STA 602 HW 11

Ryan Tang

November 18th 2022

1 Exercise 9.1

(a) Analysis We have the following graphical model, one for each swimmer $k \in \{1, \ldots, K = 4\}$. For each swimmer, we have 6 weeks of data points that record their swim time for a 50 yards run, $i \in \{1, \ldots, N = 6\}$. Here we like to fit a linear regression for each swimmer; thus, in a total of 4 linear regressions, each with its respective priors.

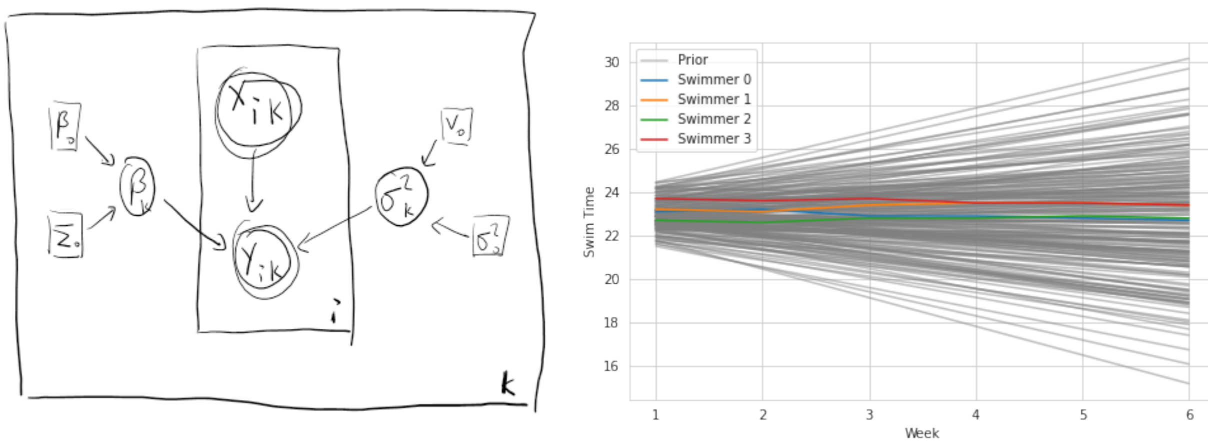

Figure 1: (Left) Typical Bayesian Linear Regression Graphical Model. (Right) Prior Believe Check. Each grey line is a sampled prediction using just the priors.

We are also given some knowledge on typical competitive times for the age group, ranging from 22 to 24 seconds. We will use this to construct out priors accordingly, and the following is the full spec of the model. Some notations, $X_k \in \mathbf{R}^{N \times p}$, $\mathbf{y}_k \in \mathbf{R}^N$, and $\beta_k \in \mathbf{R}^p$, where $p = 2$ in our case, one for the intercept, and one for the weeks.

$$
\begin{aligned}
y_{ik} &\overset{iid}{\sim} \mathcal{N}(X_k \beta_k, \sigma_k^2 \mathbf{I}_n) \\
\beta_k &\sim \mathcal{N}(\beta_o, \Sigma_o) \\
1/\sigma_k^2 = \gamma_k &\sim \text{Gamma}(\frac{\nu_o}{2}, \frac{\nu_o \sigma_o^2}{2})
\end{aligned}
$$

Here we specified $\beta_o = (23, 0)^\intercal$, $\Sigma_o = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$, $\nu_0 = 1$, $\sigma_o^2 = 1$, and the resulting prior believes is also plotted in Figure 1, right plot.

Lastly, we obtained the posterior predictive distribution using the resulting linear regression assuming all swimmers had only swam two weeks from the last recorded time. The distributions are plotted in Figure 2

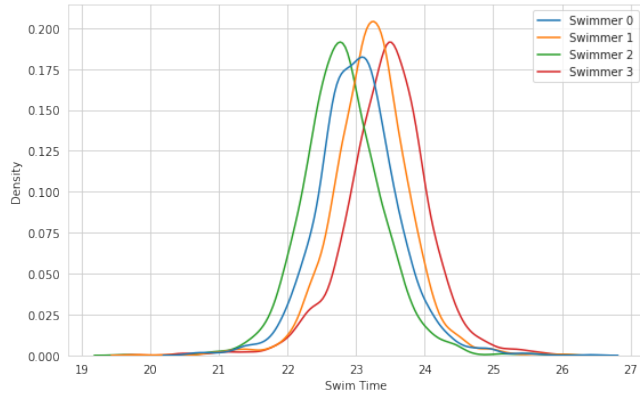

Figure 2: Posterior predictive distribution of Swim Time given only swim two weeks from the last recorded time.

(b) Recommendation In order to make a recommendation to the coach, we took the posterior predictive samples and calculated $Pr(Y_j^* = max\{Y_k\}|Y)$. We already have the samples while constructing Figure 2. Hence, the resulting probability bar chart is shown in Figure 3. According to the analysis, we should consider swimmer 3.

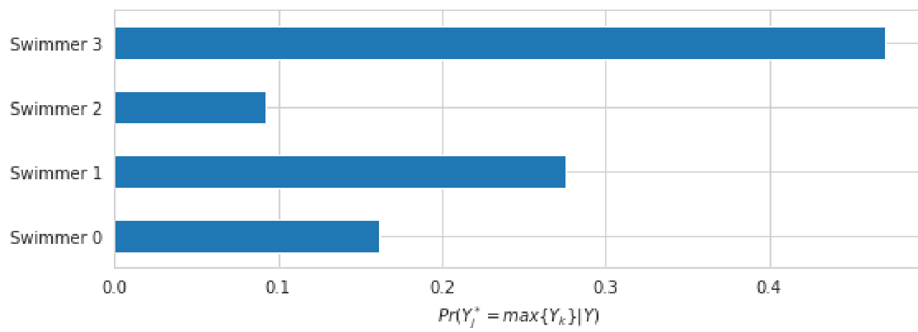

Figure 3: Probability for being the best.

2 Exercise 9.2

We are revisiting the diabetes dataset interested in making inferences on the glucose level. We used all other covariates, except 'diabetes', in the Bayesian linear regression model training. Hence the design matrix $X \in \mathbf{R}^{N\times(p+1)}$, where $N = 532$ and $p = 6$.

(a) Linear Regression First, we fit a typical regression using all covariates and a g-prior for β. Therefore, we have the following model specs.

$$\begin{aligned} y_i &\sim \mathcal{N}(X\beta, \sigma^2) \\ 1/\sigma^2 = \gamma &\sim \text{Gamma}(\frac{\nu_o}{2}, \frac{\nu_o\sigma_o^2}{2}) \\ \beta &\sim \mathcal{N}(0, g\sigma^2(X^\intercal X)^{-1}) \end{aligned}$$

We set $\nu_o = 2$, $\sigma_o^2 = 1$ and $g = n = 532$. The resulting confidence interval of all parameters are shown in the below table.

95% CI	β intercept	β npreg	β bp	β skin	β bmi	β ped	β age	σ^2
2.5%	35.12	-1.62	-0.035	-0.11	0.15	2.98	0.44	746.87
97.5%	70.04	0.30	0.43	0.51	1.16	17.94	1.08	937.10

(a) Bayesian Model Selection and Averaging (BMA) Here we like to do BMA, and the simplest way we can perform this is by assigning a binary vector $\mathbf{z} \in \mathbb{R}^p$ the same size of β, with each z_j be an independent Bernoulli random variable. Hence, the resulting model becomes as the following. And the resulting new posterior confidence interval and inclusion probabilities are also illustrated below. Note, $\circ$ is the Hadamard product. Comparing the results to the original linear regression, we can see that most of β_j got shrunk, and npreg, bp, and skin even shrunk towards 0.

$$\begin{aligned} y_i &\sim \mathcal{N}(X(\mathbf{z}\circ\beta), \sigma^2) \\ z_j &\sim \text{Bernoulli}(q) \\ 1/\sigma^2 = \gamma &\sim \text{Gamma}(\frac{\nu_o}{2}, \frac{\nu_o\sigma_o^2}{2}) \\ \beta &\sim \mathcal{N}(0, g\sigma^2(X^\intercal X)^{-1}) \end{aligned}$$

95% CI	β intercept	β npreg	β bp	β skin	β bmi	β ped	β age	σ^2
2.5%	53.81	-0.86	0.0	0.0	0.76	0.0	0.59	760.39
97.5%	74.09	0.0	0.0	0.0	1.25	14.18	0.91	916.90

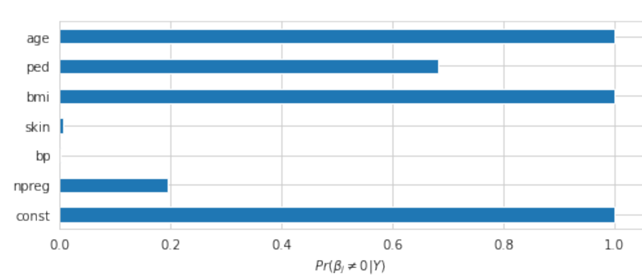

Figure 4: Posterior Parameters Inclusion Probability

3 Exercise 9.3

(a) Bayesian Linear Regression vs OLS First, we fit a typical regression using all covariates and a g-prior for β. Therefore, we have the following model specs.

$$y_i \sim \mathcal{N}(X\beta, \sigma^2)$$
$$1/\sigma^2 = \gamma \sim \text{Gamma}(\frac{\nu_o}{2}, \frac{\nu_o \sigma_o^2}{2})$$
$$\beta \sim \mathcal{N}(0, g\sigma^2(X^\intercal X)^{-1})$$

We set $\nu_o = 2$, $\sigma_o^2 = 1$ and $g = n = 532$. The resulting confidence interval of β, the marginal means, and the ordinary least squares (OLS) are shown below for comparison purposes.

95% CI	Intercept	M	So	Ed	Po1	Po2	LF	M.F	Pop	NW	U1	U2	GDP	Ineq	Prob	Time
2.5%	-0.14	0.03	-0.32	0.21	-0.02	-2.32	-0.34	-0.14	-0.30	-0.21	-0.62	0.04	-0.23	0.29	-0.53	-0.29
mean	0.0	0.28	0.0	0.54	1.46	-0.79	-0.07	0.14	-0.07	0.11	-0.27	0.36	0.23	0.71	-0.28	-0.06
97.5%	0.14	0.53	0.33	0.86	2.95	0.76	0.20	0.41	0.17	0.42	0.08	0.69	0.69	1.14	-0.04	0.17
OLS	0.0	0.29	0.0	0.54	1.47	-0.78	-0.07	0.13	-0.07	0.11	-0.27	0.37	0.24	0.73	-0.29	-0.06

We can see that the posterior marginal means of β obtained through Bayesian Linear Regression is very close to the OLS results. We also see "Ed", "Po1", and "Ineq" are strong predictors for crime rates, and "Po2" works the opposite.

(b) Prediction Comparison Again, to make a comparison, we split the crime data into 50% train and 50% test. Then, we re-trained both models, OLS and BayesLR, using the training set only, and made a prediction on the test set. Figure 5 shows the respective predictions $\hat{y}_{ols}, \hat{y}_{bayes}$ against the true y values using scatter plots.

- OLS achieved 3.14 MSE.
- BayesLR achieved 2.76 MSE, 12% improvements. Although subtle, we can certainly see that the right plot is closer to the diagonal line.

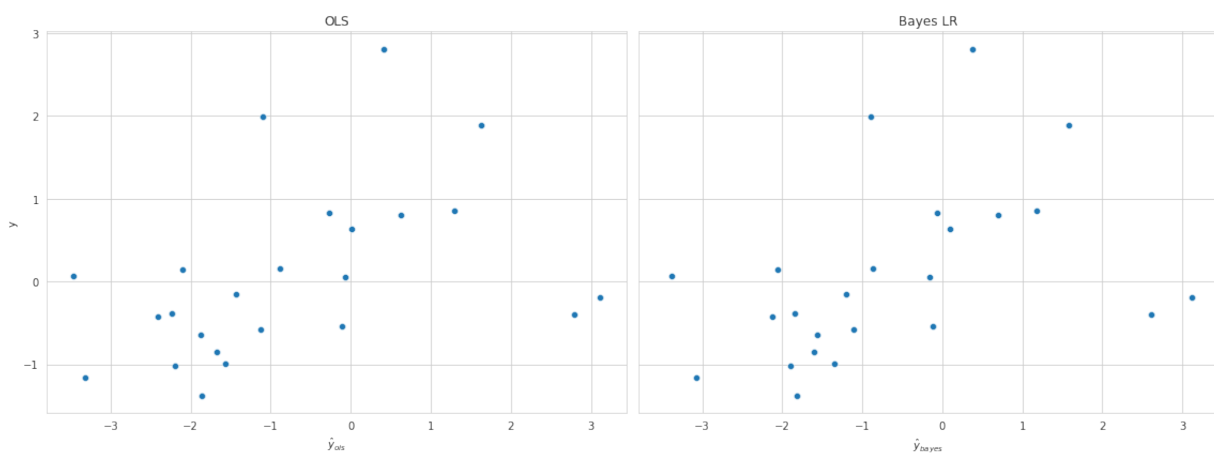

Figure 5: Prediction Comparison between OLS and BayesLR

(c) Cross Validate Lastly, we ran a 10 cross-validate on the two methods. Each fold is a random 50-50 split. The resulting average MSE for the two models is calculated. Surprisingly much lower than the previous MSE in section (b).

- OLS average MSE = 1.55
- BayesLR average MSE = 1.10