridge \( C = \frac{1}{\lambda} I \)

smoothing \( C = \frac{1}{\lambda} L \)

In practice: optimize log-marginal likelihood:

\[
\log P(Y \mid X, \lambda) = \frac{1}{2} \log \det |M| - \frac{1}{2} \tilde{Y}^\top M^{-1} \tilde{Y} + \text{const}
\]

• note there is an alternate “dual form” that is computationally more efficient when \( N > d \)
Empirical Bayes estimation

• 2-step estimation procedure

1. Estimate $\lambda$ by maximizing evidence:
   (i.e. fit the prior)
   \[ \hat{\lambda} = \arg \max_{\lambda} P(Y|X, \lambda) \]

2. Find MAP estimate for weights given $\lambda$:
   \[ \hat{w}_{EB} = \arg \max_{\vec{w}} P(\vec{w} | Y, X, \hat{\lambda}) \]

(key difference: previously we used cross-validation to set $\lambda$)

advantages: • doesn’t require partitioning data
            • more practical with multiple $\lambda$