hw11 for stat341

Zhihong Zhang

March 29th,2017

Q: 6M2 6M5 6M6 6H2 6H3

6M2. Explain the difference between model selection and model averaging. What information is lost under model selection? What information is lost under model averaging?

Solutions: difference between model selection and model averaging

model selection means choosing the model with the lowest AIC/DIC/WAIC value and then discarding the others.

Model averaging means using DIC/WAIC to construct a posterior predictive distribution that exploits what we know about relative accuracy of the models.

relative model accuracy contained in the differences among the AIC/DIC/WAIC values is lost under model selection.

information about deviance is lost under model averaging.

6M5. Provide an informal explanation of why informative priors reduce overfitting.

Solutions:

Because the informative prior expresses specific, definite information about a variable. And this information helps on posterior which improves overfitting.

6M6. Provide an information explanation of why overly informative priors result in underfitting.

Solutions:

Because overly informative priors gives too much detailed on prediction of model and it did not do well on training data.

6H2. For each model, produce a plot with model averaged mean and 97% confidence interval of the mean, superimposed on the raw data. How do predictions differ across models?

Solutions:

```
data(Howell1)
# standardize the age before fitting
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(sigma)),
        mu <- a + b1* age.s),
```

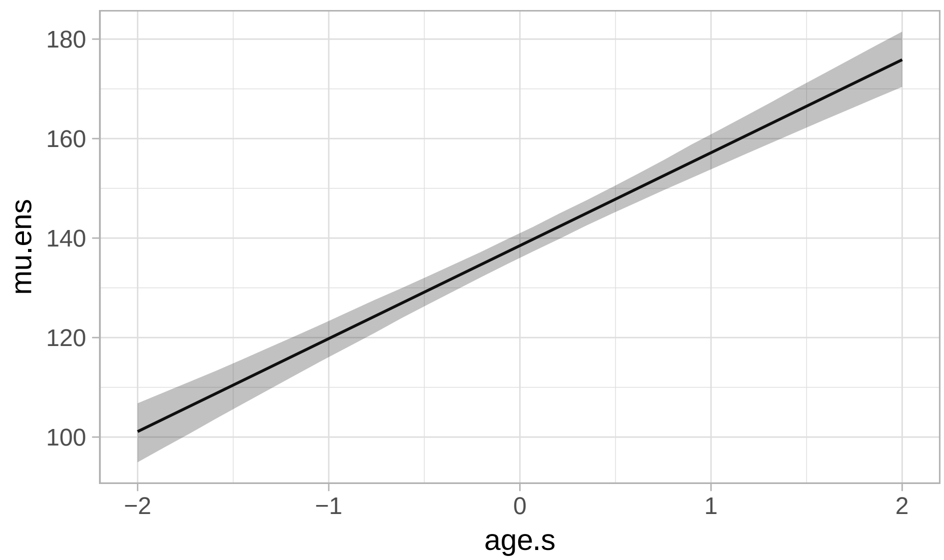
```r
  data = Howell.train ,
  start = list(
    a = mean(Howell.train$height),
    b1 = 0,
    sigma = sd(Howell.train$height)
  ),
  method = "Nelder-Mead"
)



# extract MAP estimates
theta <- coef(m1); theta
```

```
##          a         b1      sigma
## 138.451551  18.663040   2.972433
```

```r
#modelM2


model.predict <- data_frame(
  age.s = seq(from = -2, to = 2, length.out = 30),
height = 0
   # average height
)

model1.link <- link(m1,data = model.predict)
```

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

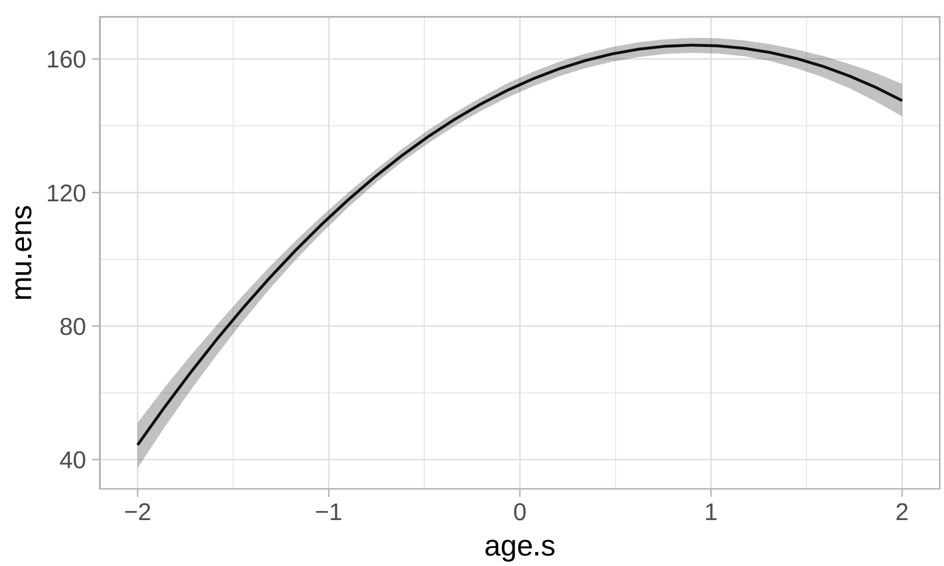
```r
model1.predict <-
  model.predict %>%
  mutate(
    mu.ens =  apply(model1.link, 2, mean),
    mu.ens.lo =  apply(model1.link, 2, PI,prob=0.97)[1,],
    mu.ens.hi =  apply(model1.link, 2, PI,prob=0.97)[2,]
  )
#for  model 1
gf_line(mu.ens ~ age.s, data = model1.predict) %>%
  gf_ribbon(mu.ens.lo + mu.ens.hi ~ age.s, data =model1.predict)
```



Similary, for other models, using the same way to plot others

For Model 2

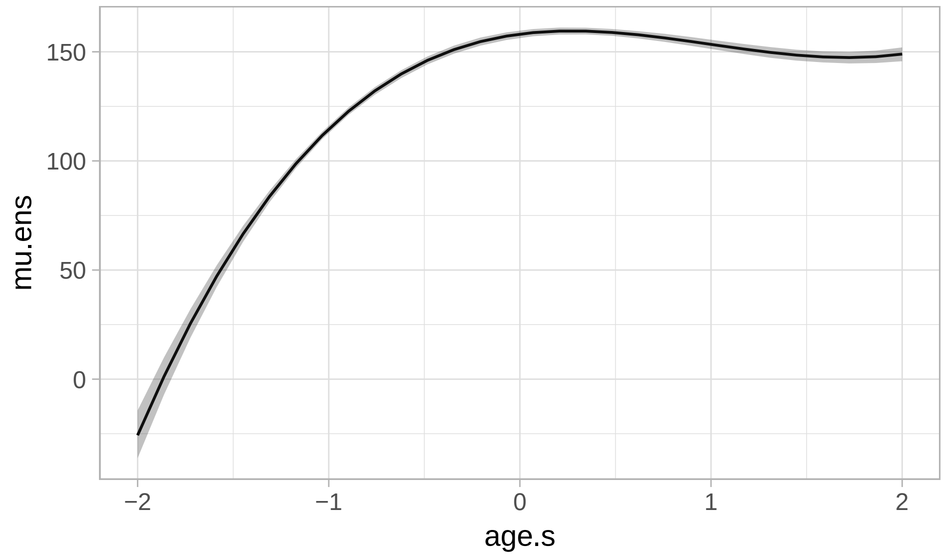
```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```



for model 3

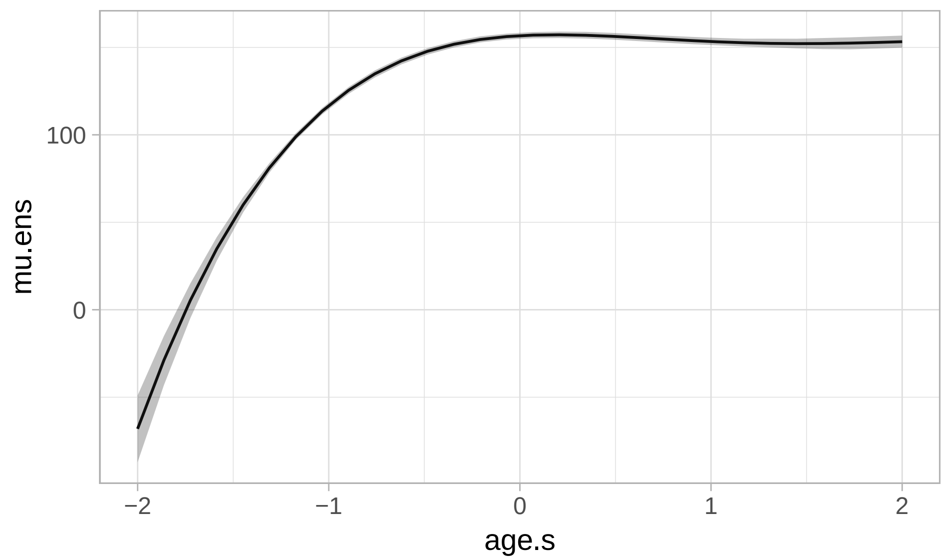
```
## [ 100 / 1000 ]
```

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



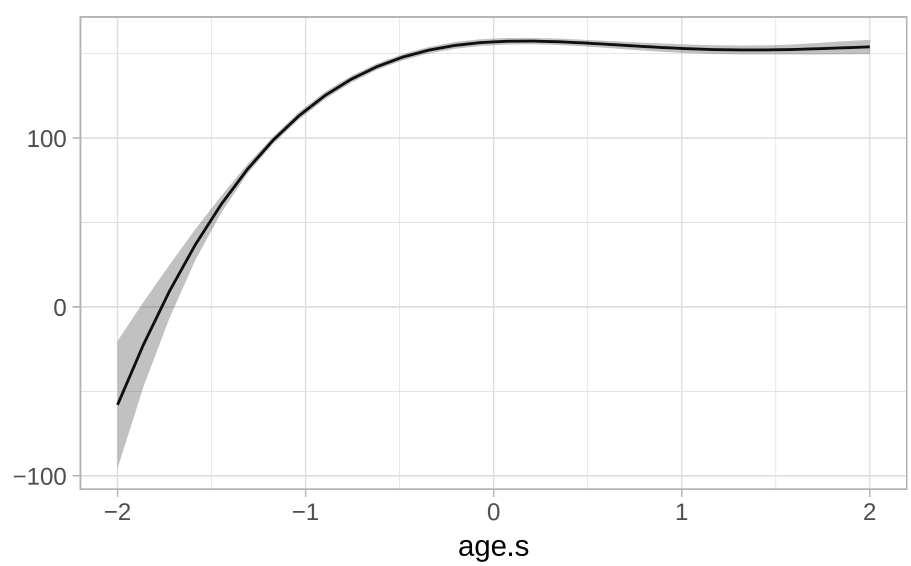
for model 4

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



for model 5

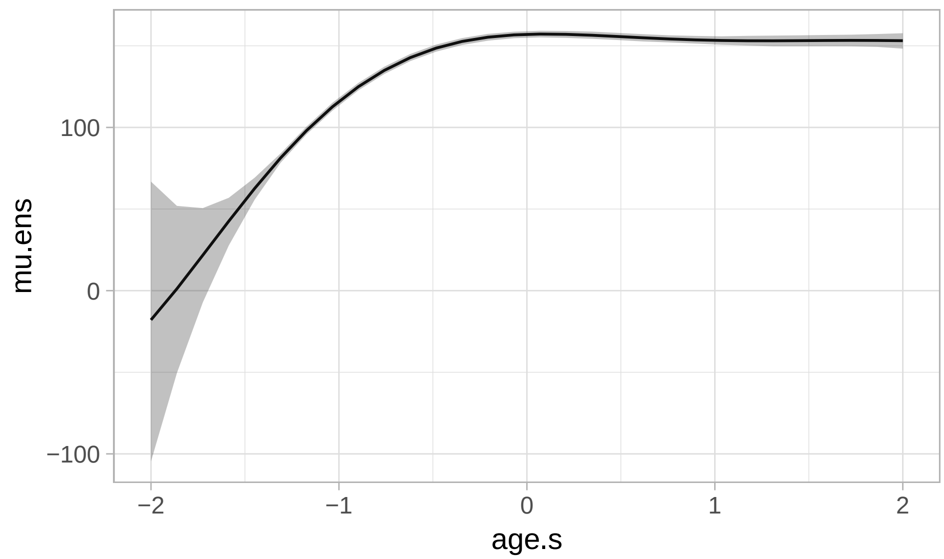
```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```



for model 6

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
```

```
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```



Compared all 6 models, the model gets uncertain on -2 to 0 of age.s range as the degree of the models increased.

6H3. Now also plot the model averaged predictions, across all models. In what ways do the averaged predictions differ from the predictions of the model with the lowest WAIC value?

Solutions: Similarily, use ensemble instead of link to plot model averaged prediction.

```
##          a         b1      sigma
## 138.451551  18.663040   2.972433
```

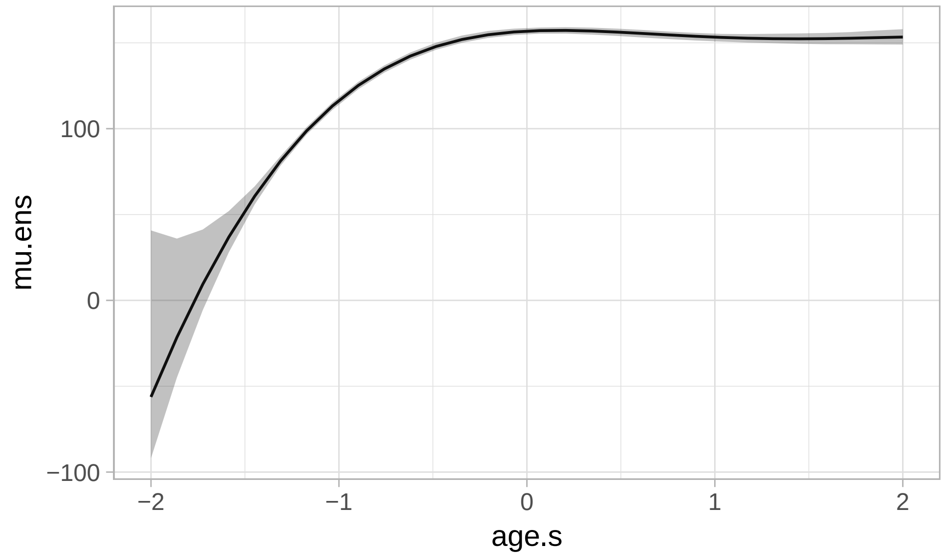
```
## Constructing posterior predictions
```

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

```
## Constructing posterior predictions
```

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
```

```
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]

## Constructing posterior predictions

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

## Constructing posterior predictions

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

## Constructing posterior predictions

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

## Constructing posterior predictions

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



This figure showed that the model is more similar to the higher degree model.