hw10 for stat341

Zhihong Zhang

March 17th,2017

Q: 6M3 6M4 6H1 6H4 6H5

6M3. When comparing models with an information criterias, why must all models be fit to exactly the same observations? What would happen to the information criterion values, if the models were fit to different numbers of observations? Perform some experiments, if you are not sure.

Solutions:

Because all of these information criterions are used to calculate the estimates of the deviations. As we expected, all models must be fit to exactly the same observations.

6M4. What happens to the effective number of parameters, as measured by DIC or WAIC, as a prior becomes more concentrated? Why? Perform some experiments, if you are not sure.

Solutions:

When prior becomes more concentrated, the estimates are shrunk to the zero. Because for a concentrated prior, they are more conservatives.

6H1. Compare the models above, using WAIC. Compare the model rankings, as well as the WAIC weights.

Solutions:

```
data(Howell1)
Howell <-
Howell1 %>% mutate(age.s = zscore(age))
set.seed(1000) # so we all get the same "random" data sets
train <- sample(1:nrow(Howell), size = nrow(Howell) / 2) # half of the rows
Howell.train <- Howell[ train, ] # put half in training set
Howell.test <- Howell[-train, ] # the other half in test set

a.start <- mean(Howell.train$age.s)
sigma.start <- log(sd(Howell.train$age.s))


#modelM1
m1 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s
  ),
  data = Howell.train ,
  start = list(a = a.start, b1 = 0, log.sigma = sigma.start)
)

#modelM2

m2 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2
```

```
  ),
  data = Howell.train ,
  start = list(a = a.start, b1 = 0,b2=0, log.sigma = sigma.start)
)


#modelM3

m3 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3
  ),
  data = Howell.train ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0, log.sigma = sigma.start)
)


#modelM4

m4 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3+b4*age.s^4
  ),
  data = Howell.train ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0,b4=0, log.sigma = sigma.start)
)


#modelM5

m5 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3+b4*age.s^4+b5*age.s^5
  ),
  data = Howell.train ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0,b4=0,b5=0, log.sigma = sigma.start)
)

#modelM6

m6 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3+b4*age.s^4+b5*age.s^5+b6*age.s^6
  ),
  data = Howell.train ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0,b4=0,b5=0,b6=0, log.sigma = sigma.start)
)

compare(m1,m2,m3,m4,m5,m6)
```

```
##        WAIC pWAIC dWAIC weight    SE    dSE
## m4 1926.1   5.7   0.0   0.56 25.50    NA
## m5 1927.5   6.4   1.3   0.28 25.28  1.19
## m6 1928.6   7.7   2.5   0.16 25.07  2.95
## m3 1952.3   5.4  26.1   0.00 24.14 10.89
## m2 2149.9   5.2 223.8   0.00 22.54 26.82
## m1 2395.4   3.4 469.3   0.00 22.90 31.14
```

Using WAIC, based on the data, m1 has the highest WAIC value. m4 has weight of 0.56 with lowest WAIC.

6H4. Compute the test-sample deviance for each model. This means calculating deviance, but using the data in d2 now. You can compute the log-likelihood of the height data with: R code sum(dnorm(d2$height , mu , sigma , log=TRUE)) 6.32 where mu is a vector of predicted means (based upon age values and MAP parameters) and sigma is the MAP standard deviation.

```r
data(Howell1)
Howell <-
Howell1 %>% mutate(age.s = zscore(age))
set.seed(1000) # so we all get the same "random" data sets
train <- sample(1:nrow(Howell), size = nrow(Howell) / 2) # half of the rows
Howell.train <- Howell[ train, ] # put half in training set
Howell.test <- Howell[-train, ] # the other half in test set

a.start <- mean(Howell.test$age.s)
sigma.start <- log(sd(Howell.test$age.s))


#modelM1
m1 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s
  ),
  data = Howell.test ,
  start = list(a = a.start, b1 = 0, log.sigma = sigma.start)
)

#modelM2

m2 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2
  ),
  data = Howell.test ,
  start = list(a = a.start, b1 = 0,b2=0, log.sigma = sigma.start)
)


#modelM3

m3 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3
```

```
  ),
  data = Howell.test ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0, log.sigma = sigma.start)
)


#modelM4

m4 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3+b4*age.s^4
  ),
  data = Howell.test ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0,b4=0, log.sigma = sigma.start)
)


#modelM5

m5 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3+b4*age.s^4+b5*age.s^5
  ),
  data = Howell.test ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0,b4=0,b5=0, log.sigma = sigma.start)
)

#modelM6

m6 <- map(
  alist(
    height ~ dnorm(mu, exp(log.sigma)),
    mu <- a + b1* age.s+b2* age.s^2+b3*age.s^3+b4*age.s^4+b5*age.s^5+b6*age.s^6
  ),
  data = Howell.test ,
  start = list(a = a.start, b1 = 0,b2=0,b3=0,b4=0,b5=0,b6=0, log.sigma = sigma.start)
)

#deviance computing
dev1<-(-2) * logLik(m1)
dev2<-(-2) * logLik(m2)
dev3<-(-2) * logLik(m3)
dev4<-(-2) * logLik(m4)
dev5<-(-2) * logLik(m5)
dev6<-(-2) * logLik(m6)

dev1

## 'log Lik.' 2420.02 (df=3)

dev2

## 'log Lik.' 2136.227 (df=4)
```

```
dev3
```

```
## 'log Lik.' 1919.874 (df=5)
```

```
dev4
```

```
## 'log Lik.' 1867.126 (df=6)
```

```
dev5
```

```
## 'log Lik.' 1863.243 (df=7)
```

```
dev6
```

```
## 'log Lik.' 1860.608 (df=8)
```

6H5. Compare the deviances from 6H4 to the WAIC values. It might be easier to compare if you subtract the smallest value in each list from the others. For example, subtract the minimum WAIC from all of the WAIC values so that the best WAIC is normalized to zero. Which model makes the best out-of-sample predictions in this case? Does WAIC do a good job of estimating the test deviance?

Solutions:

Compare the deviances from to the WAIC values,m1 is the best models.

```
dev1-WAIC(m1)
```

```
## Constructing posterior predictions
```

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```

```
## 'log Lik.' -6.455895 (df=3)
```

```
dev2-WAIC(m2)
```

```
## Constructing posterior predictions
```

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```

```
## 'log Lik.' -10.76693 (df=4)
```

```
dev3-WAIC(m3)
```

```
## Constructing posterior predictions
```

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```

'log Lik.' -11.16484 (df=5)

dev4-WAIC(m4)

Constructing posterior predictions

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```

'log Lik.' -12.63348 (df=6)

dev5-WAIC(m5)

Constructing posterior predictions

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```

'log Lik.' -14.16601 (df=7)

dev6-WAIC(m6)

Constructing posterior predictions

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
```

```
[ 900 / 1000 ]
[ 1000 / 1000 ]
```

'log Lik.' -16.74557 (df=8)