CGS698C, Assignment 05

Himanshu Yadav

Part 1: Information-theoretic measures and cross-validation

You are given 10 independent and identically distributed data points that are assumed to come from a Binomial distribution with sample size 20 and probability of success θ :

10, 15, 15, 14, 14, 14, 13, 11, 12, 16

Suppose that you build two models differing in prior knowledge about the θ parameter. Model 1 has Beta(6,6) prior for θ and model 2 has $Beta(20, 60)$ prior on θ.

Let y_i be i^{th} data point.

Model 1:

$y_i \sim Binomial(n = 20, \theta)$

$\theta \sim Beta(6, 6)$

Model 2:

$y_i \sim Binomial(n = 20, \theta)$

$\theta \sim Beta(20, 60)$

Exercise 1.1 Graph the posterior distribution of θ for each model

Exercise 1.2 Compute log pointwise predictive density (lppd) for each model

(Hint: Draw samples from the posterior distribution $\hat{p}(\theta|y)$, calculate the log predictive density for each data point y_i averaged over all samples from the posterior.

$lpd_i = \log \frac{1}{N} \sum_{j=1}^{N} p(y_i|\theta_j)$ where $\theta_j \sim \hat{p}(\theta|y)$

After you have collected log predictive density lpd_i for each datapoint, add up all the lpd_i to obtain the log pointwise predictive density $lppd$ for the model.

See example code on pages 18–20.)

Exercise 1.3 Calculate in-sample deviance for each model from the log pointwise predictive density (lppd) computed in 3.2. Use the following formula:

In-sample deviance = -2*lppd

Why are we calling this in-sample deviance?

Exercise 1.4 Based on in-sample deviance, which model is a better fit to the data?

Exercise 1.5 Suppose that you have 5 new data points: [5, 6, 10, 8, 9]. Which of your models is better at predicting new data? You can calculate out-of-sample deviance now to compare your models.

(Hint: Compute log predictive densities for each new data point; compute lppd and out-of-sample deviance, i.e., -2*lppd).

Exercise 1.6 Now suppose you do not have new data. Perform leave-one-out cross-validation (LOO-CV) to compare model 1 and model 2

(Hint: You have to again compute lppd, but this time fit the model on 9 datapoints and calculate log predictive density on remaining 1 datapoint, repeat this process 10 times such that you leave out all datapoints one by one. See example code on page 18.)

1 *Part 2: Marginal likelihood and prior sensitivity*

Consider a Binomial model with sample size n and probability of success θ and prior on θ is Beta(a,b).

The marginal likelihood of the model for k successes will be:
$\binom{n}{k}\frac{(k+a-1)!(n-k+b-1)!}{(n+a+b-1)!}$
You can use the following function to calculate the same.

```
ML_binomial <- function(k,n,a,b){
  ML <- (factorial(n)/(factorial(k)*factorial(n-k)))*(
    factorial(k+a-1)*factorial(n-k+b-1)/factorial(n+a+b-1))
  ML
}
```

Exercise 2.1 For k=2 and n=10, calculate marginal likelihood of the models having following priors on θ:

- Beta(0.1,0.4)
- Beta(1,1)
- Beta(2,6)
- Beta(6,2)
- Beta(20,60)
- Beta(60,20)

Exercise 2.2 Estimate the marginal likelihood of the model given the above prior assumptions using Monte Carlo Integration method.