Lecture 9: Uncertainty Quantification (II)

Soufiane Hayou

Last week

Bayesian approach

Model true parameters as random

$$p(\beta|S) = \frac{p(S|\beta)p_0(\beta)}{\int_\beta p(S|\beta)p_0(\beta)}$$

Model true parameters as random

$$p(\beta|S) = \frac{p(S|\beta)p_0(\beta)}{\int_\beta p(S|\beta)p_0(\beta)}$$

- Difference: β is not a fixed number, but a distribution $p(\beta|S)$
- $p_0(\beta)$: prior, $p(S|\beta)$ likelihood of getting S

Model true parameters as random

$$p(\beta|S) = \frac{p(S|\beta)p_0(\beta)}{\int_\beta p(S|\beta)p_0(\beta)}$$

- Difference: β is not a fixed number, but a distribution $p(\beta|S)$
- $p_0(\beta)$: prior, $p(S|\beta)$ likelihood of getting S

Advantage

- Easy to implement prior knowledge
- Robust against overfitting
- Give information on alternative possibilities

Bayesian perspective

Model true parameters as random

$$p(\beta|S) = \frac{p(S|\beta)p_0(\beta)}{\int_\beta p(S|\beta)p_0(\beta)}$$

- Difference: β is not a fixed number, but a distribution $p(\beta|S)$
- $p_0(\beta)$: prior, $p(S|\beta)$ likelihood of getting S

Advantage

- Easy to implement prior knowledge
- Robust against overfitting
- Give information on alternative possibilities

Disadvantage

- Expensive to compute.
- Sensitive to prior choices.

Gaussian family

- Prior $\beta \sim \mathcal{N}(\mu_0, \Sigma_0)$
- Data model is linear

$$y_i = \beta^T z_i + \epsilon_i$$

- Likelihood

$$p(\mathbf{y}|\beta) = \mathcal{N}(Z\beta, \Sigma_\epsilon)$$

Gaussian family

- Prior $\beta \sim \mathcal{N}(\mu_0, \Sigma_0)$
- Data model is linear

$$y_i = \beta^T z_i + \epsilon_i$$

- Likelihood

$$p(\mathbf{y}|\beta) = \mathcal{N}(Z\beta, \Sigma_\epsilon)$$

- The posterior is Gaussian with mean

$$\mu_p = (\Sigma_0^{-1} + Z^T \Sigma_\epsilon^{-1} Z)^{-1} (Z^T \Sigma_\epsilon^{-1} \mathbf{y} + \Sigma_0^{-1} \mu_0)$$

- And covariance

$$\Sigma_p = (\Sigma_0^{-1} + Z^T \Sigma_\epsilon^{-1} Z)^{-1}.$$

Example (Simple Example)

Suppose we want to estimate a $\beta \in R^1$

$$y_i = \beta + \epsilon_i$$

We use prior $\beta \sim \mathcal{N}(0, 1)$. Find the posterior mean and covariance of β with n data points.

$\mathcal{N}(\frac{\sum y_i}{n+1}, \frac{1}{n+1})$.

Choice of Prior is very important!

The prior can have a huge impact on the posterior!!

(found on twitter)

I was surprised when my daughter said she learned R at school yesterday, and then I remembered that she's 4 and she meant the letter.

My priors are all too skewed

Monte Carlo

Monte Carlo average

How to compute

$$\mathbb{E}[f(\beta)|S] = \int f(\beta)p(\beta|S)d\beta?$$

Monte Carlo average

How to compute

$$\mathbb{E}[f(\beta)|S] = \int f(\beta)p(\beta|S)d\beta?$$

Generate samples $\beta^i \sim p(\beta|S)$, use

$$\hat{f}_m = \frac{1}{m}\sum_{i=1}^{m} f(\beta^i)$$

- $\hat{f}_m$ is an unbiased estimator
- If β^i are i.i.d. from p, then

$$\mathrm{var}(\hat{f}_m) = \frac{1}{m}\mathrm{var}(f)$$

- How to access the uncertainty of $f(\beta)$?

Sample covariance

- How to access the uncertainty of $f(\beta)$?
- Find the posterior variance

$$\mathrm{var}[f(\beta)|S] = \int (f(\beta) - \mathbb{E}[f(\beta)|S])^2 p(\beta|S) d\beta$$

- Sample posterior variance

$$\frac{1}{m}\sum(\hat{f}_m - f(\beta^i))^2$$

Consider a general linear model

$$y_i = \beta^T h(x_i) + \eta_i, \quad \eta \sim \mathcal{N}(0, \sigma I)$$

- Suppose $\beta \sim \mathcal{N}(0, \tau I)$

Applying to linear regression

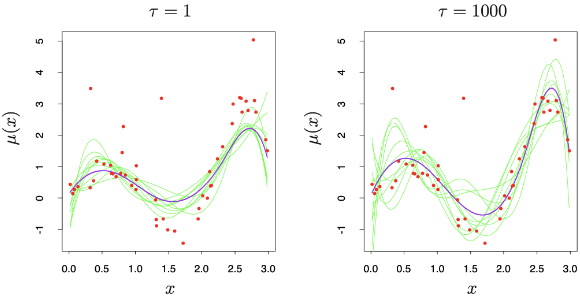
Consider a general linear model

$$y_i = \beta^T h(x_i) + \eta_i, \quad \eta \sim \mathcal{N}(0, \sigma I)$$

- Suppose $\beta \sim \mathcal{N}(0, \tau I)$
- The posterior mean

$$\hat{w}_{\text{MAP}} = (\sigma\tau^{-1} I + H^T H)^{-1} H^T \mathbf{y}.$$

- This is also the ridge regression result.



A larger τ :

- Less informative prior
- More data fitting, less structure
- Less smoothness