Module 9: Linear Regression

Rebecca C. Steorts

Hoff, Chapter 9

Final Exam

Exam III is **April 12, during class** (open note/open book)

- The material will be on modules 1 – 9.
- I will go over more details closer to the exam

Agenda

- Motivation: oxygen uptake example
- Linear regression
- Multiple and Multivariate Linear Regression
- Bayesian Linear Regression
- Background on the Euclidean norm and argmin
- Ordinary Least Squares + Exercises
- Setting Prior Parameters
- The g-prior
- How does this all fit together

Oxygen uptake case study

Experimental design: 12 male volunteers.

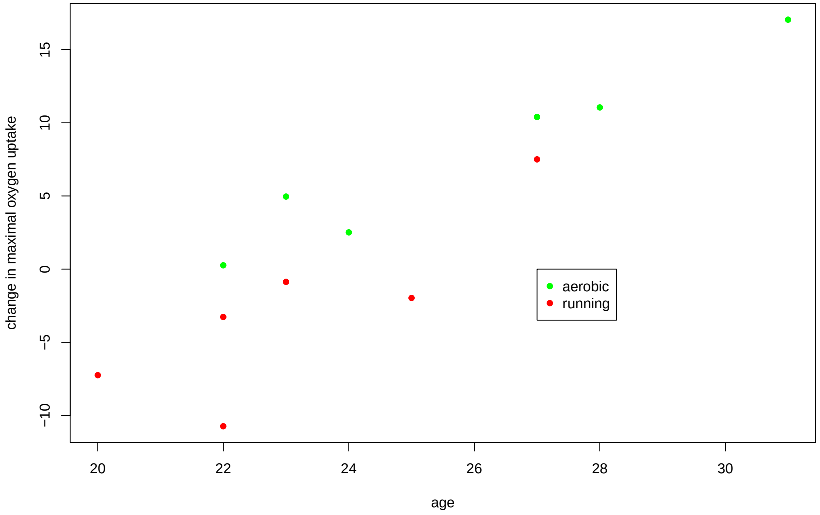
1. O_2 uptake measured at the beginning of the study
2. 6 men take part in a randomized aerobics program
3. 6 remaining men participate in a running program
4. O_2 uptake measured at end of study

What type of exercise is the most beneficial?

Data

```r
# 0 is running
# 1 is aerobic exercise
x1<-c(0,0,0,0,0,0,1,1,1,1,1,1)
# x2 is age
x2<-c(23,22,22,25,27,20,31,23,27,28,22,24)
# change in maximal oxygen uptake
y<-c(-0.87,-10.74,-3.27,-1.97,7.50,
     -7.25,17.05,4.96,10.40,11.05,0.26,2.51)
```

Exploratory Data Analysis



Data analysis

y = change in oxygen uptake (scalar)

x_1 = exercise indicator (0 for running, 1 for aerobic)

x_2 = age

How can we estimate $p(y \mid x_1, x_2)$?

Linear regression

Assume that smoothness is a function of age.

For each group,

$$y = \beta_o + \beta_1 x_2 + \epsilon$$

Linearity means **linear in the parameters** (β's).

Linear regression

We could also try the model

$$y = \beta_o + \beta_1 x_2 + \beta_2 x_2^2 + \beta_3 x_2^3 + \epsilon$$

which is also a linear regression model.

Notation

- $X_{n \times p}$: regression features or covariates (design matrix)
- $\boldsymbol{x}_i$: ith row vector of the regression covariates
- $\boldsymbol{y}_{n \times 1}$: response variable (vector)
- $\boldsymbol{\beta}_{p \times 1}$: vector of regression coefficients

Notation (continued)

$$\boldsymbol{X}_{n \times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ x_{i1} & x_{i2} & \dots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}.$$

- A column of x represents a particular covariate we might be interested in, such as age of a person.
- Denote x_i as the ith row vector of the $X_{n \times p}$ matrix.

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

Notation (continued)

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\boldsymbol{y}_{n\times 1} = X_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}_{n\times 1}$$

Regression models

How does an outcome $\boldsymbol{y}$ vary as a function of the covariates which we represent as $X_{n \times p}$ matrix?

- Can we predict $\boldsymbol{y}$ as a function of each row in the matrix $X_{n \times p}$ denoted by $\boldsymbol{x}_i$.
- Which $\boldsymbol{x}_i$'s have an effect?

Such questions can be assessed via a linear regression model $p(\boldsymbol{y} \mid X)$.

Multiple linear regression

Consider the following:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$$

where

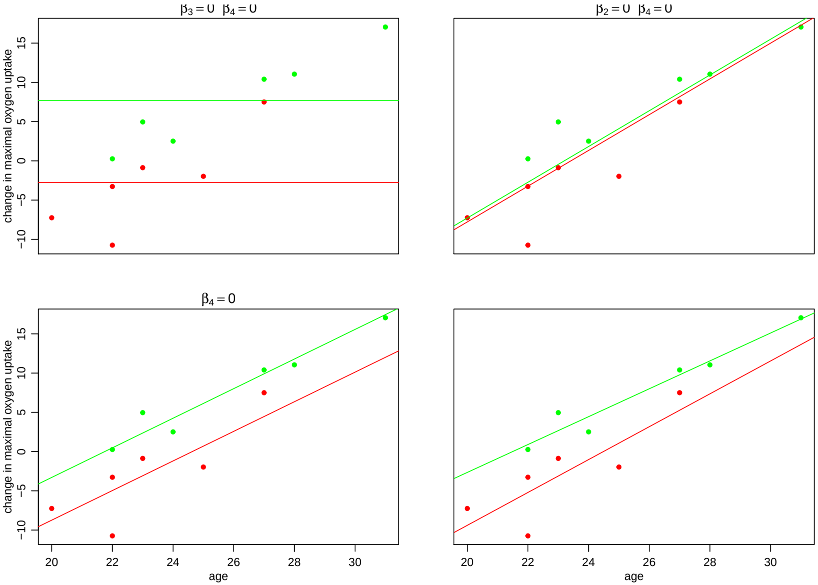
$$x_{i1} = 1 \text{ for subject } i \tag{1}$$
$$x_{i2} = 0 \text{ for running; } 1 \text{ for aerobics} \tag{2}$$
$$x_{i3} = \text{age of subject i} \tag{3}$$
$$x_{i4} = x_{i2} \times x_{i3} \tag{4}$$

Under this model,

$$E[\boldsymbol{y} \mid \boldsymbol{x}] = \beta_1 + \beta_3 \times age \text{ if } x_2 = 0$$

$$E[\boldsymbol{y} \mid \boldsymbol{x}] = (\beta_1 + \beta_2) + (\beta_3 + \beta_4) \times age \text{ if } x_2 = 1$$

Least squares regression lines



Multivariate Setup

Let's assume that we have data points (x_i, y_i) available for all $i = 1, \dots, n$.

- y is the response variable

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1}$$

- $\boldsymbol{x}_i$ is the ith row of the design matrix $X_{n \times p}$.

Consider the regression coefficients

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}_{p \times 1}$$

Normal Regression Model

The Normal regression model specifies that

- $E[Y \mid \boldsymbol{x}_i]$ is linear and
- the sampling variability around the mean is independently and identically (iid) drawn from a normal distribution

$$Y_i = \beta^T \boldsymbol{x}_i + \epsilon_i \tag{5}$$

$$\epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} \text{Normal}(0, \sigma^2) \tag{6}$$

This implies $Y_i \mid \beta, \boldsymbol{x}_i \sim \text{Normal}(\beta^T \boldsymbol{x}_i, \sigma^2)$.

Multivariate Bayesian Normal Regression Model

We can re-write this as a multivariate regression model as:

$$\boldsymbol{y} \mid X, \boldsymbol{\beta}, \sigma^2 \sim \text{MVN}(X\boldsymbol{\beta}, \sigma^2 I_p).$$

We can specify a multivariate Bayesian model as:

$$\begin{gathered} \boldsymbol{y} \mid X, \boldsymbol{\beta}, \sigma^2 \sim \text{MVN}(X\boldsymbol{\beta}, \sigma^2 I_p) \\ \boldsymbol{\beta} \sim \text{MVN}(0, \tau^2 I_p), \end{gathered}$$

where σ^2, τ^2 are known.

Bayesian Normal Regression Model

The likelihood is

$$p(y_1, \ldots, y_n \mid x_1, \ldots x_n, \boldsymbol{\beta}, \sigma^2) \tag{7}$$

$$= \prod_{i=1}^{n} p(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \beta, \sigma^2) \tag{8}$$

$$(2\pi\sigma^2)^{-n/2} \exp\{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (\boldsymbol{y}_i - \beta^T \boldsymbol{x}_i)^2\} \tag{9}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2}(\boldsymbol{y} - X\beta)^T (\sigma^2)^{-1} I_p (\boldsymbol{y} - X\beta)\} \tag{10}$$

Background

The Euclidean norm (L^2 norm or square root of the sum of squares) of $\boldsymbol{y} = (y_1, \ldots, y_n)$ is defined by

$$\|\boldsymbol{y}\|_2 = \sqrt{y_1^2 + \ldots + y_n^2}.$$

It follows that

$$\|\boldsymbol{y}\|_2^2 = y_1^2 + \ldots + y_n^2.$$

Why do we use this notation? It's compact and convenient!

Background

We would like to find

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \|\boldsymbol{y} - X\beta\|_2^2,$$

where the $\arg\min$ (the arguments of the minima) are the points or elements of the domains of some function as which the functions values are minimized.

Ordinary Least Squares

We can estimate the coefficients $\hat{\beta} \in \mathbb{R}^p$ by least squares:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2$$

One can show that

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y}$$

The fitted values are

$$\hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}} = X(X^T X)^{-1} X^T \boldsymbol{y}$$

This is a linear function of $\boldsymbol{y}$, $\hat{y} = H\boldsymbol{y}$, where $H = X(X^T X)^{-1} X^T$ is sometimes called the **hat matrix**.

Exercise 1 (OLS)

Let SSR denote sum of squared residuals.

$$\min_{\beta} SSR(\beta) = \min_{\beta} \|\boldsymbol{y} - X\beta\|_2^2$$

Show that

$$\hat{\beta} = (X^T X)^{-1} X^T \boldsymbol{y}.$$

Ordinary Least squares estimation

Proof: Observe

$$
\begin{aligned}
\frac{\partial SSR(\beta)}{\partial \boldsymbol{\beta}} &:= \frac{\partial \|\mathbf{y} - X\beta\|_2^2}{\partial \boldsymbol{\beta}} && (11)\\
&= \frac{\partial (\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta)}{\partial \boldsymbol{\beta}} && (12)\\
&= \frac{\partial \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T X^T\mathbf{y} + \hat{\beta}^T(X^TX)\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} && (13)\\
&= -2X^T\mathbf{y} + 2X^TX\beta && (14)
\end{aligned}
$$

This implies $-X^T\mathbf{y} + X^TX\beta = 0 \implies \hat{\beta}_{ols} = (X^TX)^{-1}X^T\mathbf{y}$.

This is called the **ordinary least squares estimator**. How do we know it is unique?

Exercise 2 (OLS)

Show that

$$\hat{\beta} \sim MVN(\beta, \sigma^2 (X^T X)^{-1}).$$

Ordinary Least squares estimation

Proof: Recall

$$\hat{\beta} = (X^T X)^{-1} X^T \boldsymbol{Y}.$$

$$E(\hat{\boldsymbol{\beta}}) = E[(X^T X)^{-1} X^T \boldsymbol{Y}] = (X^T X)^{-1} X^T E[\boldsymbol{Y}] = (X^T X)^{-1} X^T X \boldsymbol{\beta}.$$

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}\{(X^T X)^{-1} X^T \boldsymbol{Y}\} && (15)\\
&= (X^T X)^{-1} X^T \sigma^2 I_n X (X^T X)^{-1} && (16)\\
&= \sigma^2 (X^T X)^{-1} && (17)
\end{aligned}$$

$$\hat{\boldsymbol{\beta}} \sim MVN(\boldsymbol{\beta}, \sigma^2 (X^T X)^{-1}).$$

Recall data set up

```r
# running is 0, 1 is aerobic
x1<-c(0,0,0,0,0,0,1,1,1,1,1,1)
# age
x2<-c(23,22,22,25,27,20,31,23,27,28,22,24)
# change in maximal oxygen uptake
y<-c(-0.87,-10.74,-3.27,-1.97,7.50,
     -7.25,17.05,4.96,10.40,11.05,0.26,2.51)
```

Recall data set up

```
(x3 <- x2) #age
```

```
##  [1] 23 22 22 25 27 20 31 23 27 28 22 24
```

```
(x2 <- x1) #aerobic versus running
```

```
##  [1] 0 0 0 0 0 0 1 1 1 1 1 1
```

```
(x1<- seq(1:length(x2))) #index of person
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12
```

```
(x4 <- x2*x3)
```

```
##  [1]  0  0  0  0  0  0 31 23 27 28 22 24
```

Recall data set up

```
(X <- cbind(x1,x2,x3,x4))
```

```
##       x1 x2 x3 x4
##  [1,]  1  0 23  0
##  [2,]  2  0 22  0
##  [3,]  3  0 22  0
##  [4,]  4  0 25  0
##  [5,]  5  0 27  0
##  [6,]  6  0 20  0
##  [7,]  7  1 31 31
##  [8,]  8  1 23 23
##  [9,]  9  1 27 27
## [10,] 10  1 28 28
## [11,] 11  1 22 22
## [12,] 12  1 24 24
```

OLS estimation in R

```r
## using the lm function
fit.ols<-lm(y~ X[,2] + X[,3] +X[,4])
summary(fit.ols)$coef
```

```
##                Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) -51.2939459 12.2522126 -4.1865047 0.003052321
## X[, 2]       13.1070904 15.7619762  0.8315639 0.429775106
## X[, 3]        2.0947027  0.5263585  3.9796120 0.004063901
## X[, 4]       -0.3182438  0.6498086 -0.4897500 0.637457484
```

Exercise 3 (Multivariate inference for regression models)

Let $\boldsymbol{y} = (y_1, \ldots y_n)_{n\times 1}$ and

$$\boldsymbol{y} \mid \boldsymbol{\beta} \sim \text{MVN}(X\boldsymbol{\beta}, \sigma^2 I) \tag{18}$$

$$\boldsymbol{\beta} \sim \text{MVN}(\boldsymbol{\beta}_0, \Sigma_0) \tag{19}$$

Show that the posterior is

$$\boldsymbol{\beta} \mid \boldsymbol{y}, X \sim \text{MVN}(\boldsymbol{\beta}_n, \Sigma_n), \text{ where}$$

$$\boldsymbol{\beta}_n = E[\boldsymbol{\beta} \ \mid \boldsymbol{y}, \boldsymbol{X}, \sigma^2] = (\Sigma_o^{-1} + (X^T X)/\sigma^2)^{-1}(\Sigma_o^{-1}\boldsymbol{\beta}_0 + \boldsymbol{X}^T\boldsymbol{y}/\sigma^2)$$
$$\Sigma_n = \text{Var}[\boldsymbol{\beta} \mid \boldsymbol{y}, X, \sigma^2] = (\Sigma_o^{-1} + (X^T X)/\sigma^2)^{-1}$$

Remark: If $\Sigma_o^{-1} << (X^T X)^{-1}$ then $\boldsymbol{\beta}_n \approx \hat{\boldsymbol{\beta}}_{ols}$

If $\Sigma_o^{-1} >> (X^T X)^{-1}$ then $\boldsymbol{\beta}_n \approx \boldsymbol{\beta}_0$

Exercise 3 Solution

Let's start with the likelihood:

$$\boldsymbol{y} \mid \beta \sim \text{MVN}(X\beta, \sigma^2 I) \tag{20}$$

$$\begin{aligned}
p(\boldsymbol{y} \mid \beta) &\propto \exp\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta^T x_i)^2\} && (21)\\
&\propto \exp\{\frac{-1}{2\sigma^2}(\boldsymbol{y} - X\beta)^T(\boldsymbol{y} - X\beta)\} && (22)\\
&\propto \exp\{\frac{-1}{2\sigma^2}[\boldsymbol{y}^T\boldsymbol{y} + \beta^T X^T X\beta - 2\beta^T X^T \boldsymbol{y}]\} && (23)\\
&\propto \exp\{\frac{-1}{2\sigma^2}[\beta^T X^T X\beta - 2\beta^T X^T \boldsymbol{y}]\} && (24)\\
&\propto \exp\{-\frac{1}{2}[\beta^T \frac{X^T X}{\sigma^2}\beta - 2\beta^T \frac{X^T \boldsymbol{y}}{\sigma^2}]\} && (25)
\end{aligned}$$

This implies that $A_1 = \frac{X^T X}{\sigma^2}$ and $b_1 = \frac{X^T \boldsymbol{y}}{\sigma^2}$.

Exercise 3 Solution

Now, let's consider the prior.

$$\beta \sim \text{MVN}(\beta_0, \Sigma_0) \tag{26}$$

$$\begin{aligned}
p(\beta) &\propto \exp\{-\frac{1}{2}(\beta - \beta_0)^T \Sigma_0^{-1} (\beta - \beta_0)\} && (27) \\
&\propto \exp\{-\frac{1}{2}[\beta^T \Sigma_0^{-1} \beta + \beta_0^T \Sigma_0^{-1} \beta_0 - 2\beta^T \Sigma_0^{-1} \beta_0]\} && (28) \\
&\propto \exp\{-\frac{1}{2}[\beta^T \Sigma_0^{-1} \beta - 2\beta^T \Sigma_0^{-1} \beta_0]\} && (29)
\end{aligned}$$

This implies that $A_o = \Sigma_0^{-1}$ and $b_o = \Sigma_0^{-1}\beta_0$.

Exercise 3 Solution

The posterior is

$$p(\beta \mid \boldsymbol{y}) \propto \exp\{-\frac{1}{2}[\beta^T \frac{X^T X}{\sigma^2} \beta - 2\beta^T \frac{X^T \boldsymbol{y}}{\sigma^2}]\} \tag{30}$$

$$\times \exp\{-\frac{1}{2}[\beta^T \Sigma_0^{-1} \beta - 2\beta^T \Sigma_0^{-1} \beta_0]\} \tag{31}$$

$$= \exp\{-\frac{1}{2}[\beta^T A_1 \beta - 2\beta^T b_1]\} \tag{32}$$

$$\times \quad \exp\{-\frac{1}{2}[\beta^T A_o \beta - 2\beta^T b_o]\} \tag{33}$$

$$= \exp\{-\frac{1}{2}[\beta^T (A_o + A_1) \beta - 2\beta^T (b_o + b_1)]\} \tag{34}$$

Exercise 3 Solution

Let $A_n = A_o + A_1$ and $b_n = b_o + b_1$.

Using the kernel of the multivariate normal, we can now find the posterior mean and the posterior covariance:

$$A_n = A_o + A_1 = \Sigma_0^{-1} + \frac{X^T X}{\sigma^2}.$$

$$b_n = b_o + b_1 = \Sigma_0^{-1}\beta_0 + \frac{X^T \boldsymbol{y}}{\sigma^2}.$$

$$\beta \mid \boldsymbol{y} \sim MVN(A_n^{-1} b_n, A_n^{-1}) =: MVN(\beta_n, \Sigma_n),$$

where

$$\mu_n = (\Sigma_0^{-1} + \frac{X^T X}{\sigma^2})^{-1}(\Sigma_0^{-1}\beta_0 + \frac{X^T \boldsymbol{y}}{\sigma^2})$$

and

$$\Sigma_n = (\Sigma_0^{-1} + \frac{X^T X}{\sigma^2})^{-1}.$$

Multivariate inference for regression models

The posterior from Exercise 3 can be shown to be

$$\boldsymbol{\beta} \mid \boldsymbol{y}, X \sim \text{MVN}(\boldsymbol{\beta}_n, \Sigma_n)$$

where

$$\boldsymbol{\beta}_n = E[\beta \ \mid \boldsymbol{y}, \boldsymbol{X}, \sigma^2] = (\Sigma_o^{-1} + (X^T X)/\sigma^2)^{-1}(\Sigma_o^{-1}\boldsymbol{\beta}_0 + X^T \boldsymbol{y}/\sigma^2)$$

$$\Sigma_n = \text{Var}[\beta \mid \boldsymbol{y}, \boldsymbol{X}, \sigma^2] = (\Sigma_o^{-1} + (X^T X)/\sigma^2)^{-1}$$

Setting prior parameters

How would you set the prior parameters for

- σ^2
- Σ_o
- β_0

Setting prior parameters

- Estimate σ^2 by $\dfrac{y^T y - \hat{\beta}_{ols}^T X^T y}{n-(p+1)}$ because this is an unbiased estimator of σ^2.
- Set
$$\Sigma_o^{-1} = \frac{(X^T X)}{n\sigma^2},$$
which is known as the unit information prior (Kass and Wasserman, 1995).
- Set $\beta_0 = \hat{\beta}_{ols}$. (This centers the prior distribution of β around the OLS estimate).

Why are these reasonable choices?

Setting prior parameters

- Do you think that the posterior would be sensitive to the choice of these parameters?
- How could you improve upon our choices regarding priors on β_0 and Σ_0?

The g-prior

To improve things by doing the **least amount of calculus**, we can put a *g-prior* on β (not β_0).

The g-prior on β has the following form:

$$\beta \mid \boldsymbol{X}, \sigma^2 \sim MVN(0, g\, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}),$$

where g is a constant, such as $g = n$.

It can be shown that (Zellner, 1986):

1. g shrinks the coefficients and can prevent overfitting to the data
2. if $g = n$, then as n increases, inference approximates that using $\hat{\beta}_{ols}$

The g-prior

Under the g-prior, it follows that

$$\boldsymbol{\beta}_n = E[\beta \mid \boldsymbol{y}, \boldsymbol{X}, \sigma^2] \tag{35}$$
$$= \left(\frac{X^T X}{g\sigma^2} + \frac{X^T X}{\sigma^2} \right)^{-1} \frac{X^T y}{\sigma^2} \tag{36}$$
$$= \frac{g}{g+1} (X^T X)^{-1} X^T y = \frac{g}{g+1} \hat{\beta}_{ols} \tag{37}$$

$$\Sigma_n = \mathrm{Var}[\beta \mid \boldsymbol{y}, \boldsymbol{X}, \sigma^2] \tag{38}$$
$$= \left(\frac{X^T X}{g\sigma^2} + \frac{X^T X}{\sigma^2} \right)^{-1} = \frac{g}{g+1} \sigma^2 (X^T X)^{-1} \tag{39}$$
$$= \frac{g}{g+1} \mathrm{Var}[\hat{\beta}_{ols}] \tag{40}$$

Prior on Σ_0

What prior would you place on Σ_0 and why?

Next steps

- How do all these concepts fit together? How can you build a hierarchical model using linear regression and the tools that you've learned?
- I recommend doing the derivations from this module on your own.
- I recommend reading through Hoff to solidify you knowledge. This material is around page 153, but chapter 9 is helpful regarding being complementary to this material.
- You could also code this up to further solidify you knowledge of this, but you'll get practice on this with lab 10 and homework 8.

Linear Regression Applied to Swimming (Lab 10)

- We will consider Exercise 9.1 in Hoff very closely to illustrate linear regression.
- The data set we consider contains times (in seconds) of four high school swimmers swimming 50 yards.
- There are 6 times for each student, taken every two weeks.
- Each row corresponds to a swimmer and a higher column index indicates a later date.
- This corresponds with Lab 10 and Homework 8 (the last homework)!

Data set

```
read.table("data/swim.dat",header=FALSE)
```

```
## Warning in read.table("data/swim.dat", header = FALSE):
## found by readTableHeader on 'data/swim.dat'

##     V1   V2   V3   V4   V5   V6
## 1 23.1 23.2 22.9 22.9 22.8 22.7
## 2 23.2 23.1 23.4 23.5 23.5 23.4
## 3 22.7 22.6 22.8 22.8 22.9 22.8
## 4 23.7 23.6 23.7 23.5 23.5 23.4
```

Full conditionals (Task 1)

We will fit a separate linear regression model for each swimmer, with swimming time as the response and week as the explanatory variable. Let $y_i \in \mathbb{R}^6$ be the 6 recorded times for swimmer i. Let

$$X_i = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ \dots & \\ 1 & 9 \\ 1 & 11 \end{bmatrix}$$

be the design matrix for swimmer i. Then we use the following linear regression model:

$$\begin{aligned} Y_i &\sim \mathcal{N}_6\left(X_i\beta_i, \tau_i^{-1}\mathcal{I}_6\right) \\ \beta_i &\sim \mathcal{N}_2\left(\beta_0, \Sigma_0\right) \\ \tau_i &\sim \mathrm{Gamma}(a, b). \end{aligned}$$

Derive full conditionals for β_i and τ_i.

Solution (Task 1)

The conditional posterior for β_i is multivariate normal with

$$\mathbb{V}[\beta_i \mid Y_i, X_i, \tau_i] = (\Sigma_0^{-1} + \tau_i X_i^T X_i)^{-1}$$
$$\mathbb{E}[\beta_i \mid Y_i, X_i, \tau_i] = (\Sigma_0^{-1} + \tau_i X_i^T X_i)^{-1}(\Sigma_0^{-1}\beta_0 + \tau_i X_i^T Y_i).$$

while

$$\tau_i \mid Y_i, X_i, \beta \sim \text{Gamma}\left(a + 3\,,\; b + \frac{(Y_i - X_i\beta_i)^T(Y_i - X_i\beta_i)}{2}\right).$$

These can be found in in Hoff in section 9.2.1.

I highly recommend that you derive these as practice for the final exam.

Task 2

Complete the prior specification by choosing a, b, β_0, and Σ_0. Let your choices be informed by the fact that times for this age group tend to be between 22 and 24 seconds.

Solution (Task 2)

Choose $a = b = 0.1$ so as to be somewhat uninformative.

Choose $\beta_0 = [23 \ \ 0]^T$ with

$$\Sigma_0 = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}.$$

This centers the intercept at 23 (the middle of the given range) and the slope at 0 (so we are assuming no increase) but we choose the variance to be a bit large to err on the side of being less informative.

Gibbs sampler (Task 3)

Code a Gibbs sampler to fit each of the models. For each swimmer i, obtain draws from the posterior predictive distribution for y_i^*, the time of swimmer i if they were to swim two weeks from the last recorded time.

Posterior Prediction (Task 4)

The coach has to decide which swimmer should compete in a meet two weeks from the last recorded time. Using the posterior predictive distributions, compute $\Pr\left\{y_i^* = \max\left(y_1^*, y_2^*, y_3^*, y_4^*\right)\right\}$ for each swimmer i and use these probabilities to make a recommendation to the coach.

Final Grades

I am proposing to drop your lowest exam grade (out of Exam I, Exam II, and the Exam III).

- Homework: 30 percent
- Highest Exam: 35
- Second Highest Exam: 35
- So your **two highest exam scores** will be weighted evenly and you lowest exam score will be completely dropped.
- It's highly recommended that you take Exam III!

Course Evaluations

- I would be very appreciative if you would fill out the course evaluations
- They are located at: duke.evaluationkit.com
- If there is a 100 percent response rate, I will give everyone in the course 1 point on their final grade.