Bayesian Linear Regression

Some different plausible lines \((w_1 x + w_2 x^2)\)

Prior (Example) \(p(w) = N(w; 0, \sigma w^2 I)\)

\(\Rightarrow\) Broad range functions plausible before seeing data

Posterior \(\sim\) likelihood

\[ p(w | D) \propto p(w) p(y | x, w) \]

\[ = N(w; w_N, V_N) \]

For Gaussian prior and noise

Negative slope exceedingly implausible

\(w_N\), slope

Negative intercept possible.
Probabilistic Prediction

\[ f(x) = w^T x = x^T w \]

\[ p(f(x) | Data) = N(f; w_N^T x, x^T V_N x) \]

\[ p(y | Data) = N(y; \mu, \sigma_y^2) \]

Questions

Uncertainty \( x^T V_N x \) grows with \( x \)

1. Why in figure is most certain region at \( x > 0 \) (around \( x = 3 \))?

2. What do contours of \( x^T V_N x \) look like?

A

\[ \begin{array}{c}
\begin{array}{c}
\cdots
\end{array}
\end{array} \]

B

\[ \begin{array}{c}
\begin{array}{c}
\cdots
\end{array}
\end{array} \]

C

\[ \begin{array}{c}
\begin{array}{c}
\cdots
\end{array}
\end{array} \]

D Other

\[ ??? \]
$V_N$ was posterior covariance of $w$

Contours $N(w; w_N, V_N)$

$-\frac{1}{2} (w - w_N)^T V_N^{-1} (w - w_N)$
Decisions - Loss function

\[ L(y, \hat{y}) \]

\[ \uparrow \quad \text{"point estimate"} / \quad \text{"guess"} \]

Loss what happens

Minimize expected loss

\[ c = \int L(y, \hat{y})p(y | \text{Data}) dy \]

\[ = \mathbb{E}_{p(y | \text{Data})} [L(y, \hat{y})] \]

Find \( \hat{y} \) that minimizes \( c \)

E.g. square loss \( L(y, \hat{y}) = (y - \hat{y})^2 \)

\[ \frac{dc}{d\hat{y}} = \mathbb{E} [2(y - \hat{y})] \]

\[ = 0 \text{ if } \mathbb{E}[y] = \hat{y} \]

Estimate \( \hat{y} = \text{mean belief} \).
Overfitting

Bayesian don't fit

$\hat{w} = \arg\min_w \text{cost}(w)$
do n't do

Compute beliefs $p(w|D)$ ... decision

"Underfitting"

Over simple models

$\Rightarrow$ Over-confident

Residuals

$\Rightarrow$ Tell us things are wrong

"Model checking"
Bayesian methods with lots of parameters

\[ \phi_1, \phi_2, \phi_3, \phi_4, \phi_5 \rightarrow \]

Observe data \downarrow

Samples from prior

Sample from posterior

Extreme flexible model:

\[ \ldots \]

\[ 10^6 \text{ basis functions} \]

Can model:

\[ f(x) \uparrow \]

\[ 10^{-6} \]
If prior \( p(w_k) = N(w_k; 0, \sigma^2_w) \)

Independent \( p(w) = N(w; 0, \sigma^2_w) \)

What's the posterior?
Probabilistic model choice

Bayes classifiers

Model for class 1
\[ p(x \mid M_1) \]

Contours of \[ p(x \mid M_2) \]

Regression

\[ p(y \mid X, M) = \int p(y, w \mid X, M) \, dw \]

\[ = \int p(y \mid X, w, M) \, p(w \mid X, M) \, dw \]

Marginal

Likelihood of Model, scores how good model is.