## STA 360/602L: Module 6.1

BAYESIAN LINEAR REGRESSION

DR. OLANREWAJU MICHAEL AKANDE



### MOTIVATING EXAMPLE

- Let's consider the problem of predicting swimming times for high school swimmers to swim 50 yards.
- We have data collected on four students, each with six times taken (every two weeks).
- Suppose the coach of the team wants to use the data to recommend one of the swimmers to compete in a swim meet in two weeks time.
- Since we want to predict swimming times given week, one option would be regression models.
- In a typical regression setup, we store the predictor variables in a matrix  $X_{n\times p}$ , so n is the number of observations and p is the number of variables.
- You should all know how to write down and fit linear regression models of the most common forms, so let's only review the most important details.



### NORMAL REGRESSION MODEL

lacktriangle The model assumes the following distribution for a response variable  $Y_i$  given multiple covariates/predictors  $m{x}_i = (1, x_{i1}, x_{i2}, \dots, x_{i(p-1)})$ .

$$Y_i = eta_0 + eta_1 x_{i1} + eta_2 x_{i2} + \ldots + eta_{p-1} x_{i(p-1)} + \epsilon_i; \quad \epsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2).$$

or in vector form for the parameters,

$$Y_i = oldsymbol{eta}^T oldsymbol{x}_i + \epsilon_i; \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),$$

where 
$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}).$$

We can also write the model as:

$$Y_i \overset{iid}{\sim} \mathcal{N}(oldsymbol{eta}^Toldsymbol{x}_i, \sigma^2); \ p(y_i|oldsymbol{x}_i) = \mathcal{N}(oldsymbol{eta}^Toldsymbol{x}_i, \sigma^2).$$

lacksquare That is, the model assumes  $\mathbb{E}[Y|oldsymbol{x}]$  is linear.

### LIKELIHOOD

lacksquare Given that we have  $Y_i \overset{iid}{\sim} \mathcal{N}(oldsymbol{eta}^Toldsymbol{x}_i, \sigma^2)$ , the likelihood is

$$egin{aligned} p(y_i,\dots,y_n|m{x}_1,\dots,m{x}_n,m{eta},\sigma^2) &= \prod_{i=1}^n p(y_i|m{x}_i,m{eta},\sigma^2) \ &= \prod_{i=1}^n rac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-rac{1}{2\sigma^2}(y_i-m{eta}^Tm{x}_i)^2
ight\} \ &\propto (\sigma^2)^{-rac{n}{2}} \exp\left\{-rac{1}{2\sigma^2}\sum_{i=1}^n (y_i-m{eta}^Tm{x}_i)^2
ight\}. \end{aligned}$$

- From all our work with normal models, we already know it would be convenient to specify a (multivariate) normal prior on  $\beta$  and a gamma prior on  $1/\sigma^2$ , so let's start there.
- Two things to immediately notice:
  - since  $\beta$  is a vector, it might actually be better to rewrite this kernel in multivariate form altogether, and
  - when combining this likelihood with the prior kernel, we will need to find a way to detach  $\beta$  from  $x_i$ .



### **M**ULTIVARIATE FORM

Let

$$oldsymbol{Y} = egin{bmatrix} Y_1 \ Y_2 \ dots \ Y_n \end{bmatrix} oldsymbol{X} = egin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1(p-1)} \ 1 & x_{21} & x_{22} & \dots & x_{2(p-1)} \ dots & dots & dots & dots \ 1 & x_{n1} & x_{n2} & \dots & x_{n(p-1)} \end{bmatrix} oldsymbol{eta} = egin{bmatrix} eta_0 \ eta_1 \ eta_2 \ dots \ eta_{n-1} \end{bmatrix} oldsymbol{\epsilon} = egin{bmatrix} \epsilon_1 \ eta_2 \ dots \ eta_{n-1} \end{bmatrix} oldsymbol{I} = egin{bmatrix} 1 & 0 & \dots & 0 \ 0 & 1 & \dots & 0 \ 0 & 1 & \dots & 0 \ dots & dots & dots & dots \ 0 & 0 & \dots & 1 \end{bmatrix}$$

Then, we can write the model as

$$oldsymbol{Y} = oldsymbol{X}oldsymbol{eta} + oldsymbol{\epsilon}; \ \ oldsymbol{\epsilon} \sim \mathcal{N}_n(0, \sigma^2 oldsymbol{I}_{n imes n}).$$

That is, in multivariate form, we have

$$oldsymbol{Y} \sim \mathcal{N}_n(oldsymbol{X}oldsymbol{eta}, \sigma^2oldsymbol{I}_{n imes n}).$$

### FREQUENTIST ESTIMATION RECAP

• OLS estimate of  $\beta$  is given by

$$\hat{oldsymbol{eta}}_{ ext{ols}} = \left(oldsymbol{X}^Toldsymbol{X}
ight)^{-1}oldsymbol{X}^Toldsymbol{y}.$$

Predictions can then be written as

$$\hat{m{y}} = m{X}\hat{m{eta}}_{ ext{ols}} = m{X}\left[ig(m{X}^Tm{X}ig)^{-1}m{X}^Tm{y}
ight] = \left[m{X}ig(m{X}^Tm{X}ig)^{-1}m{X}^T
ight]m{y}.$$

■ The variance of the OLS estimates of all p coefficients is

$$\mathbb{V}ar\left[\hat{oldsymbol{eta}}_{ ext{ols}}
ight] = \sigma^2ig(oldsymbol{X}^Toldsymbol{X}ig)^{-1}.$$

Finally,

$$s_e^2 = rac{(oldsymbol{y} - oldsymbol{X}\hat{oldsymbol{eta}}_{ ext{ols}})^T(oldsymbol{y} - oldsymbol{X}\hat{oldsymbol{eta}}_{ ext{ols}})}{n-p}.$$

#### BAYESIAN SPECIFICATION

The likelihood for the regression model becomes

$$egin{aligned} p(oldsymbol{y}|oldsymbol{X},oldsymbol{eta},\sigma^2) &\propto (\sigma^2)^{-rac{n}{2}} \exp\left\{-rac{1}{2\sigma^2}(oldsymbol{y}-oldsymbol{X}oldsymbol{eta})^T(oldsymbol{y}-oldsymbol{X}oldsymbol{eta}) 
ight. \ &\propto (\sigma^2)^{-rac{n}{2}} \exp\left\{-rac{1}{2\sigma^2}ig[oldsymbol{y}^Toldsymbol{y}-2oldsymbol{eta}^Toldsymbol{X}^Toldsymbol{y}+oldsymbol{eta}^Toldsymbol{X}^Toldsymbol{X}oldsymbol{eta}
ight]
ight\}. \end{aligned}$$

• We can start with the following semi-conjugate prior for  $\beta$ :

$$\pi(oldsymbol{eta}) = \mathcal{N}_p(oldsymbol{\mu}_0, \Sigma_0).$$

That is, the pdf is

$$\pi(oldsymbol{eta}) = (2\pi)^{-rac{p}{2}} |\Sigma_0|^{-rac{1}{2}} \exp\left\{-rac{1}{2}(oldsymbol{eta} - oldsymbol{\mu}_0)^T \Sigma_0^{-1} (oldsymbol{eta} - oldsymbol{\mu}_0)
ight\}.$$

Recall from our multivariate normal model that we can write this pdf as

$$\pi(oldsymbol{eta}) \propto \exp\left\{-rac{1}{2}oldsymbol{eta}^T\Sigma_0^{-1}oldsymbol{eta} + oldsymbol{eta}^T\Sigma_0^{-1}oldsymbol{\mu}_0
ight\}.$$

### MULTIVARIATE NORMAL MODEL RECAP

- To avoid doing all work from scratch, we can leverage results from the multivariate normal model.
- lacksquare In particular, recall that if  $oldsymbol{Y} \sim \mathcal{N}_p(oldsymbol{ heta}, \Sigma)$ ,

$$p(oldsymbol{y}|oldsymbol{ heta},\Sigma)\propto \exp\left\{-rac{1}{2}oldsymbol{ heta}^T(\Sigma^{-1})oldsymbol{ heta}+oldsymbol{ heta}^T(\Sigma^{-1}ar{oldsymbol{y}})
ight\}$$

and

$$\pi(oldsymbol{ heta}) \propto \exp\left\{-rac{1}{2}oldsymbol{ heta}^T\Lambda_0^{-1}oldsymbol{ heta} + oldsymbol{ heta}^T\Lambda_0^{-1}oldsymbol{\mu}_0
ight\}$$

Then

$$\pi(m{ heta}|\Sigma,m{y}) \propto \exp\left\{-rac{1}{2}m{ heta}^T\left[\Lambda_0^{-1}+\Sigma^{-1}
ight]m{ heta}+m{ heta}^T\left[\Lambda_0^{-1}m{\mu}_0+\Sigma^{-1}ar{m{y}}
ight]
ight\} \;\equiv\; \mathcal{N}_p(m{\mu}_n,\Lambda_n)$$

where

$$egin{aligned} \Lambda_n &= \left[\Lambda_0^{-1} + \Sigma^{-1}
ight]^{-1} \ oldsymbol{\mu}_n &= \Lambda_n \left[\Lambda_0^{-1} oldsymbol{\mu}_0 + \Sigma^{-1} ar{oldsymbol{y}}
ight]. \end{aligned}$$

• For inference on  $\beta$ , rewrite the likelihood as

$$egin{aligned} p(oldsymbol{y}|oldsymbol{X},oldsymbol{eta},\sigma^2) &\propto (\sigma^2)^{-rac{n}{2}} \exp\left\{-rac{1}{2\sigma^2}ig[oldsymbol{y}^Toldsymbol{y} - 2oldsymbol{eta}^Toldsymbol{X}^Toldsymbol{y} + oldsymbol{eta}^Toldsymbol{X}^Toldsymbol{X}etaig]
ight\} \ &\propto \exp\left\{-rac{1}{2}oldsymbol{eta}^Tigg(rac{1}{\sigma^2}oldsymbol{X}^Toldsymbol{X}igg)oldsymbol{eta} + oldsymbol{eta}^Tigg(rac{1}{\sigma^2}oldsymbol{X}^Toldsymbol{y}igg)
ight\}. \end{aligned}$$

Again, with the prior written as

$$\pi(oldsymbol{eta}) \propto \exp\left\{-rac{1}{2}oldsymbol{eta}^T\Sigma_0^{-1}oldsymbol{eta} + oldsymbol{eta}^T\Sigma_0^{-1}oldsymbol{\mu}_0
ight\},$$

both forms look like what we have on the previous page. It is then easy to read off the full conditional for  $\beta$ .

That is,

$$egin{aligned} \pi(oldsymbol{eta}|oldsymbol{y},oldsymbol{X},\sigma^2)&\propto p(oldsymbol{y}|oldsymbol{X},oldsymbol{eta},\sigma^2)\cdot\pi(oldsymbol{eta}) \ &\propto \exp\left\{-rac{1}{2}oldsymbol{eta}^T\left[\Sigma_0^{-1}+rac{1}{\sigma^2}oldsymbol{X}^Toldsymbol{X}
ight]oldsymbol{eta}+oldsymbol{eta}^T\left[\Sigma_0^{-1}oldsymbol{\mu}_0+rac{1}{\sigma^2}oldsymbol{X}^Toldsymbol{y}
ight]
ight\} \ &\equiv \mathcal{N}_p(oldsymbol{\mu}_n,\Sigma_n). \end{aligned}$$

Comparing this to the prior

$$\pi(oldsymbol{eta}) \propto \exp\left\{-rac{1}{2}oldsymbol{eta}^T\Sigma_0^{-1}oldsymbol{eta} + oldsymbol{eta}^T\Sigma_0^{-1}oldsymbol{\mu}_0
ight\},$$

means

$$egin{aligned} \Sigma_n &= \left[ \Sigma_0^{-1} + rac{1}{\sigma^2} oldsymbol{X}^T oldsymbol{X} 
ight]^{-1} \ oldsymbol{\mu}_n &= \Sigma_n \left[ \Sigma_0^{-1} oldsymbol{\mu}_0 + rac{1}{\sigma^2} oldsymbol{X}^T oldsymbol{y} 
ight]. \end{aligned}$$

• Next, we move to  $\sigma^2$ . From previous work, we already know the inversegamma distribution with be semi-conjugate.

$$lacksquare ext{First, recall that } \mathcal{IG}(y;a,b) \equiv rac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-rac{b}{y}} \, .$$

$$lacksquare$$
 So, if we set  $\pi(\sigma^2)=\mathcal{IG}\left(rac{
u_0}{2},rac{
u_0\sigma_0^2}{2}
ight)$ , we have

$$egin{aligned} \pi(\sigma^2|m{y},m{X},m{eta})&\propto p(m{y}|m{X},m{eta},\sigma^2)\cdot\pi(\sigma^2) \ &\propto (\sigma^2)^{-rac{n}{2}}\exp\left\{-\left(rac{1}{\sigma^2}
ight)rac{(m{y}-m{X}m{eta})^T(m{y}-m{X}m{eta})}{2}
ight\} \ &\qquad ext{} \times (\sigma^2)^{-\left(rac{
u_0}{2}+1
ight)}e^{-\left(rac{1}{\sigma^2}
ight)\left[rac{
u_0\sigma_0^2}{2}
ight]} \end{aligned}$$

That is,

$$egin{aligned} \pi(\sigma^2|m{y},m{X},m{eta}) &\propto (\sigma^2)^{-rac{n}{2}} \exp\left\{-\left(rac{1}{\sigma^2}
ight)rac{(m{y}-m{X}m{eta})^T(m{y}-m{X}m{eta})}{2}
ight\} \ & imes (\sigma^2)^{-\left(rac{
u_0}{2}+1
ight)}e^{-\left(rac{1}{\sigma^2}
ight)\left[rac{
u_0\sigma_0^2}{2}
ight]} \ &\propto (\sigma^2)^{-\left(rac{
u_0+n}{2}+1
ight)}e^{-\left(rac{1}{\sigma^2}
ight)\left[rac{
u_0\sigma_0^2+(m{y}-m{X}m{eta})^T(m{y}-m{X}m{eta})}{2}
ight]} \ &\equiv \mathcal{I}\mathcal{G}\left(rac{
u_n}{2},rac{
u_n\sigma_n^2}{2}
ight), \end{aligned}$$

where

$$egin{aligned} 
u_n &= 
u_0 + n; \quad \sigma_n^2 = rac{1}{
u_n} igl[ 
u_0 \sigma_0^2 + (oldsymbol{y} - oldsymbol{X} oldsymbol{eta})^T (oldsymbol{y} - oldsymbol{X} oldsymbol{eta}) igr] = rac{1}{
u_n} igl[ 
u_0 \sigma_0^2 + ext{SSR}(oldsymbol{eta}) igr] \,. \end{aligned}$$

 $= (y - X\beta)^T (y - X\beta)$  is the sum of squares of the residuals (SSR).

# WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!

