Modelling game outcomes

Two player game → win/lose (no draws)

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
<th>Winner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fred</td>
<td>Nina</td>
<td>Nina</td>
</tr>
<tr>
<td>Jo</td>
<td>Nina</td>
<td>Jo</td>
</tr>
<tr>
<td>Nina</td>
<td>Nina</td>
<td>Nina</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Fred</td>
<td>Nina</td>
<td>Fred</td>
</tr>
</tbody>
</table>

How might we predict winner given players?
Can we rank players, best → worst?
Think of simple approaches first.
Each player "skill" parameter

\[ p(y=1 \mid x, w) = \sigma (w_{\text{fred}} + w_{\text{Nina}}^2) \]

player 1 wins (fred)

"skill fred"

"Badness
Nina"

\[ \sigma (w_{\text{fred}} - w_{\text{Nina}} + b) \]
Approx. Normalizer $p(D|M)$

$$p(w|D) = \frac{p(w, D)}{p(D)} \propto N(w; w^*, H^{-1})$$

$$= \frac{1H^{1/2}}{(2\pi)^{D/2}} e^{-\frac{1}{2}(w-w^*)' H (w-w^*)}$$

Evaluate approx. at $w = w^*$:

$$\frac{p(w^*, D)}{p(D)} \propto \frac{1H^{1/2}}{(2\pi)^{D/2}}$$

# parameters

$P(D) = \frac{p(w^*, D)}{p(w^*|D)} \approx \frac{P(w^*, D)}{1H^{1/2}}$

$N(w; w^*, 1/H)$

$P(D)$ from Laplace approximation:

A) Too Big; B) Too Small; C) Correct; E) ???

L27 (3)

2019 L26 (4)
\[ p(D) = \int p(D, w) \, dw \]
\[ = \int p(D|w) p(w) \, dw \]
Bayesian Logistic Regression

Decision boundaries:
\[ p(y=1|\mathbf{x}, \mathbf{w}) = \sigma(\mathbf{w}^T \mathbf{x}) \]
for different plausible \( \mathbf{w} \)

\[ \sigma(\hat{\mathbf{w}}^T \mathbf{x}) = 0.9 \]

\[ \sigma(\hat{\mathbf{w}}^T \mathbf{x}) = 0.5 \]

Predictive distribution
\[ \sigma(\hat{\mathbf{w}}^T \mathbf{x}) = 0.1 \]
for a single fitted \( \hat{\mathbf{w}} \), eg L2 regularised fit

\[ p(y=1|\mathbf{x}, D) = 0.9 \]

\[ p(y=1|\mathbf{x}, D) = 1/2 \]

Training data
Plausible weights described by Posterior

\[ p(w | D, M) = \frac{P(D | w, M) p(w | M)}{P(D | M)} \]

\[ = \int p(D | w, M) p(w | M) \, dw \]

Train \( E \{ x^n, y^n \} \)

\[ \int p(D | M) \, dw \]

M, model choices, hyperparameters, basis fns

(Often miss out)

Likelihood

\[ P(D | w) = \prod_{n} p(x^{(n)} | w) p(y^{(n)} | w, x^{(n)}) \]

Prior: \( p(w) = N(w; 0, \sigma_w^2 \mathbb{I}) \)

Marginal Likelihood \( p(D) \) or \( p(D | M) \) can compare models

Predictions

\[ p(y=1 | x, D) = \int p(y=1 | w, x, D) \, dw \]

\[ = \int p(y=1 | w, x, D) p(w | x, D) \, dw \]

\[ \sigma(w^T x) \] posterior
Predictions for logistic regression

\[ p(y=1 \mid x, D) = \int \sigma(w^T x) N(w; w^*, H^{-1}) dw \]

Laplace approx.

\[ = \mathbb{E}_{N(w; w^*, H^{-1})} \left[ \sigma(w^T x) \right] \]

(Could do Monte Carlo)

Average under "activation" \( w^T x = a \)

\[ = \mathbb{E}_{p(a)} \left[ \sigma(a) \right] \]

\[ N(a; w^*^T x, x^T H^{-1} x) \]

Could solve numerically

\[ \approx \sigma \left( x^T w^* \right) \]

\[ \left( \begin{array}{c} 1 \\ \sqrt{1 + \frac{1}{8} x^T H^{-1} x} \end{array} \right) \]

Homework: check!

Bishop p220
Murphy
§8.4.4.2

2019 L27 3
Laplace Approximation

\[ \log p(w, D) \]

Hessian, \( H_{ij} = \frac{\partial^2 \log p(w, D)}{\partial w_i \partial w_j} \)

Quadratic fit, tiny curvature

Posterior

\[ p(w \mid D) \approx N(w; w^*, H^{-1}) \]

Marginal Likelihood

\[ p(D) = \frac{p(w^*, D)}{p(w^* \mid D)} \approx \frac{p(w^*, D)}{N(w^*; w^*, H^{-1})} \]

Areas under both curves = 1

**Quiz**

Is \( p(D) \) approx.?

A) Too big
B) Too small
C) \( \propto \) Correct
D) ???

2019 L27 ⑧