BDA - Assignment 9

Anonymous

Contents

Decision analysis for the factory data (3p)

```r
library(aaltobda)
data("factory")
library(rstan)
```

```
## Loading required package: StanHeaders

## Loading required package: ggplot2

## rstan (Version 2.21.2, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
```

```r
set.seed(123)
library(markmyassignment)
assignment_path <- paste("https://github.com/avehtari/BDA_course_Aalto/",
                         "blob/master/assignments/tests/assignment9.yml",
                         sep="")
set_assignment(assignment_path)
```

```
## Assignment set:
## assignment9: Bayesian Data Analysis: Assignment 9
## The assignment contain the following task:
## - utility
```

As noticed in the previous assignment, the hierarchical model fits best with the dataset, so use it to compute the utilities.

1. For each of the six machines, compute and report the expected utility of one product of that machine.

Hierarchical model in stan

```
data {
  int < lower =0 > N; // number of measurements
```

```
  int < lower =0 > J; // number of machines
  vector[J] y[N];
}
parameters {
  real mu; // hyper-parameter 1
  real<lower=0> tau; // hyper-parameter 2
  vector[J+1] theta; // separate mean parameter theta for each 7 machines
  real<lower=0> sigma; // common sigma parameter for all7 machines
}
model {
  // Weakly informative priors
  mu ~ normal (0, 100); // hyperprior for mu
  tau ~ inv_chi_square(0.1); // hyperprior for tau
  for ( j in 1: (J+1) ){ // Adding the 7th machine
    theta [j] ~ normal (mu, tau);
  }
  sigma ~ inv_chi_square(0.1);
  // likelihood
  for ( j in 1: J )
    y[ ,j ] ~ normal (theta[j], sigma);
}
generated quantities {
  vector[J+1] ypred ;
  for (j in 1:(J+1)) // Adding the 7th machine
    ypred[j] = normal_rng(theta[j], sigma);
}
```

Utility function

```
utility <- function(draws){
  util <- length(draws[draws >= 85])*(94) + length(draws[draws < 85]) * (-106)
  return(util/length(draws))
}

mark_my_assignment()
```

```
## v |  OK F W S | Context
##
/ |   0       | utility()
v |   5       | utility()
##
## == Results ===============================
## OK:       5
## Failed:   0
## Warnings: 0
## Skipped:  0
## Good work!
```

The expected utilities for one product of each machine

```
factory_data <- list(y = factory, N = nrow(factory), J = ncol(factory))
fit <- rstan::sampling(hierarchical_model, data = factory_data, refresh=0)
```

```
## Warning: There were 34 divergent transitions after warmup. See
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
## to find out why this is a problem and how to eliminate them.

## Warning: Examine the pairs() plot to diagnose sampling problems

## Warning: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and tail quant
## Running the chains for more iterations may help. See
## http://mc-stan.org/misc/warnings.html#tail-ess
```

```r
monitor(fit)
```

```
## Inference for the input samples (4 chains: each with iter = 2000; warmup = 0):
##
##                 Q5    Q50    Q95   Mean   SD  Rhat Bulk_ESS Tail_ESS
## mu            82.3   92.6  102.7   92.6  6.5  1.00     2334     1846
## tau            4.1   11.5   24.9   12.7  7.1  1.01      479      226
## theta[1]      69.8   81.2   92.1   81.1  6.9  1.00     1143     1175
## theta[2]      91.8  102.1  112.3  102.1  6.3  1.00     1581     2310
## theta[3]      79.2   89.3   99.1   89.3  6.0  1.00     2685     2571
## theta[4]      94.4  106.1  117.1  106.0  6.9  1.00     1134     1124
## theta[5]      80.5   90.9  100.7   90.7  6.1  1.00     2709     2597
## theta[6]      78.1   88.0   97.7   88.0  6.0  1.00     2088     2432
## theta[7]      68.3   92.6  118.9   93.1 16.4  1.00     2333     1744
## sigma         11.8   14.9   19.3   15.1  2.3  1.00     1923     2671
## ypred[1]      53.7   80.8  108.8   81.1 16.8  1.00     3346     4016
## ypred[2]      75.0  101.4  129.2  101.6 16.8  1.00     3795     3918
## ypred[3]      62.0   89.1  116.4   89.0 16.2  1.00     4233     3739
## ypred[4]      77.9  106.2  133.2  105.9 16.9  1.00     2425     2842
## ypred[5]      65.0   91.1  117.9   91.0 16.5  1.00     3667     3818
## ypred[6]      60.3   88.3  115.7   88.2 16.8  1.00     3874     3608
## ypred[7]      57.7   92.8  128.5   92.8 22.3  1.00     2751     2261
## lp__        -122.7 -116.9 -113.0 -117.2  3.0  1.01      560      206
##
## For each parameter, Bulk_ESS and Tail_ESS are crude measures of
## effective sample size for bulk and tail quantities respectively (an ESS > 100
## per chain is considered good), and Rhat is the potential scale reduction
## factor on rank normalized split chains (at convergence, Rhat <= 1.05).
```

```r
df <- as.data.frame(fit)

paste("The expeted utilities are...")
```

```
## [1] "The expeted utilities are..."
```

```r
paste("Machine 1:", utility(df$'ypred[1]'))
```

```
## [1] "Machine 1: -26.45"
```

```r
paste("Machine 2:", utility(df$'ypred[2]'))
```

```
## [1] "Machine 2: 62.65"
```

```r
paste("Machine 3:", utility(df$'ypred[3]'))
```

```
## [1] "Machine 3: 14.05"
```

```r
paste("Machine 4:", utility(df$'ypred[4]'))
```

```
## [1] "Machine 4: 73"
```

```r
paste("Machine 5:", utility(df$'ypred[5]'))
```

```
## [1] "Machine 5: 22.35"
```

```r
paste("Machine 6:", utility(df$'ypred[6]'))
```

```
## [1] "Machine 6: 10.55"
```

2. Rank the machines based on the expected utilities (from worst to best). Also briefly explain what the utility values tell about the quality of these machines. E.g. Tell which machines are profitable and which are not.

Ranking from worst to best

1, 6, 3, 5, 2, 4

Discussion about utility values

The expected utility value tells how much each finished product by the machine is expected to make. If the value is negative, the product produced by the machine is expected to be unprofitable for the company.

Based on the calculated expected utilities with our hierarchical model, all the machines except the machine 1 are profitable for the factory and each finished product by these machines can be expected to have sufficient quality.

Machine 1 on the other hand, is expected to produce insufficient quality products and isn't profitable for the company in the long run.

3. Compute and report the expected utility of the products of a new (7th) machine.

```r
paste("The expeted utility for new (7th) machine is", utility(df$'ypred[7]'))
```

```
## [1] "The expeted utility for new (7th) machine is 23.25"
```

4. Based on your analysis, discuss briefly whether the company owner should buy a new (7th) machine.

Based on the calculated ecpected utility for the new machine and knowing that the factory owner only cares about money (sad but i suppose this is kind of realistic), I would recommend him to invest into this new machine. Based on the analysis above the machine would make profit for the company in the long run.