Statistical modeling and analysis of neural data (NEU 560)

Lecture 19
Thurs 11/5

Expectation Maximization

Jonathan Pillow
Fall 2020
Derive the Laplace approximation for the Bernoulli GLM with logistic link function: 
(and a flat prior)

• Model is given by:
  
  \[
P(y = 1|x, w) = p = \frac{e^{xw}}{1 + e^{xw}}
  \]
  
  \[
P(y = 0|x, w) = 1 - p = \frac{1}{1 + e^{xw}}
  \]

• Which can be written compactly as:
  
  \[
P(y|x, w) = \frac{e^{ywx}}{1 + e^{xw}}
  \]

• Likelihood is therefore:
  
  \[
P(Y|X, w) = \prod_{i=1}^{n} \frac{e^{y_i x_i w}}{1 + e^{x_i w}}
  \]
warmup problem

- Likelihood is therefore:
  \[ P(Y|X, w) = \prod_{i=1}^{n} \frac{e^{y_i x_i w}}{1 + e^{x_i w}} \]

1. compute log-likelihood
2. take 2nd derivative w.r.t. w
warmup problem

- Likelihood is therefore:

\[ P(Y|X, w) = \prod_{i=1}^{n} \frac{e^{y_i x_i w}}{1 + e^{x_i w}} \]

1. compute log-likelihood
2. take 2nd derivative w.r.t. w
3. Set covariance: \( C = -H^{-1} \)
warmup problem

• Likelihood is therefore:

\[ P(Y|X, w) = \prod_{i=1}^{n} \frac{e^{y_i x_i w}}{1 + e^{x_i w}} \]

• log-likelihood:

\[ L = \log P(Y|X, w) = \sum y_i x_i w - \sum \log(1 + e^{x_i w}) \]

\[ = w \left( \sum y_i x_i \right) - \sum \log(1 + e^{x_i w}) \]

1st derivative:

\[ \frac{d}{dw} L = \left( \sum y_i x_i \right) - \sum \frac{e^{x_i w}}{1 + e^{x_i w}} x_i \]
warmup problem

• Likelihood is therefore: \[ P(Y|X,w) = \prod_{i=1}^{n} \frac{e^{y_i x_i w}}{1 + e^{x_i w}} \]

• log-likelihood:

\[
L = \log P(Y|X,w) = \sum y_i x_i w - \sum \log(1 + e^{x_i w})
= w \left( \sum y_i x_i \right) - \sum \log(1 + e^{x_i w})
\]

1st derivative:

\[
\frac{d}{dw} L = \left( \sum y_i x_i \right) - \sum \frac{e^{x_i w}}{1 + e^{x_i w}} x_i
\]

2nd derivative:

\[
\frac{d^2}{dw^2} L = - \sum \frac{e^{x_i w}}{(1 + e^{x_i w})^2} x_i^2
\]
warmup problem

Laplace Approximation

\[ P(w|X,Y) \approx \mathcal{N}\left( \hat{w}_{map}, \frac{1}{\sum \frac{e^{x_i w}}{(1+e^{x_i w})^2} x_i^2} \right) \]

2nd derivative:

\[ \frac{d^2}{dw^2} L = - \sum \frac{e^{x_i w}}{(1 + e^{x_i w})^2} x_i^2 \]
warmup problem

Laplace Approximation

\[ P(w|X, Y) \approx \mathcal{N}\left( \hat{w}_{map}, \sum \frac{1}{(1+e^{x_i^Tw})^2} x_i^2 \right) \]

matrix version

\[ P(\bar{w}|X, Y) \approx \mathcal{N}\left( \hat{w}_{map}, \left( \sum \frac{e^{x_i^Tw}}{(1+e^{x_i^Tw})^2} x_i x_i^T \right)^{-1} \right) \]
Derivation of Laplace Approximation for Bernoulli GLM with logistic nonlinearity (aka logistic regression)

Log-likelihood:

\[ L = \log P(Y|X, \omega) = \sum_{i=1}^{n} y_i x_i \omega - \sum_{i=1}^{n} \log (1 + e^{x_i \omega}) \]

1st Deriv: \[ \frac{\partial}{\partial \omega} L = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \left( \frac{e^{x_i \omega}}{1 + e^{x_i \omega}} \right) x_i \]

2nd Deriv:

\[ \frac{\partial^2}{\partial \omega^2} L = -\sum_{i=1}^{n} \left( \frac{(1 + e^{x_i \omega}) x_i - e^{x_i \omega} (1 + e^{x_i \omega}) x_i}{(1 + e^{x_i \omega})^2} \right) \]

\[ = -\sum_{i=1}^{n} \left( \frac{e^{x_i \omega} + 2 x_i \omega - 2 x_i \omega}{(1 + e^{x_i \omega})^2} \right) \]

\[ = -\sum_{i=1}^{n} \left( \frac{e^{x_i \omega}}{(1 + e^{x_i \omega})^2} \right) x_i^2 \]

Thus: \[ p(\omega|X,Y) \approx N(\hat{\omega}_{\text{MAP}}, \left( \sum_{i=1}^{n} \frac{e^{x_i \omega}}{(1 + e^{x_i \omega})^2} \right)^{-1}) \]