CS 1675 Spring 2022 Homework: 08

Assigned March 24, 2022; Due: March 30, 2022

Sameera Boppana

Submission time: March 30, 2022 at 11:00PM EST

Collaborators

Include the names of your collaborators here.

Jeffery Janotka

Overview

This homework assignment is focused on model complexity and the influence of the prior **regularization** strength. You will fit non-Bayesian and Bayesian linear models, compare them, and make predictions to visualize the trends. You will use multiple prior *strengths* to study the impact on the coefficient posteriors and on the posterior predictive distributions.

You are also introduced to non-Bayesian regularization with Lasso regression via the `glmnet` package. If you do not have `glmnet` installed please download it before starting the assignment.

IMPORTANT: code chunks are created for you. Each code chunk has `eval=FALSE` set in the chunk options. You **MUST** change it to be `eval=TRUE` in order for the code chunks to be evaluated when rendering the document.

You are allowed to add as many code chunks as you see fit to answer the questions.

Load packages

This assignment will use packages from the `tidyverse` suite as well as the `coefplot` package. Those packages are imported for you below.

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.6     ✓ dplyr   1.0.8
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## ── Conflicts ─────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(coefplot)
```

```
## Warning: package 'coefplot' was built under R version 4.0.5
```

This assignment also uses the `splines` and `MASS` packages. Both are installed with base `R` and so you do not need to download any additional packages to complete the assignment.

The last question in the assignment uses the `glmnet` package. As stated previously, please download and install `glmnet` if you do not currently have it.

Problem 01

You will fit and compare 6 models of varying complexity using non-Bayesian methods. The unknown parameters will be be estimated by finding their Maximum Likelihood Estimates (MLE). You are allowed to the `lm()` function for this problem.

The data are loaded in the code chunk and a glimpse is shown for you below. There are 2 continuous inputs, `x1` and `x2`, and a continuous response `y`.

```
data_url <- 'https://raw.githubusercontent.com/jyurko/CS_1675_Spring_2022/main/HW/08/hw08_data.csv'

df <- readr::read_csv(data_url, col_names = TRUE)
```

```
## Rows: 100 Columns: 3
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## dbl (3): x1, x2, y
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df %>% glimpse()
```

```
## Rows: 100
## Columns: 3
## $ x1 <dbl> -0.30923281, 0.63127211, -0.68276690, 0.26930562, 0.37252021, 1.296…
## $ x2 <dbl> 0.308779853, -0.547919793, 2.166449412, 1.209703658, 0.785485991, -…
## $ y  <dbl> 0.43636596, 1.37562976, -0.84366730, -0.43080811, 0.77456951, 1.361…
```

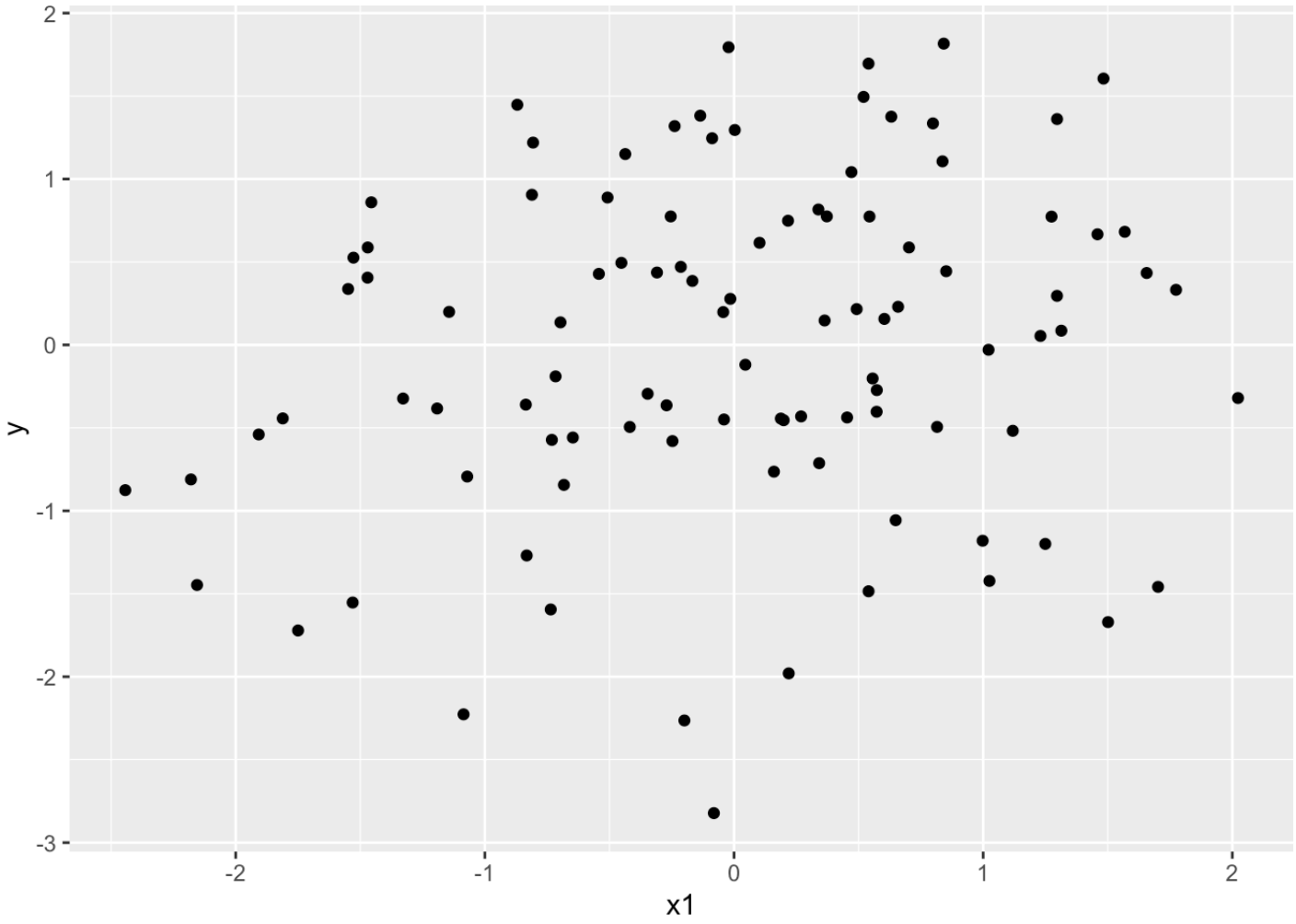
1a)

Create a scatter plot between the response, `y`, and each input using `ggplot()`.

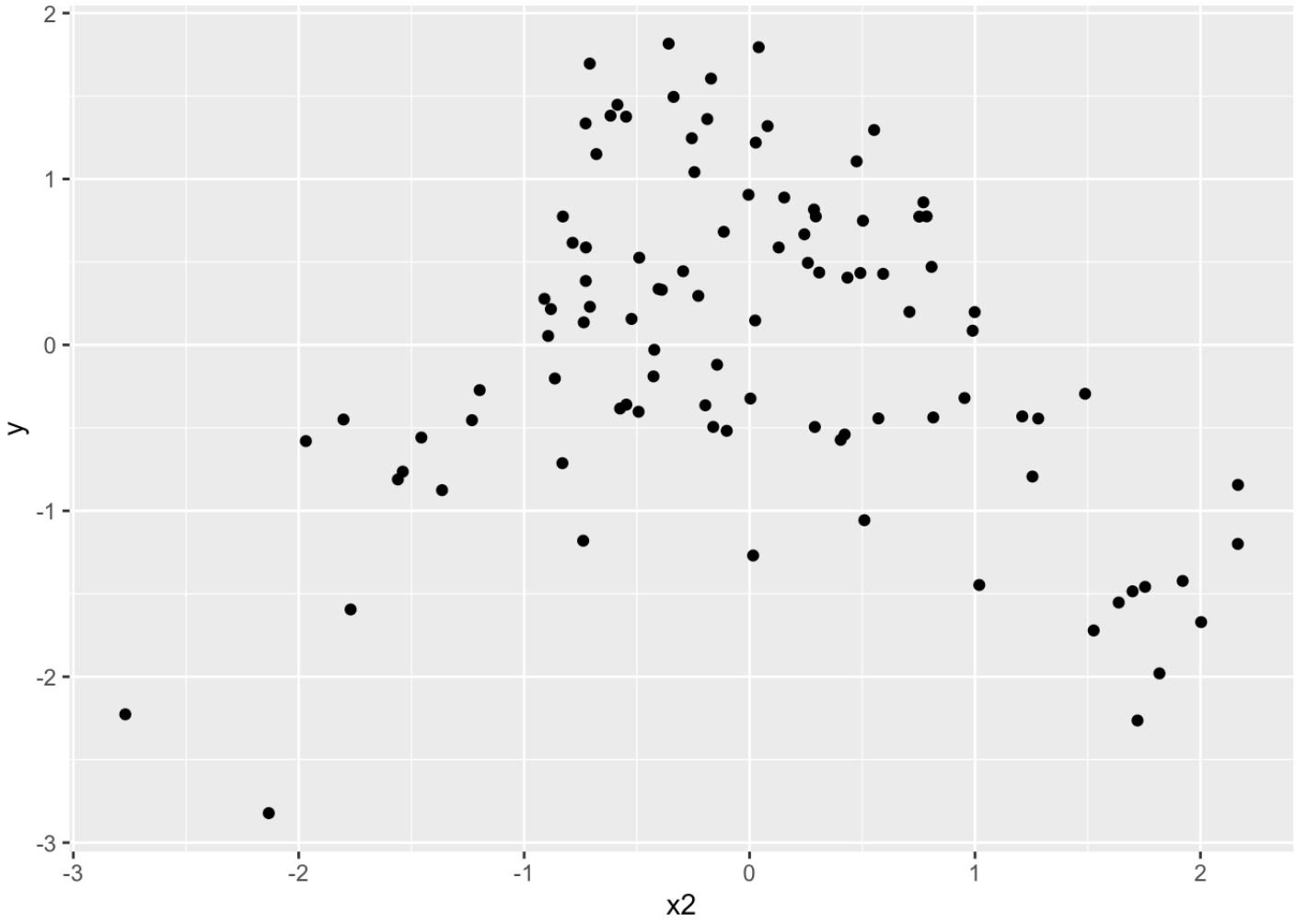
Based on the visualizations, do you think there are trends between either input and the response?

SOLUTION

```
df %>%
  ggplot(x = x, y = y) +
  geom_point(mapping = aes(x = x1, y))
```




```
df %>%
  ggplot(x = x2, y = y) +
  geom_point(mapping = aes(x = x2, y))
```



The relationship between x1 and y seems to be a little unclear at the moment, but with x2 and y there appears to be a quadratic realationship.

1b)

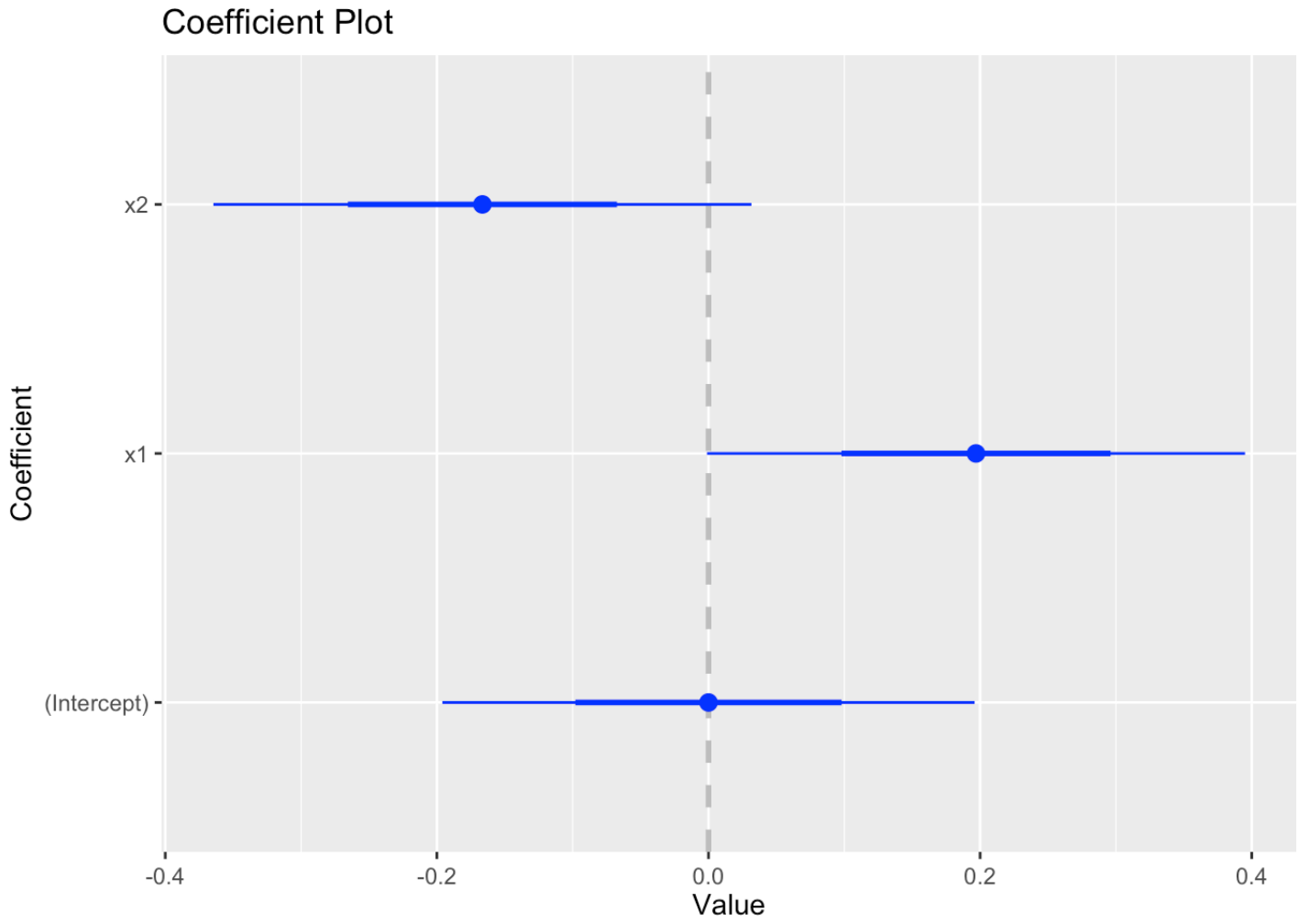
You will fit multiple models of varying complexity in this problem. You will start with *linear additive features*.

Fit a model with linear additive features to predict the response, `y`. Use the formula interface and the `lm()` function to fit the model. Assign the result to the `mod01` object.

Visualize the coefficient summaries with the `coefplot()` function. Are any of the features statistically significant?

SOLUTION

```
mod01 <- lm(y ~ x1 + x2, data = df)
coefplot(mod01)
```



No, none of the features of statistically significant. All confidence intervals contain 0, meaning that it is plausible for that particular coefficient to be 0.

1c)

As discussed in lecture, we can derive features from inputs. We have worked with polynomial features and spline-based features in previous assignments. Features can also be derived as the products between different inputs. A feature calculated as the product of multiple inputs is usually referred to as the **interaction** between those inputs.

In the formula interface, a product of two inputs is denoted by the `:`. And so if we wanted to include just the multiplication of `x1` and `x2` in a model we would type, `x1:x2`. We can then include **main-effect** terms by including the additive features within the formula. Thus, the formula for a model with additive features and the interaction between `x1` and `x2` is:

```
y ~ x1 + x2 + x1:x2
```

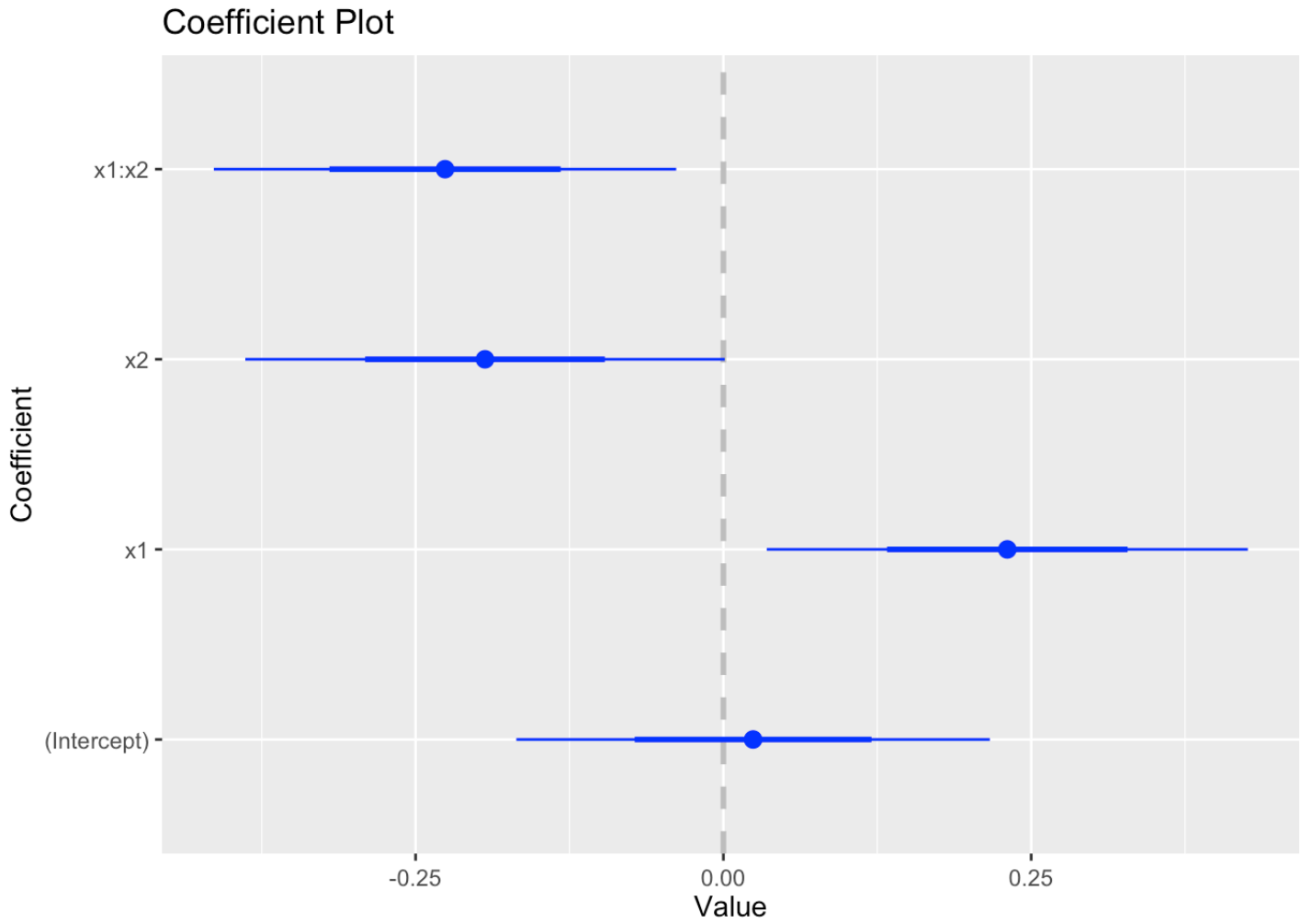
However, the formula interface provides a short-cut to create main effects and interaction features. In the formula interface, the `*` operator will generate all main-effects and all interactions for us.

Fit a model with all main-effect and all-interaction features between `x1` and `x2` using the short-cut `*` operator within the formula interface. Assign the result to the `mod02` object.

Visualize the coefficient summaries with the `coefplot()` function. How many features are present in the model? Are any of the features statistically significant?

SOLUTION

```
mod02 <- lm(y ~ x1 * x2, data = df)
coefplot(mod02)
```



With the interaction term present, there are 4 features (intercept, x1, x2, x1*x2) Now, x1 and the interaction term x1*x2 are considered to be statistically significant because their confidence intervals do not contian 0.

1d)

The `*` operator will interact more than just inputs. We can interact expressions or groups of features together. To interact one group of features by another group of features, we just need to enclose each group by parenthesis, `()`, and separate them by the `*` operator. The line of code below shows how this works with

the `<expression 1>` and `<expression 2>` as place holders for any expression we want to use.

```
(<expression 1>) * (<expression 2>)
```

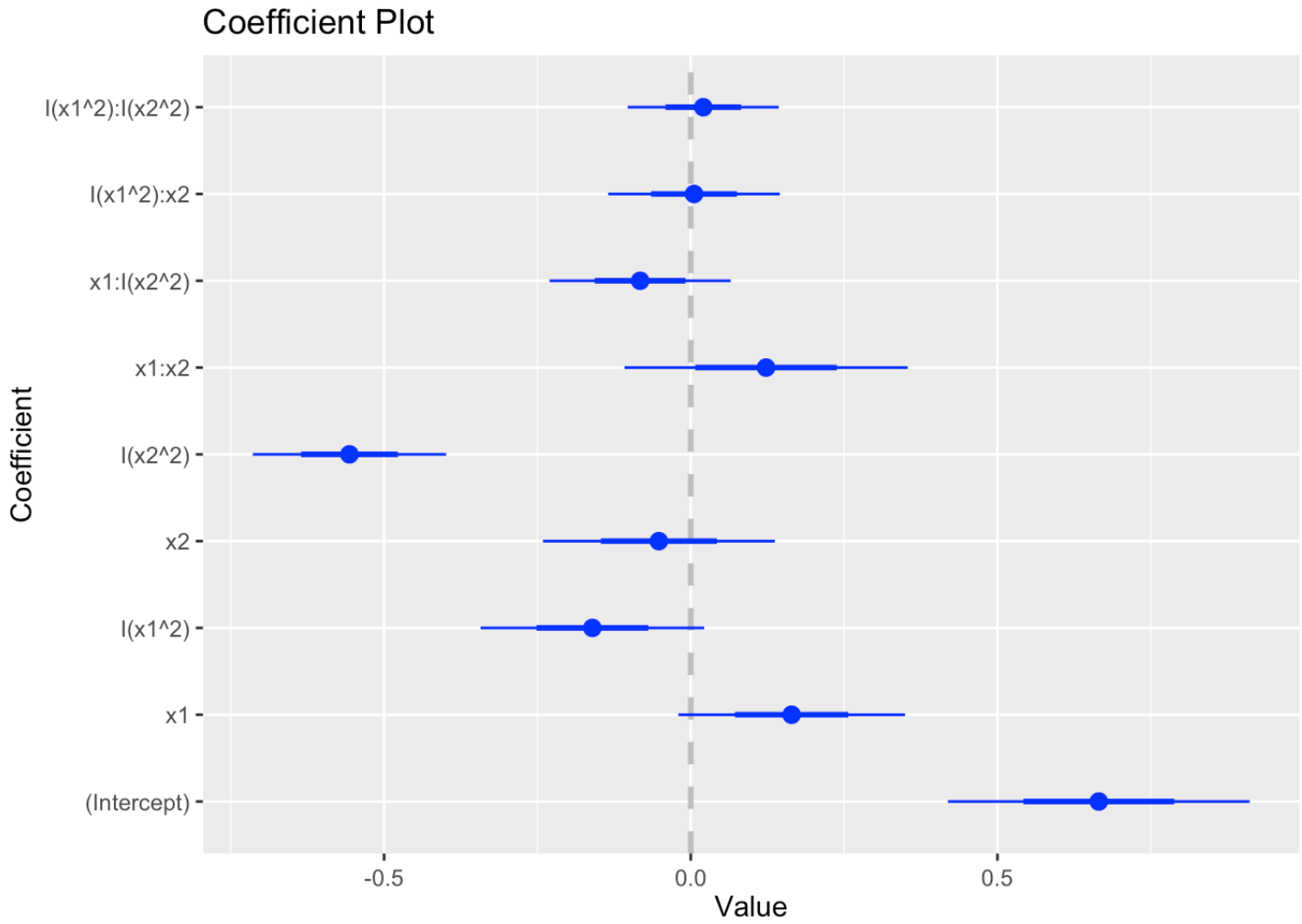
Fit a model which interacts linear and quadratic features from `x1` with linear and quadratic features from `x2`. Assign the result to the `mod03` object.

Visualize the coefficient summaries with the `coefplot()` function. How many features are present in the model? Are any of the features statistically significant?

HINT: Remember to use the `I()` function when typing polynomials in the formula interface.

SOLUTION

```
mod03 <- lm(y ~ (x1 + I(x1^2)) * (x2 + I(x2^2)), data = df)
coefplot(mod03)
```



There are 9 features (intercept + x1,x2, x1^2, x2^2, x1:x2, x1:x2^2, x1^2: x2, $x1^{2:x2}2$) The intercept and x2^2 features are statistically significant.

1e)

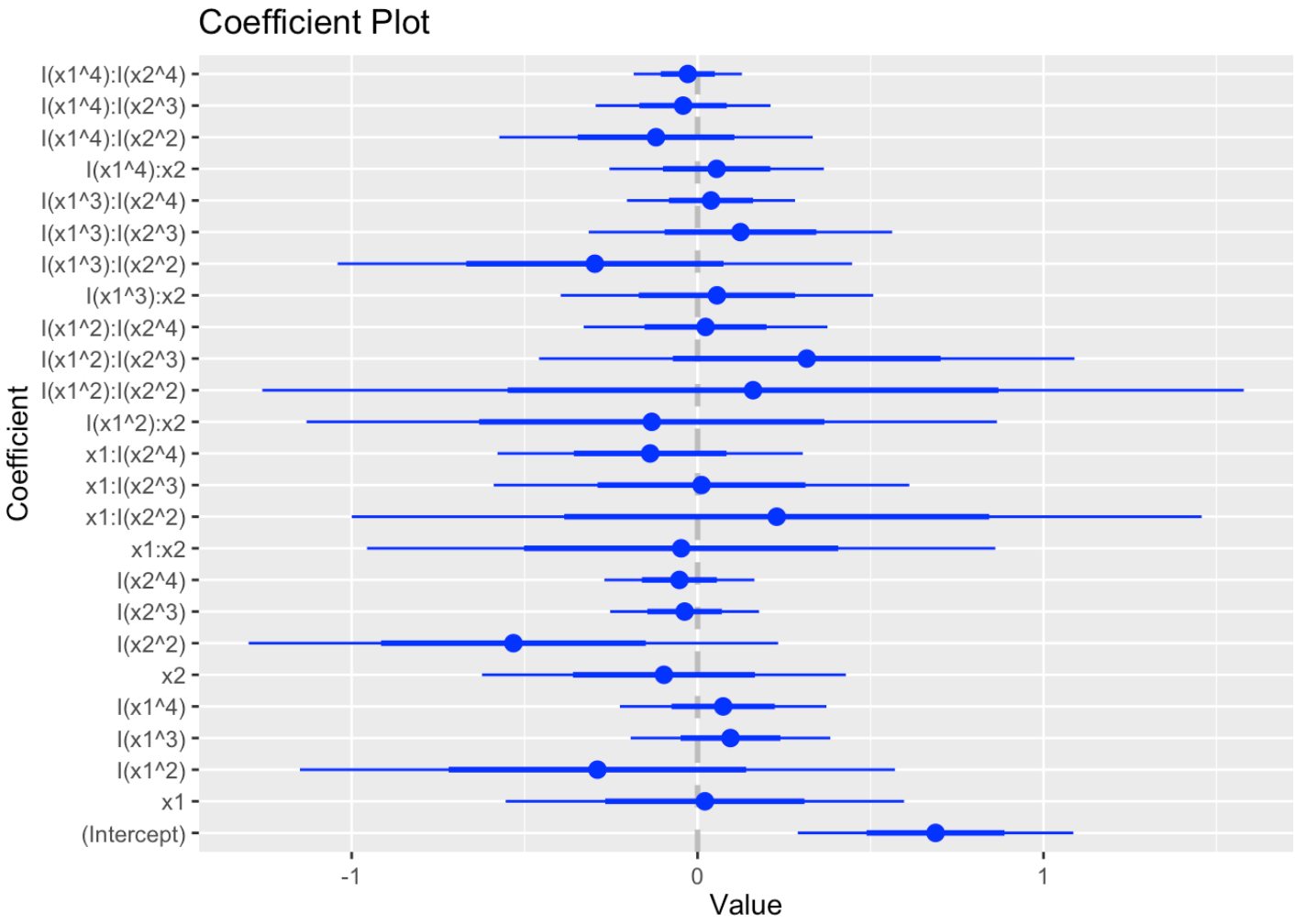
Let's now try a more complicated model.

Fit a model which interacts linear, quadratic, cubic, quartic (4th degree) polynomial features from `x1` with linear, quadratic, cubic, and quartic (4th degree) polynomial features from `x2`. Assign the result to the `mod04` object.

Visualize the coefficient summaries with the `coefplot()` function. Are any of the features statistically significant?

SOLUTION

```
mod04 <- lm(y ~ (x1 + I(x1^2) + I(x1^3) + I(x1^4)) * (x2 + I(x2^2) + I(x2^3) + I(x2^4
)), data = df)
coefplot(mod04)
```



Only the intercept is statistically significant.

1f)

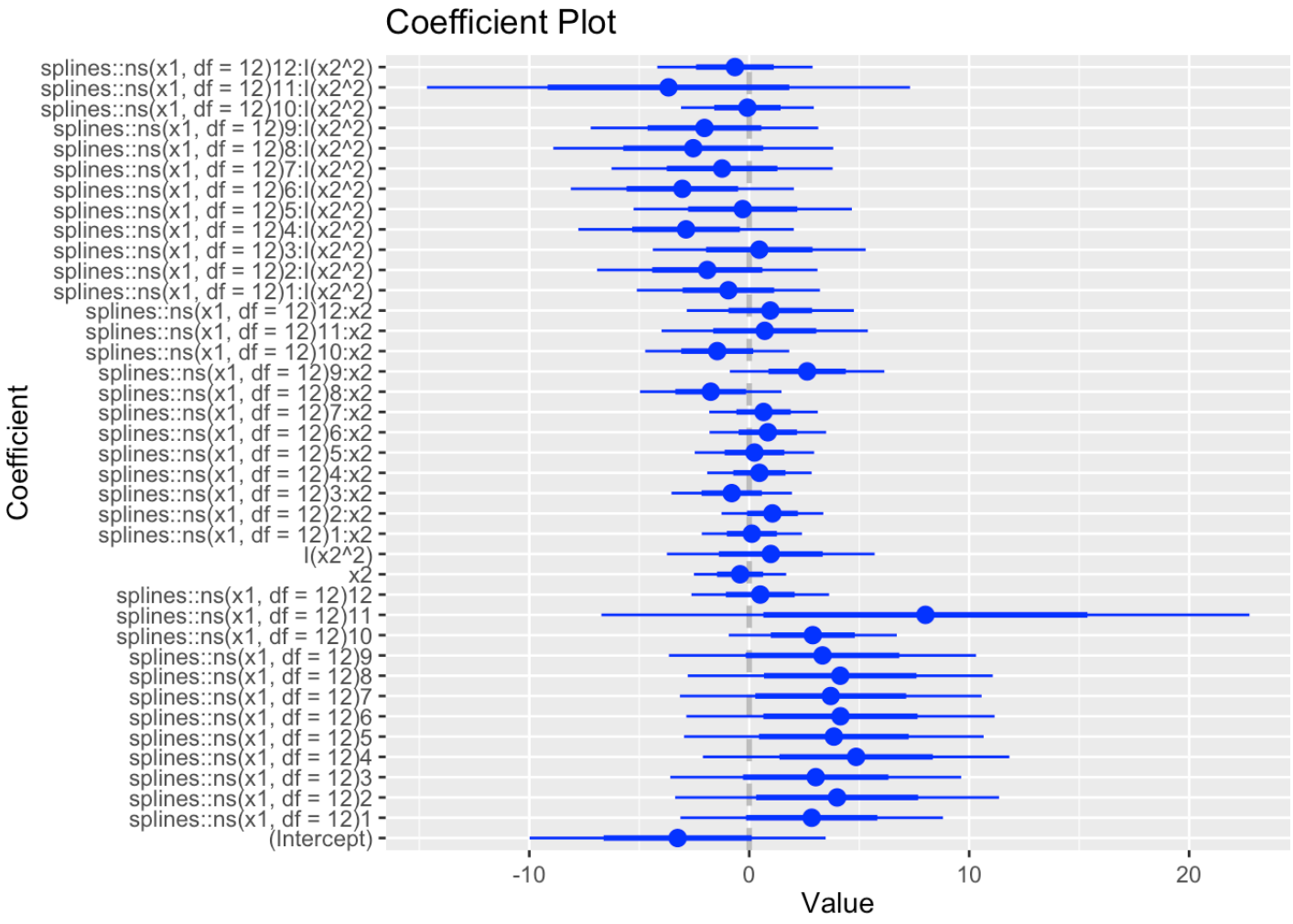
Let's try using spline based features. We will use a high degree-of-freedom natural spline applied to `x1` and interact those features with polynomial features derived from `x2`.

Fit a model which interacts 12 degree-of-freedom natural spline from `x1` with linear and quadrtic polyonomial features from `x2`. Assign the result to `mod05`.

Visualize the coefficient summaries with the `coefplot()` function. Are any of the features statistically significant?

SOLUTION

```
mod05 <- lm(y ~ (splines::ns(x1, df = 12)) * (x2 + I(x2^2)), data = df)
coefplot(mod05)
```



None of the features are significant.

1g)

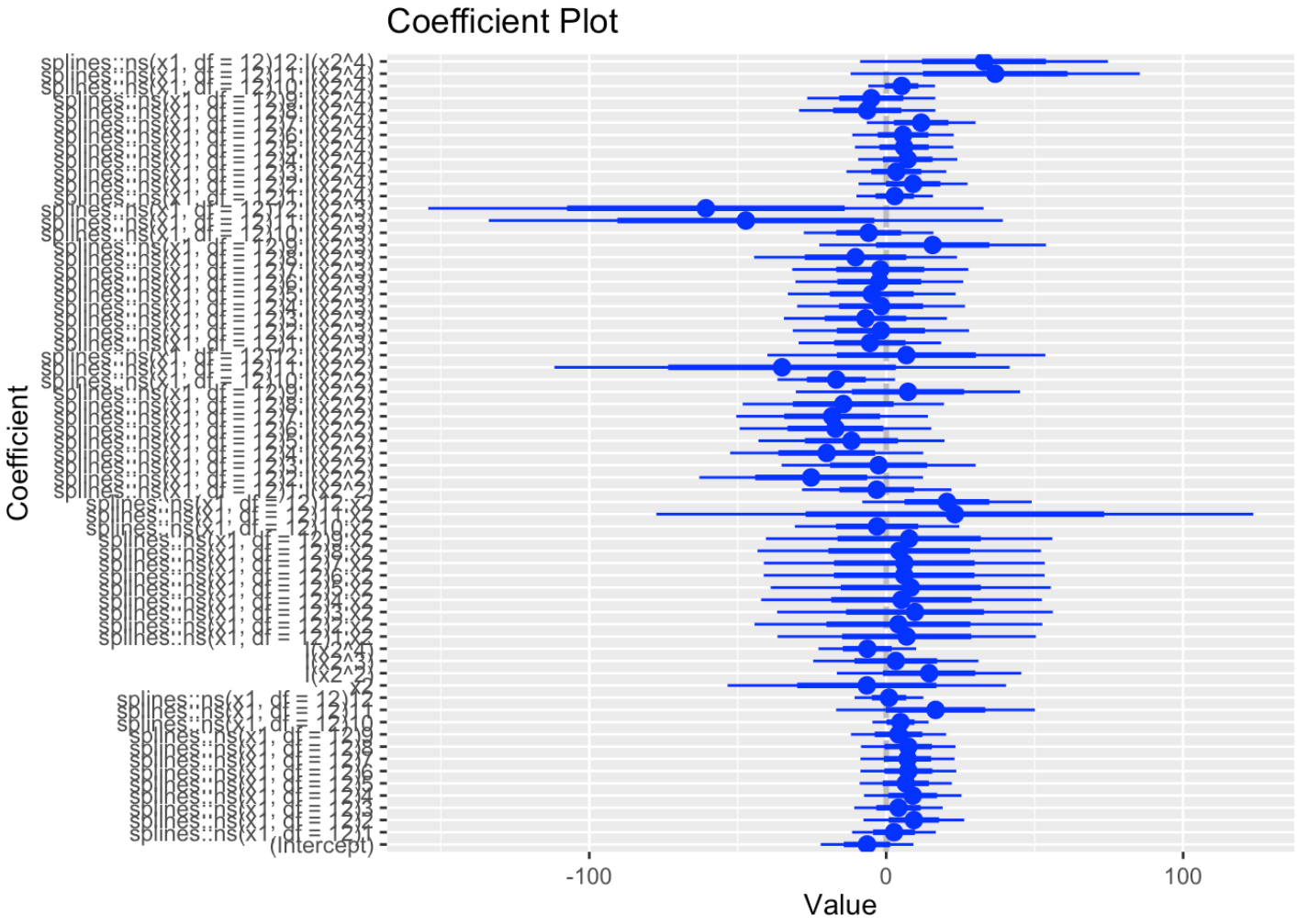
Let's fit one final model.

Fit a model which interacts 12 degree-of-freedom natural spline from `x1` with linear, quadrtic, cubic, and quartic (4th degree) polyonomial features from `x2`. Assign the result to `mod05`.

Visualize the coefficient summaries with the `coefplot()` function. Are any of the features statistically significant?

SOLUTION

```
mod06 <- lm(y ~ (splines::ns(x1, df = 12)) * (x2 + I(x2^2) + I(x2^3) + I(x2^4)), data
= df)
coefplot(mod06)
```



None of the features are statistically significant.

1h)

Now that you have fit multiple models of varying complexity, it is time to identify the best performing model.

Identify the best model considering training set only performance metrics. Which model is best according to R-squared? Which model is best according to AIC? Which model is best according to BIC?

HINT: The `brooom::glance()` function can be helpful here. The `broom` package is installed with `tidyverse` and so you should have it already.

SOLUTION

```
broom::glance(mod01)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1    0.0594        0.0401 0.980      3.07  0.0512     2  -138.  285.  295.
## # … with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
broom::glance(mod02)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.113        0.0853 0.956      4.08 0.00900     3  -135.  281.  294.
## # … with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
broom::glance(mod03)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.squared sigma statistic  p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.547         0.507 0.702      13.7 7.25e-13     8  -102.  224.  250.
## # … with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
broom::glance(mod04)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.squared sigma statistic     p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>       <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.599         0.470 0.728      4.66 0.000000151    24  -95.7  243.  311.
## # … with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
broom::glance(mod05)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.squared sigma statistic    p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>      <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.699         0.512 0.699      3.73 0.00000235    38  -81.3  243.  347.
## # … with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
broom::glance(mod06)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.782         0.383 0.785      1.96  0.0164    64  -65.3  263.  434.
## # … with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

R-squared: Model 6 AIC: Model 3 BIC: Model 3

Problem 02

Now that you know which model is best, let's visualize the predictive trends from the six models. This will help us better understand their performance and behavior.

2a)

You will define a prediction or visualization test grid. This grid will allow you to visualize behavior with respect to `x1` for multiple values of `x2` .

Create a grid of input values where `x1` consists of 101 evenly spaced points between -3.2 and 3.2 and `x2` is 9 evenly spaced points between -3 and 3. The `expand.grid()` function is started for you and the data type conversion is provided to force the result to be a `tibble` .

SOLUTION

```
viz_grid <- expand.grid(x1 = seq(-3.2, 3.2, length.out = 101),
                        x2 = seq(-3,3,length.out = 9) ,
                        KEEP.OUT.ATTRS = FALSE,
                        stringsAsFactors = FALSE) %>%
  as.data.frame() %>% tibble::as_tibble()
```

2b)

You will make predictions for each of the models and visualize their trends. A function, `tidy_predict()`, is created for you which assembles the predicted mean trend, the confidence interval, and the prediction interval into a `tibble` for you. The result include the input values to streamline making the visualizations.

```
tidy_predict <- function(mod, xnew)
{
  pred_df <- predict(mod, xnew, interval = "confidence") %>%
    as.data.frame() %>% tibble::as_tibble() %>%
    dplyr::select(pred = fit, ci_lwr = lwr, ci_upr = upr) %>%
    bind_cols(predict(mod, xnew, interval = 'prediction') %>%
                as.data.frame() %>% tibble::as_tibble() %>%
                dplyr::select(pred_lwr = lwr, pred_upr = upr))

  xnew %>% bind_cols(pred_df)
}
```

The first argument to the `tidy_predict()` function is a `lm()` model object and the second argument is new or test dataframe of inputs. When working with `lm()` and its `predict()` method, the functions will create the test design matrix consistent with the training design basis. It does so via the model object's formula which is contained within the `lm()` model object. The `lm()` object therefore takes care of the heavy lifting for us!

Make predictions with each of the six models you fit in Problem 01 using the visualization grid, `viz_grid`. The predictions should be assigned to the variables `pred_lm_01` through `pred_lm_06` where the number is consistent with the model number fit previously.

SOLUTION

```
pred_lm_01 <- tidy_predict(mod01, viz_grid)

pred_lm_02 <- tidy_predict(mod02, viz_grid)

pred_lm_03 <- tidy_predict(mod03, viz_grid)

pred_lm_04 <- tidy_predict(mod04, viz_grid)

pred_lm_05 <- tidy_predict(mod05, viz_grid)

pred_lm_06 <- tidy_predict(mod06, viz_grid)
```

2c)

You will now visualize the predictive trends and the confidence and prediction intervals for each model. The `pred` column in of each `pred_lm_` objects is the predictive mean trend. The `ci_lwr` and `ci_upr` columns are the lower and upper bounds of the confidence interval, respectively. The `pred_lwr` and `pred_upr`

columns are the lower and upper bounds of the prediction interval, respectively.

You will use `ggplot()` to visualize the predictions. You will use `geom_line()` to visualize the mean trend and `geom_ribbon()` to visualize the uncertainty intervals.

Visualize the predictions of each model on the visualization grid. Pipe the `pred_lm_` object to `ggplot()` and map the `x1` variable to the x-aesthetic. Add three geometric object layers. The first and second layers are each `geom_ribbon()` and the third layer is `geom_line()`. In the `geom_line()` layer map the `pred` variable to the `y` aesthetic. In the first `geom_ribbon()` layer, map `pred_lwr` and `pred_upr` to the `ymin` and `ymax` aesthetics, respectively. Hard code the `fill` to be orange in the first `geom_ribbon()` layer (outside the `aes()` call). In the second `geom_ribbon()` layer, map `ci_lwr` and `ci_upr` to the `ymin` and `ymax` aesthetics, respectively. Hard code the `fill` to be `grey` in the second `geom_ribbon()` layer (outside the `aes()` call). Include `facet_wrap()` with the facets with controlled by the `x2` variable.

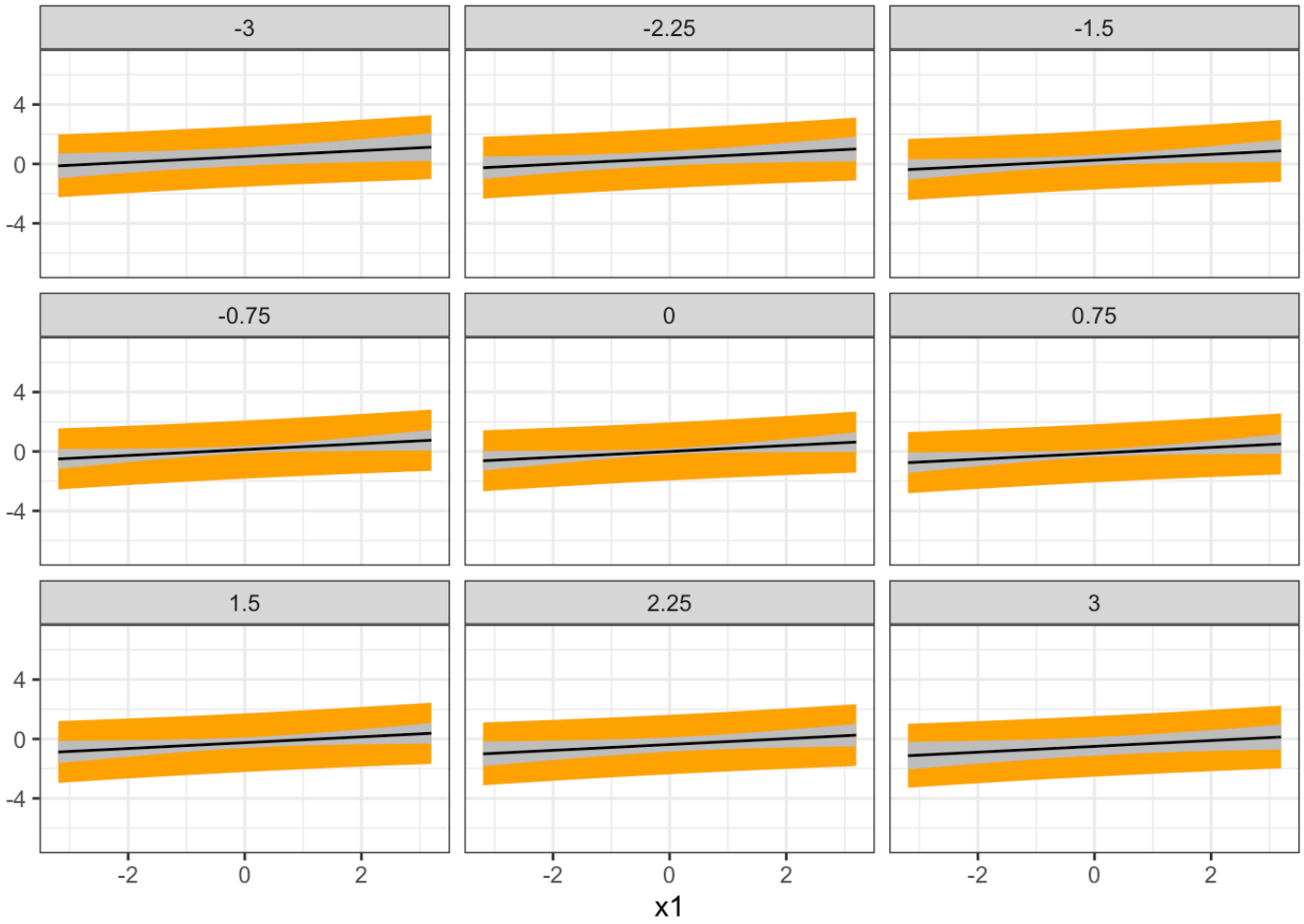
To help compare the visualizations across models include a `coord_cartesian()` layer with the `ylim` argument set to `c(-7,7)`.

Each model's prediction visualization should be created in a separate code chunk.

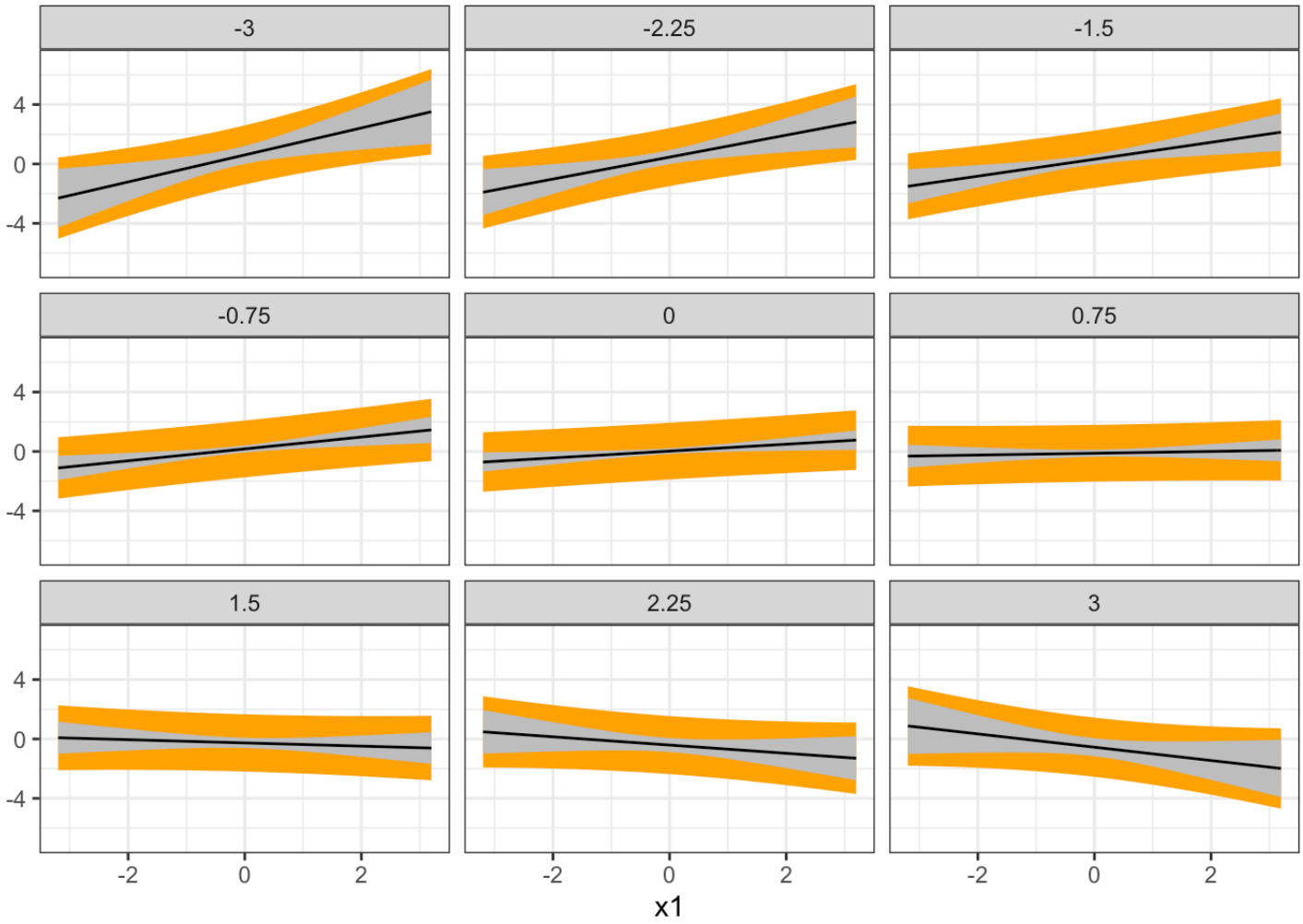
SOLUTION

Create separate code chunks for each visualization.

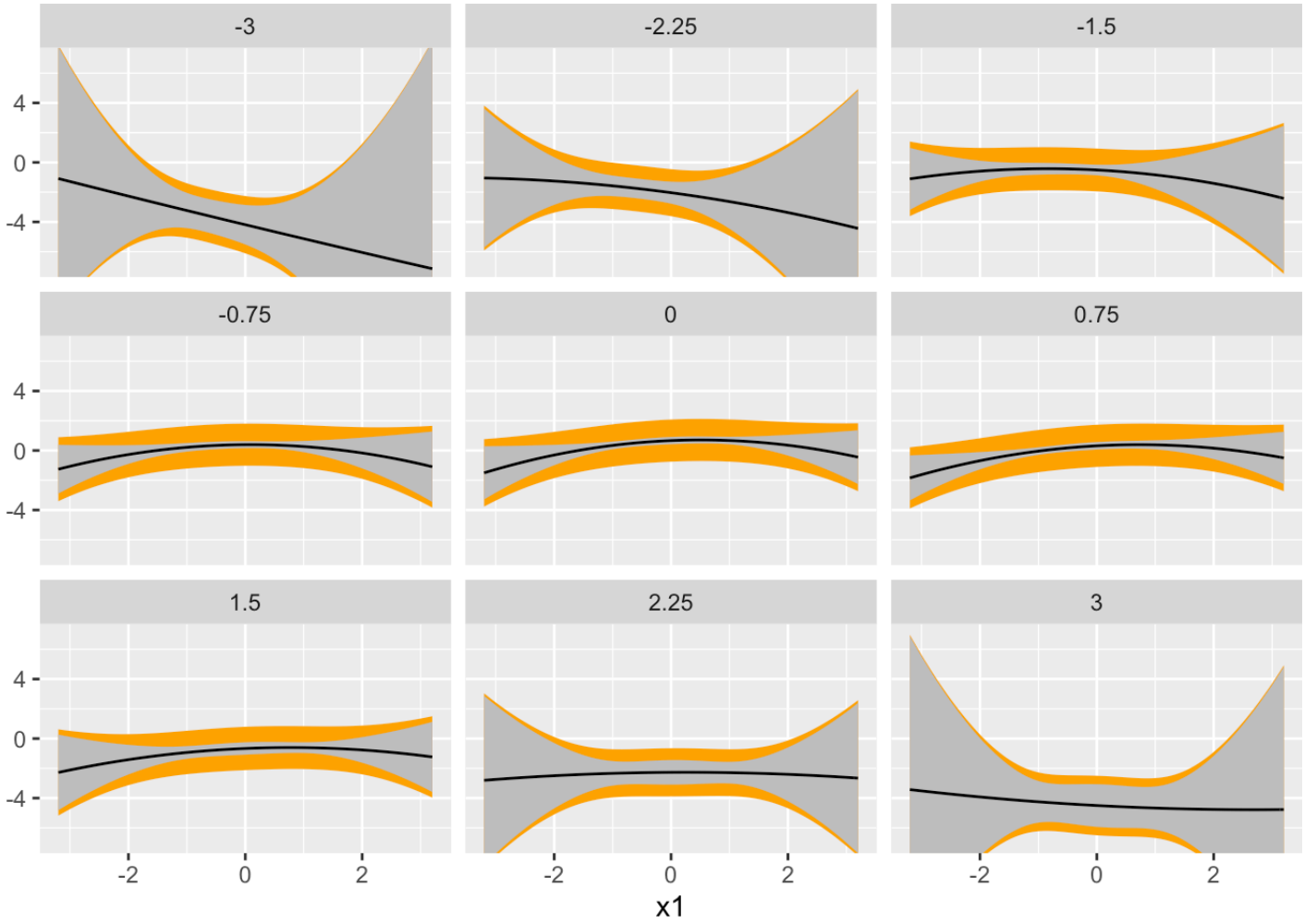
```
pred_lm_01 %>% ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = pred_lwr, ymax = pred_upr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = ci_lwr, ymax = ci_upr), fill = 'grey') +
  geom_line(y = pred_lm_01$pred)+
  facet_wrap(facets = pred_lm_01$x2) +
  theme_bw() +
  coord_cartesian(ylim = c(-7,7))
```



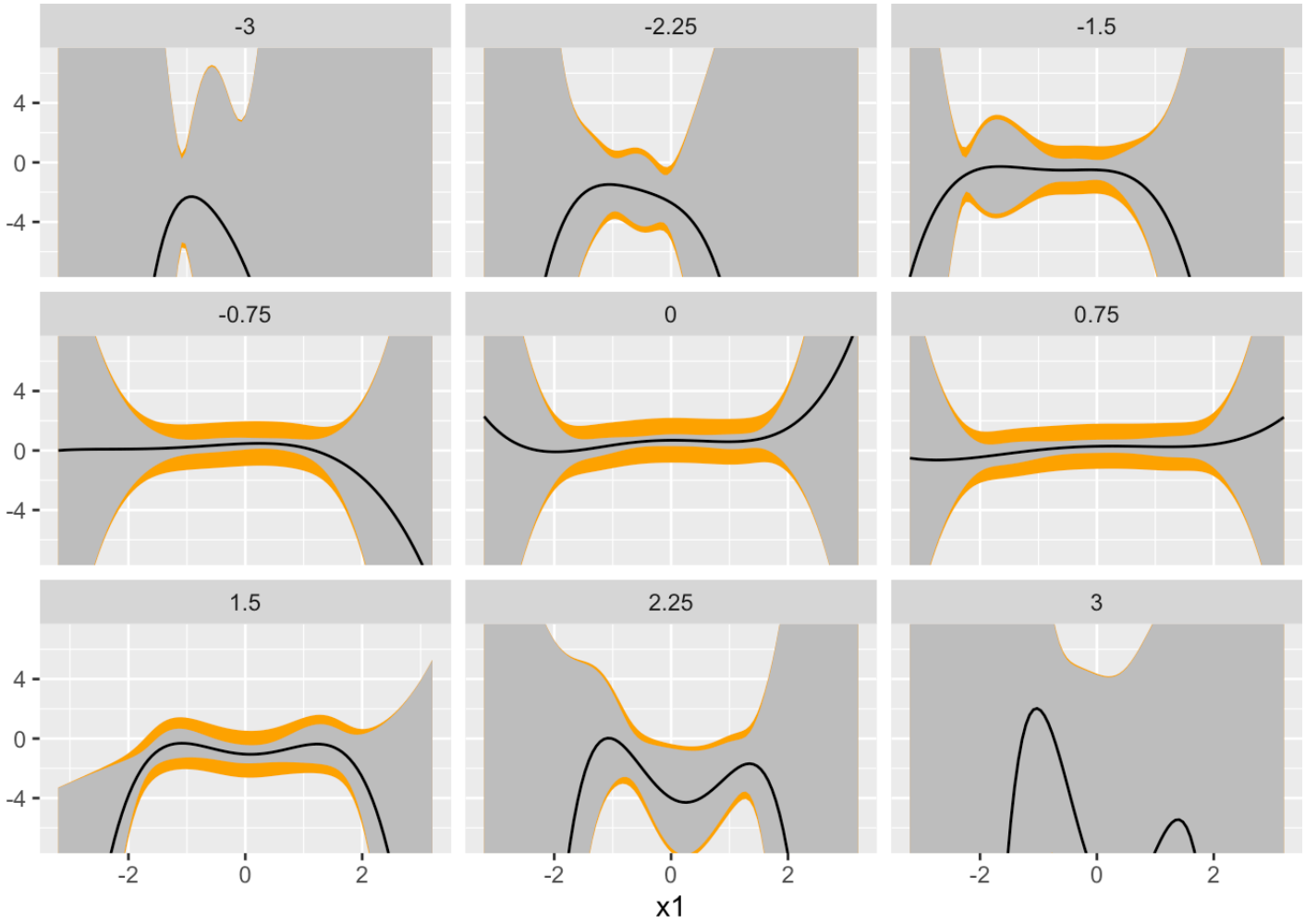

```
pred_lm_02 %>% ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = pred_lwr, ymax = pred_upr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = ci_lwr, ymax = ci_upr), fill = 'grey') +
  geom_line(y = pred_lm_02$pred)+
  facet_wrap(facets = pred_lm_02$x2)+
  theme_bw() +
  coord_cartesian(ylim = c(-7,7))
```



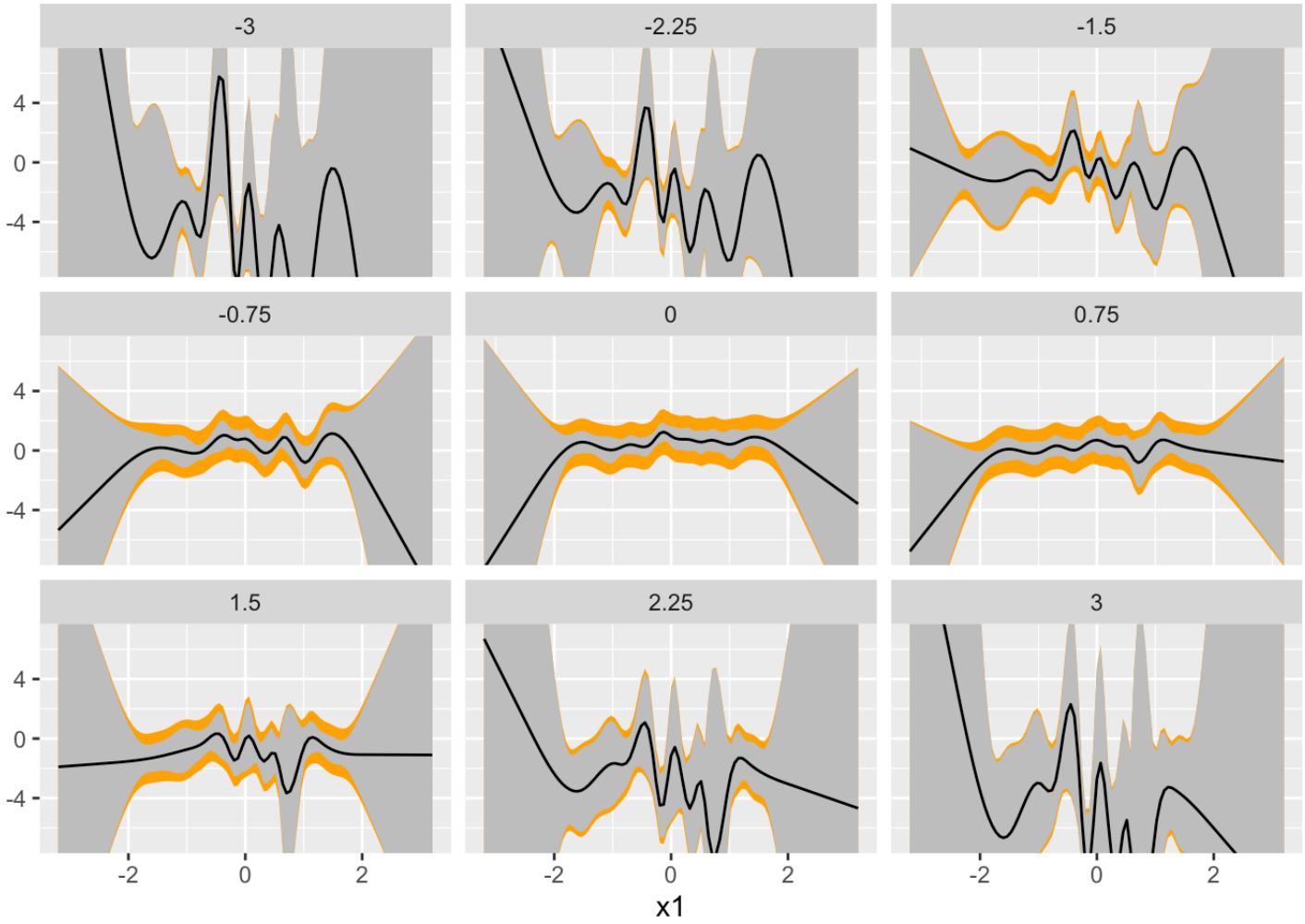

```
pred_lm_03 %>% ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = pred_lwr, ymax = pred_upr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = ci_lwr, ymax = ci_upr), fill = 'grey') +
  geom_line(y = pred_lm_03$pred)+
  facet_wrap(facets = pred_lm_03$x2)+
  coord_cartesian(ylim = c(-7,7))
```



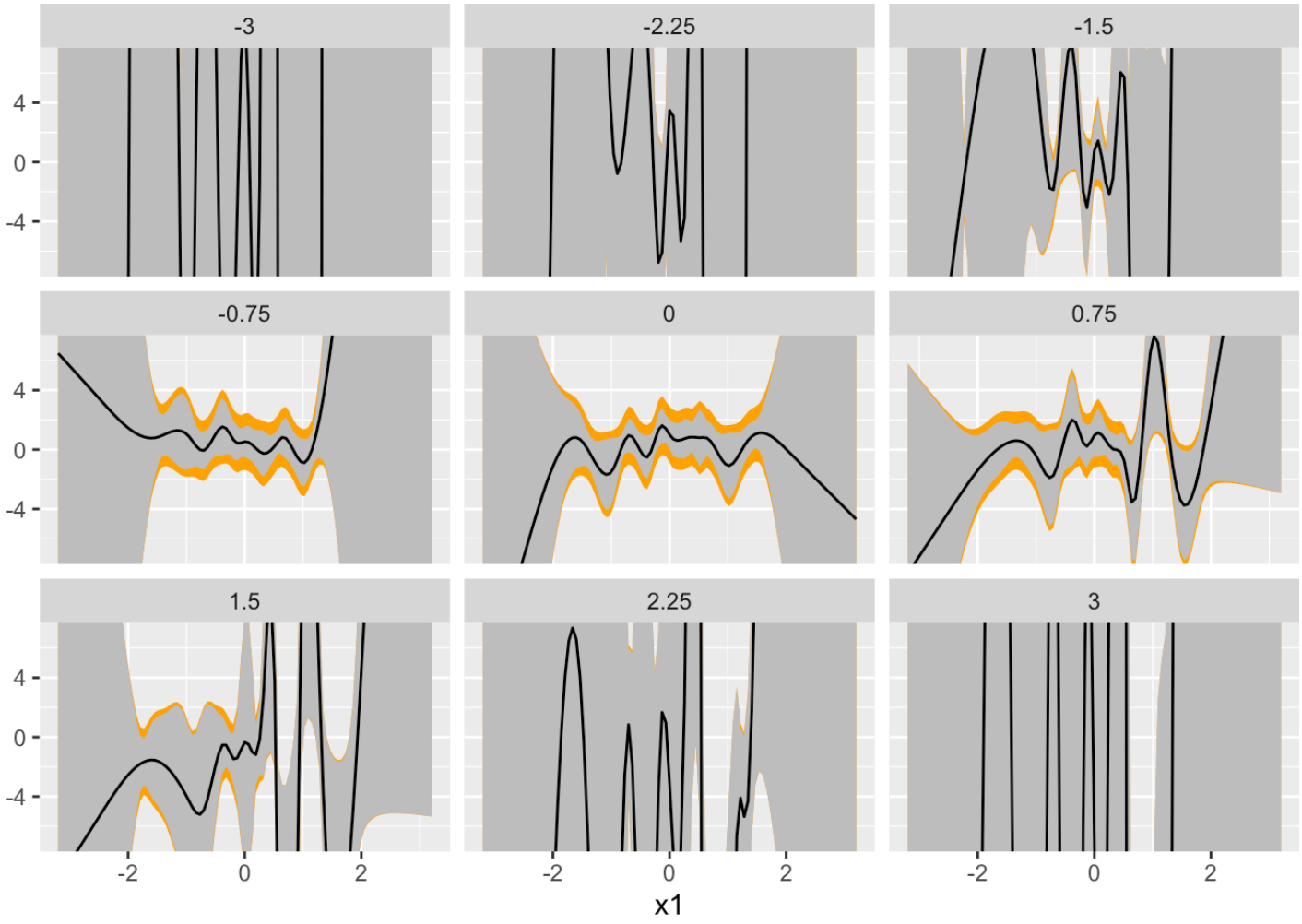

```
pred_lm_04 %>% ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = pred_lwr, ymax = pred_upr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = ci_lwr, ymax = ci_upr), fill = 'grey') +
  geom_line(y = pred_lm_04$pred)+
  facet_wrap(facets = pred_lm_04$x2)+
  coord_cartesian(ylim = c(-7,7))
```




```
pred_lm_05 %>% ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = pred_lwr, ymax = pred_upr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = ci_lwr, ymax = ci_upr), fill = 'grey') +
  geom_line(y = pred_lm_05$pred)+
  facet_wrap(facets = pred_lm_05$x2)+
  coord_cartesian(ylim =c(-7,7))
```




```
pred_lm_06 %>% ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = pred_lwr, ymax = pred_upr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = ci_lwr, ymax = ci_upr), fill = 'grey') +
  geom_line(y = pred_lm_06$pred)+
  facet_wrap(facets = pred_lm_06$x2)+
  coord_cartesian(ylim = c(-7,7))
```



2d)

Do you feel the predictions are consistent with the model performance rankings based on AIC/BIC? What is the defining characteristic of the models considered to be the worst by AIC/BIC?

SOLUTION

The defining characteristic of the models considered to be the worst by AIC/BIC are that they have the most features. AIC and BIC penalize the models for adding more features.

Problem 03

Now that you have fit non-Bayesian linear models with maximum likelihood estimation, it is time to use Bayesian models to understand the influence of the prior on the model behavior.

Regardless of your answers in Problem 02 you will only work with model 3 and model 6 in this problem.

3a)

You will perform the Bayesian analysis using the Laplace Approximation just as you did in the previous assignment. You will define the log-posterior function just as you did in the previous assignment and so before doing so you must create the list of required information. This list will include the observed response, the design matrix, and the prior specification. You will use independent Gaussian priors on the regression parameters with a shared prior mean and shared prior standard deviation. You will use an Exponential prior on the unknown likelihood noise (the σ parameter).

Complete the two code chunks below. In the first, create the design matrix following `mod03` 's formula, and assign the object to the `X03` variable. Complete the `info_03_weak` list by assigning the response to `yobs` and the design matrix to `design_matrix` . Specify the shared prior mean, `mu_beta` , to be 0, the shared prior standard deviation, `tau_beta` , as 50, and the rate parameter on the noise, `sigma_rate` , to be 1.

Complete the second code chunk with the same prior specification. The second code chunk however requires that you create the design matrix associated with `mod06` 's formula and assign the object to the `X06` variable. Assign `X06` to the `design_matrix` field of the `info_06_weak` list.

SOLUTION

```
X03 <- model.matrix(lm(df$y ~ (df$x1 + I(df$x1^2)) * (df$x2 + I(df$x2^2))))

info_03_weak <- list(
  yobs = df$y,
  design_matrix = X03,
  mu_beta = 0,
  tau_beta = 50,
  sigma_rate = 1
)
```

```
X06 <- model.matrix(lm(df$y ~ (splines::ns(df$x1, df = 12)) * (df$x2 + I(df$x2^2) + I
(df$x2^3) + I(df$x2^4))))

info_06_weak <- list(
  yobs = df$y,
  design_matrix = X06,
  mu_beta = 0,
  tau_beta = 50,
  sigma_rate = 1
)
```

3b)

You will now define the log-posterior function `lm_logpost()` . You will continue to use the log-transformation on σ, and so you will actually define the log-posterior in terms of the mean trend $\boldsymbol{\beta}$-parameters and the unbounded noise parameter, $\varphi = \log[\sigma]$.

The comments in the code chunk below tell you what you need to fill in. The unknown parameters to learn are contained within the first input argument, `unknowns`. You will assume that the unknown $\boldsymbol{\beta}$-parameters are listed before the unknown φ parameter in the `unknowns` vector. You must specify the number of $\boldsymbol{\beta}$ parameters programmatically to allow scaling up your function to an arbitrary number of unknowns. You will assume that all variables contained in the `my_info` list (the second argument to `lm_logpost()`) are the same fields in the `info_03_weak` list you defined in Problem 3a).

Define the log-posterior function by completing the code chunk below. You must calculate the mean trend, `mu`, using matrix math between the design matrix and the unknown $\boldsymbol{\beta}$ column vector.

HINT: This function should look very famaliar…

SOLUTION

```
lm_logpost <- function(unknowns, my_info)
{
  # specify the number of unknown beta parameters
  length_beta <- ncol(my_info$design_matrix)

  # extract the beta parameters from the `unknowns` vector
  beta_v <- unknowns[1:length_beta]

  # extract the unbounded noise parameter, varphi
  lik_varphi <- unknowns[length_beta + 1]

  # back-transform from varphi to sigma
  lik_sigma <- exp(lik_varphi)

  # extract design matrix
  X <- my_info$design_matrix

  # calculate the linear predictor
  mu <- as.vector(X %*% as.matrix(beta_v))

  # evaluate the log-likelihood
  log_lik <- sum(dnorm(x = my_info$yobs,
                       mean = mu,
                       sd = lik_sigma,
                       log = TRUE))

  # evaluate the log-prior
  log_prior_beta <- sum(dnorm(x = beta_v,
                              mean = my_info$mu_beta,
                              sd = my_info$tau_beta,
                              log = TRUE))

  log_prior_sigma <- dexp(x = lik_sigma,
                          rate = my_info$sigma_rate,
                          log = TRUE)

  # add the mean trend prior and noise prior together
  log_prior <- log_prior_beta + log_prior_sigma

  # account for the transformation
  log_derive_adjust <- lik_varphi

  # sum together
  (log_lik + log_prior + log_derive_adjust)
}
```

3c)

The `my_laplace()` function is defined for you in the code chunk below. This function executes the laplace approximation and returns the object consisting of the posterior mode, posterior covariance matrix, and the log-evidence.

```
my_laplace <- function(start_guess, logpost_func, ...)
{
  # code adapted from the `LearnBayes`` function `laplace()`
  fit <- optim(start_guess,
               logpost_func,
               gr = NULL,
               ...,
               method = "BFGS",
               hessian = TRUE,
               control = list(fnscale = -1, maxit = 1001))

  mode <- fit$par
  post_var_matrix <- -solve(fit$hessian)
  p <- length(mode)
  int <- p/2 * log(2 * pi) + 0.5 * log(det(post_var_matrix)) + logpost_func(mode, ...
)
  # package all of the results into a list
  list(mode = mode,
       var_matrix = post_var_matrix,
       log_evidence = int,
       converge = ifelse(fit$convergence == 0,
                         "YES",
                         "NO"),
       iter_counts = as.numeric(fit$counts[1]))
}
```

Execute the Laplace Approximation for the model 3 formulation and the model 6 formulation. Assign the model 3 result to the `laplace_03_weak` object, and assign the model 6 result to the `laplace_06_weak` object. Check that the optimization scheme converged.

SOLUTION

```
### add more code chunks if you like
laplace_03_weak <- my_laplace(rep(0,ncol(X03) + 1), lm_logpost, info_03_weak)
laplace_03_weak$converge
```

```
## [1] "YES"
```

```
laplace_06_weak <- my_laplace(rep(0,ncol(X06) + 1), lm_logpost, info_06_weak)
laplace_06_weak$converge
```

```
## [1] "YES"
```

3d)

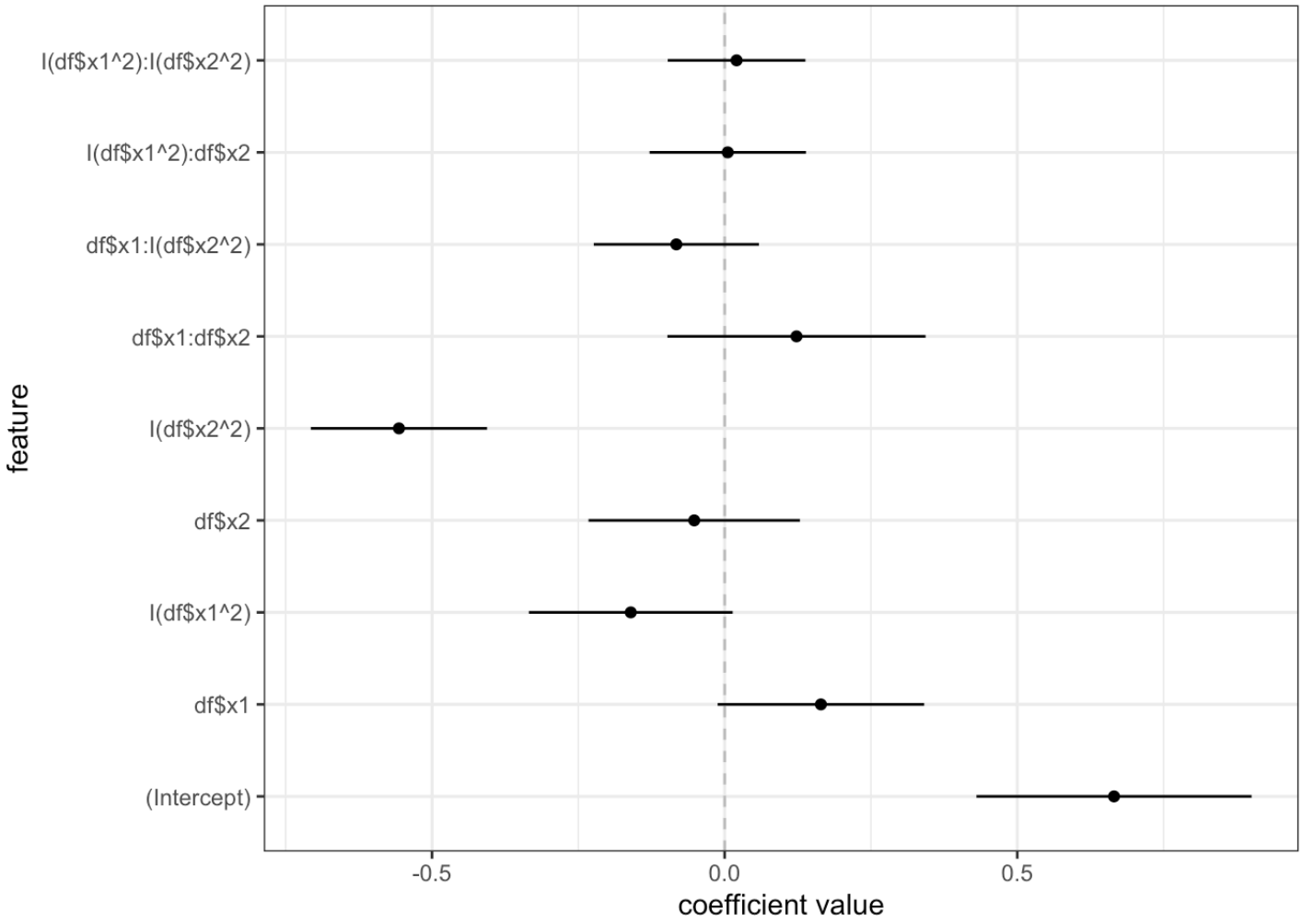
A function is defined for you in the code chunk below. This function creates a coefficient summary plot in the style of the `coefplot()` function, but uses the Bayesian results from the Laplace Approximation. The first argument is the vector of posterior means, and the second argument is the vector of posterior standard deviations. The third argument is the name of the feature associated with each coefficient.

```
viz_post_coefs <- function(post_means, post_sds, xnames)
{
  tibble::tibble(
    mu = post_means,
    sd = post_sds,
    x = xnames
  ) %>%
    mutate(x = factor(x, levels = xnames)) %>%
    ggplot(mapping = aes(x = x)) +
    geom_hline(yintercept = 0, color = 'grey', linetype = 'dashed') +
    geom_point(mapping = aes(y = mu)) +
    geom_linerange(mapping = aes(ymin = mu - 2 * sd,
                                 ymax = mu + 2 * sd,
                                 group = x)) +
    labs(x = 'feature', y = 'coefficient value') +
    coord_flip() +
    theme_bw()
}
```

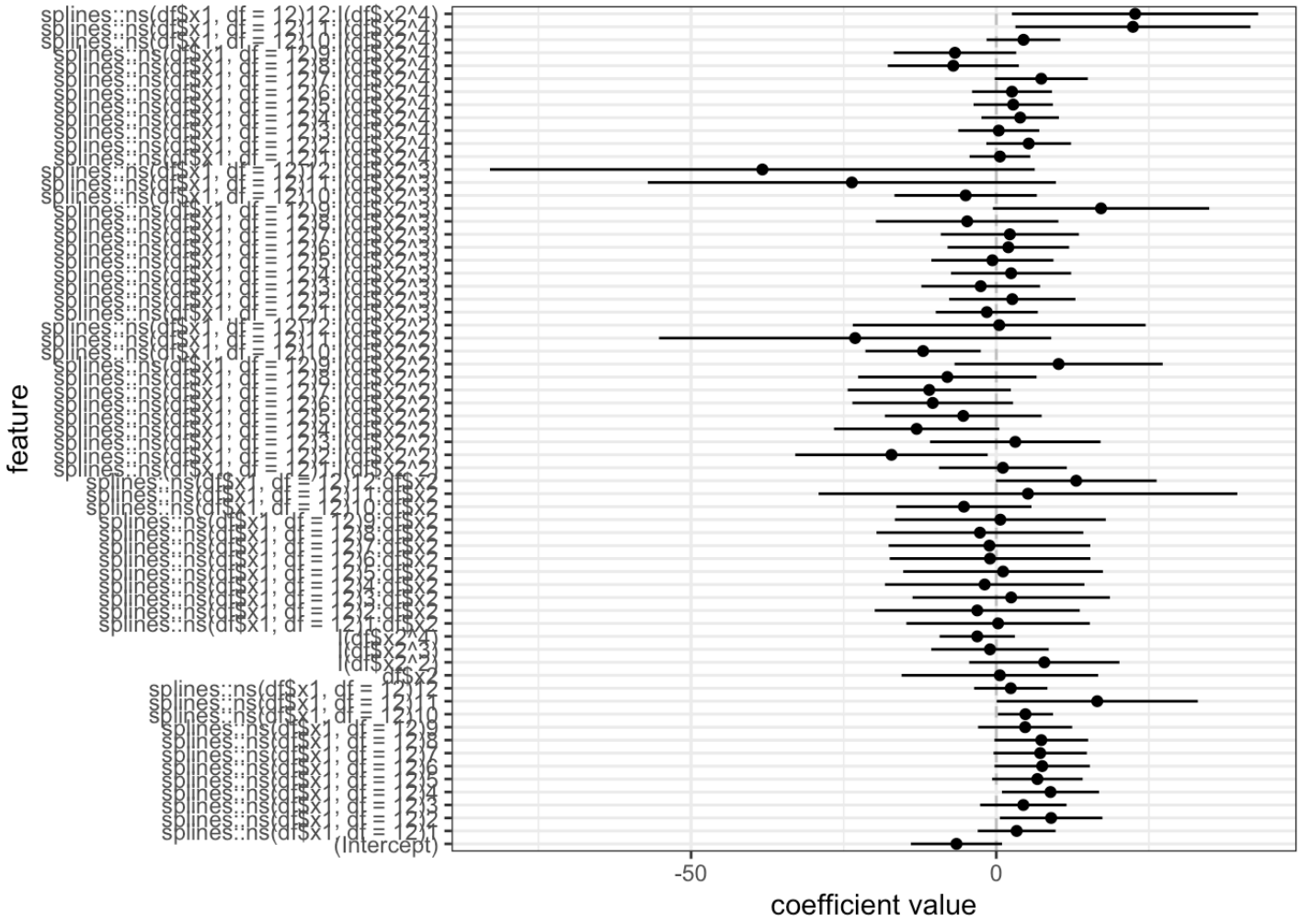
Create the posterior summary visualization figure for model 3 and model 6. You must provide the posterior means and standard deviations of the regression coefficients (the β parameters). Do NOT include the φ parameter. The feature names associated with the coefficients can be extracted from the design matrix using the `colnames()` function.

SOLUTION

```
### make the posterior coefficient visualization for model 3
sd <- c(sqrt(diag(laplace_03_weak$var_matrix)))
sd <- sd[-length(sd)]
viz_post_coefs(laplace_03_weak$mode[-length(laplace_03_weak$mode)],sd, colnames(X03))
```




```
### make the posterior coefficient visualization for model 6
sd <- c(sqrt(diag(laplace_06_weak$var_matrix)))
sd <- sd[-length(sd)]
viz_post_coefs(laplace_06_weak$mode[-length(laplace_06_weak$mode)],sd, colnames(X06))
```



3e)

Use the Bayes Factor to identify the better of the models.

SOLUTION

```
all_models <- list(laplace_03_weak, laplace_06_weak)
model_evidence <- purrr::map_dbl(all_models, "log_evidence")

model_weights <- model_evidence[1] / model_evidence[2]
model_weights
```

```
## [1] 0.448033
```

According to Bayes Factor, model 6 is better than model 3, since bayes factor is less than 1, meaning model 6 has more weight than model 3.

3f)

You fit the Bayesian models assuming a diffuse or *weak* prior. Let's now try a more informative or *strong* prior by reducing the prior standard deviation on the regression coefficients from 50 to 1. The prior mean will still be zero.

Complete the first code chunk below, which defines the list of required information for both the model 3 and model 6 formulations using the strong prior on the regression coefficients. All other information, data and the σ prior, are the same as before.

Run the Laplace Approximation using the strong prior for both the model 3 and model 6 formulations. Assign the results to `laplace_03_strong` and `laplace_06_strong`.

Confirm that the optimizations converged for both laplace approximation results.

SOLUTION

Define the lists of required information for the strong prior.

```
info_03_strong <- list(
  yobs = df$y,
  design_matrix = X03,
  mu_beta = 0,
  tau_beta = 1,
  sigma_rate = 1
)

info_06_strong <- list(
  yobs = df$y,
  design_matrix = X06,
  mu_beta = 0,
  tau_beta = 1,
  sigma_rate = 1
)
```

Execute the Laplace Approximation.

```
### add more code chunks if you like
laplace_03_strong <- my_laplace(rep(0,ncol(X03) + 1), lm_logpost, info_03_strong)
laplace_03_strong$converge
```

```
## [1] "YES"
```

```
laplace_06_strong <- my_laplace(rep(0,ncol(X06) + 1), lm_logpost, info_06_strong)
laplace_06_strong$converge
```

```
## [1] "YES"
```

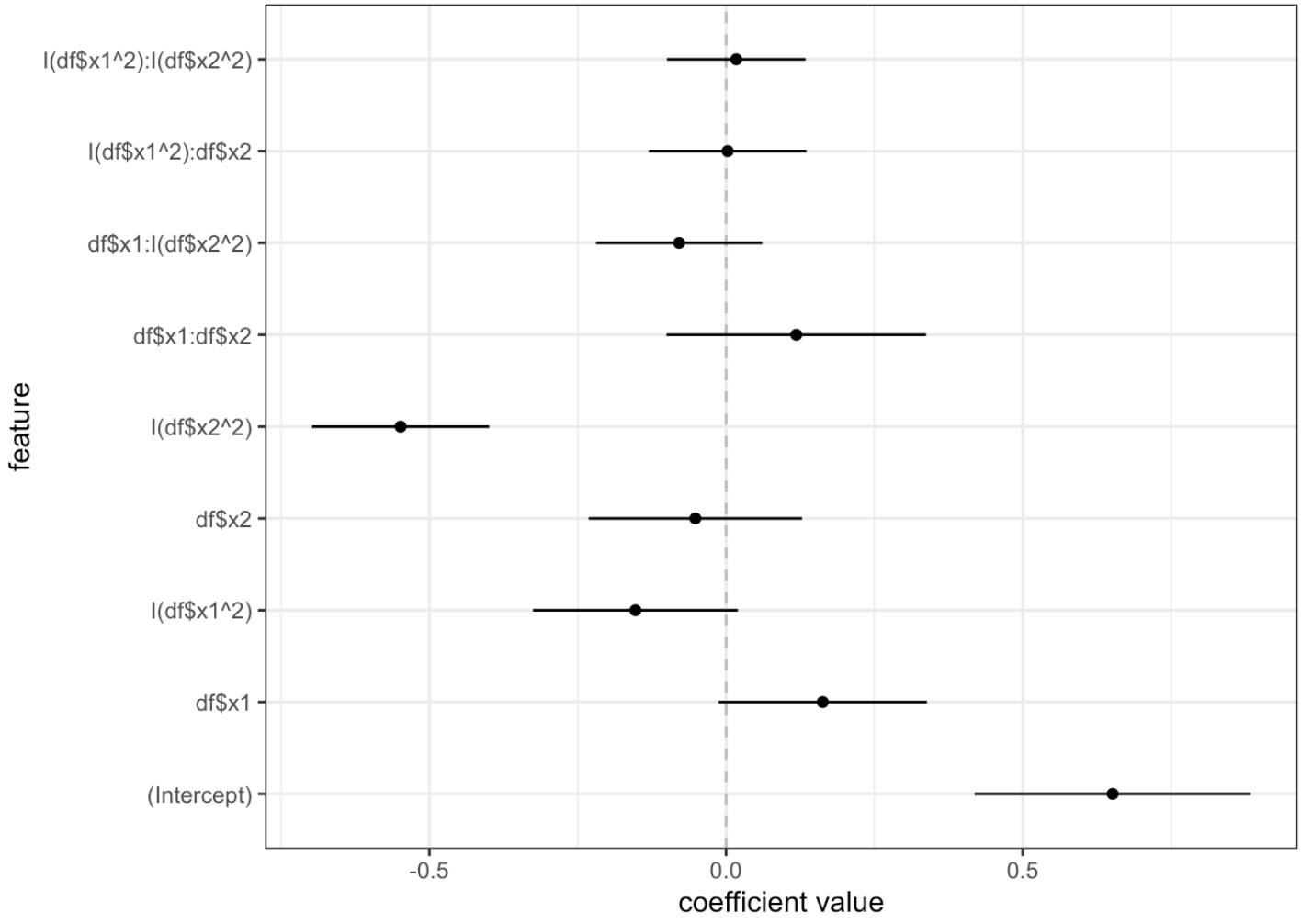
3g)

Use the `viz_post_coefs()` function to visualize the posterior coefficient summaries for model 3 and model 6, based on the strong prior specification.

SOLUTION

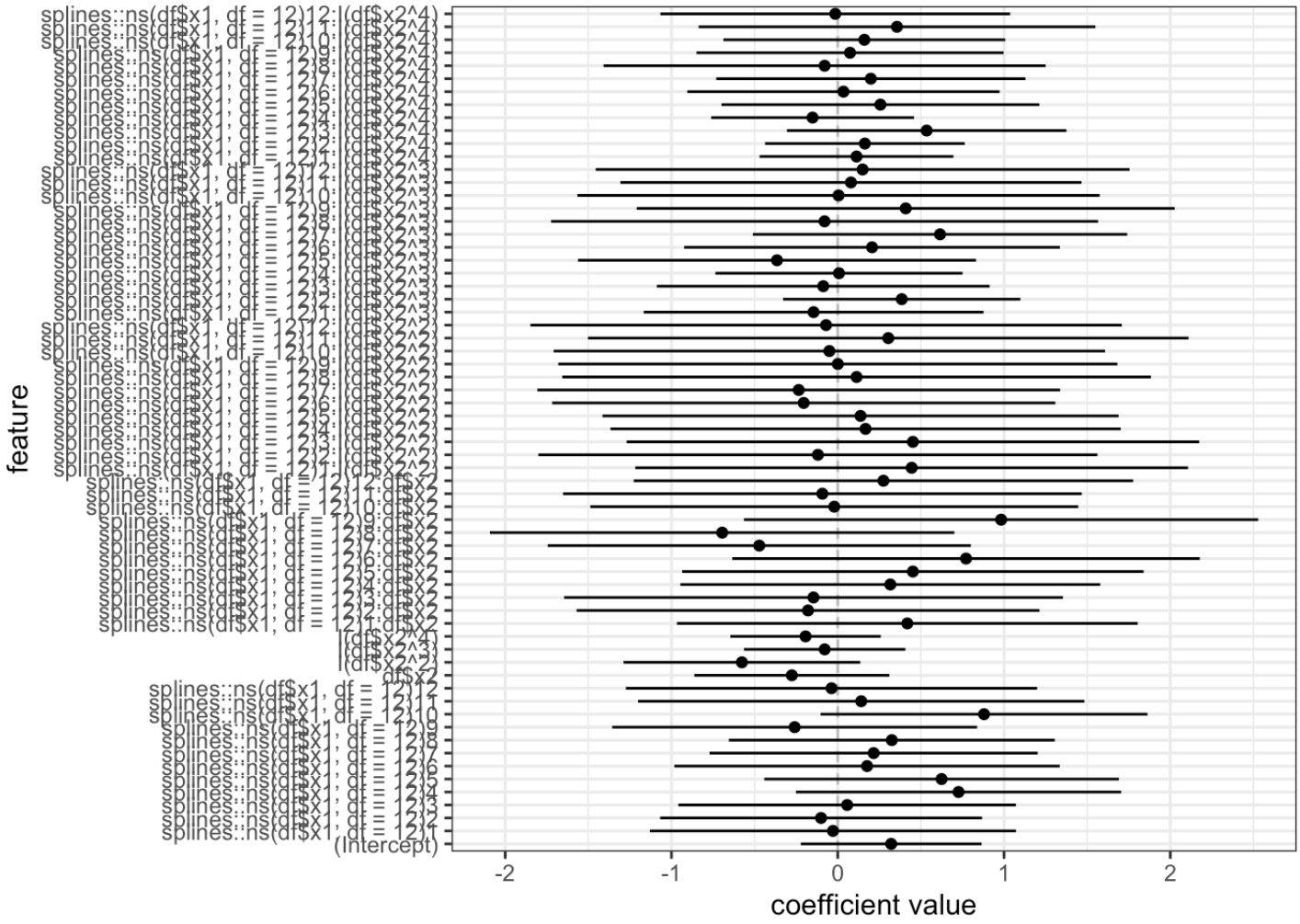
```
### add more code chunks if you like
sd <- c(sqrt(diag(laplace_03_strong$var_matrix)))
sd <- sd[-length(sd)]

viz_post_coefs(laplace_03_strong$mode[-length(laplace_03_strong$mode)],sd, colnames(X
03))
```




```
sd <- c(sqrt(diag(laplace_06_strong$var_matrix)))
sd <- sd[-length(sd)]

viz_post_coefs(laplace_06_strong$mode[-length(laplace_06_strong$mode)],sd, colnames(X
06))
```



3h)

You will fit one more set of Bayesian models with a very strong prior on the regression coefficients. The prior standard deviation will be equal to 1/50.

Complete the first code chunk below, which defines the list of required information for both the model 3 and model 6 formulations using the very strong prior on the regression coefficients. All other information, data and the σ prior, are the same as before.

Run the Laplace Approximation using the strong prior for both the model 3 and model 6 formulations. Assign the results to `laplace_03_very_strong` and `laplace_06_very_strong`.

Confirm that the optimizations converged for both laplace approximation results.

SOLUTION

```
info_03_very_strong <- list(
  yobs = df$y,
  design_matrix = X03,
  mu_beta = 0,
  tau_beta = 1/50,
  sigma_rate = 1
)

info_06_very_strong <- list(
  yobs = df$y,
  design_matrix = X06,
  mu_beta = 0,
  tau_beta = 1/50,
  sigma_rate = 1
)
```

Execute the Laplace Approximation.

```
### add more code chunks if you like
laplace_03_very_strong <- my_laplace(rep(0,ncol(X03) + 1), lm_logpost, info_03_very_s
trong)
laplace_03_very_strong$converge
```

```
## [1] "YES"
```

```
laplace_06_very_strong <- my_laplace(rep(0,ncol(X06) + 1), lm_logpost, info_06_very_s
trong)
laplace_06_very_strong$converge
```

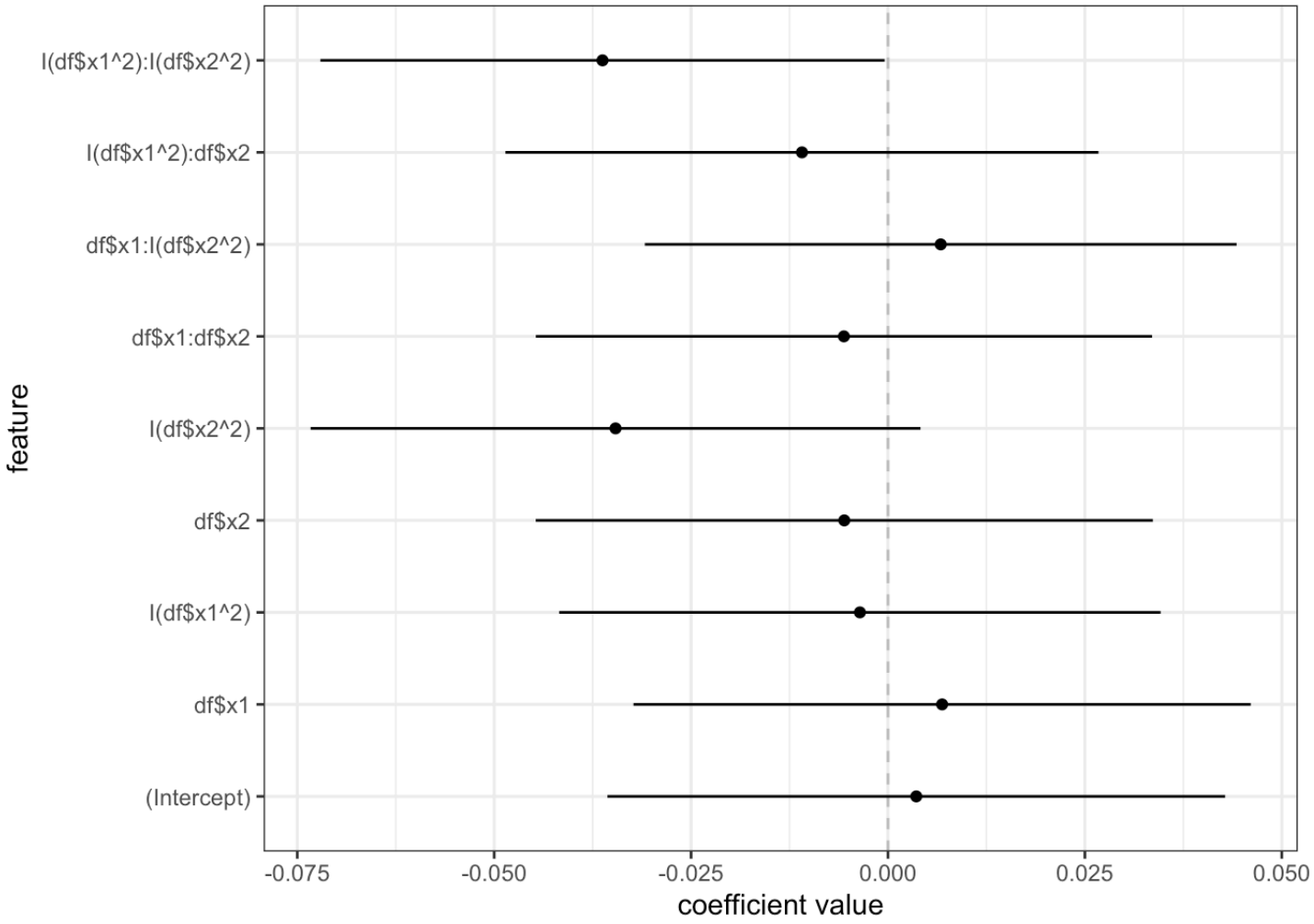
```
## [1] "YES"
```

3i)

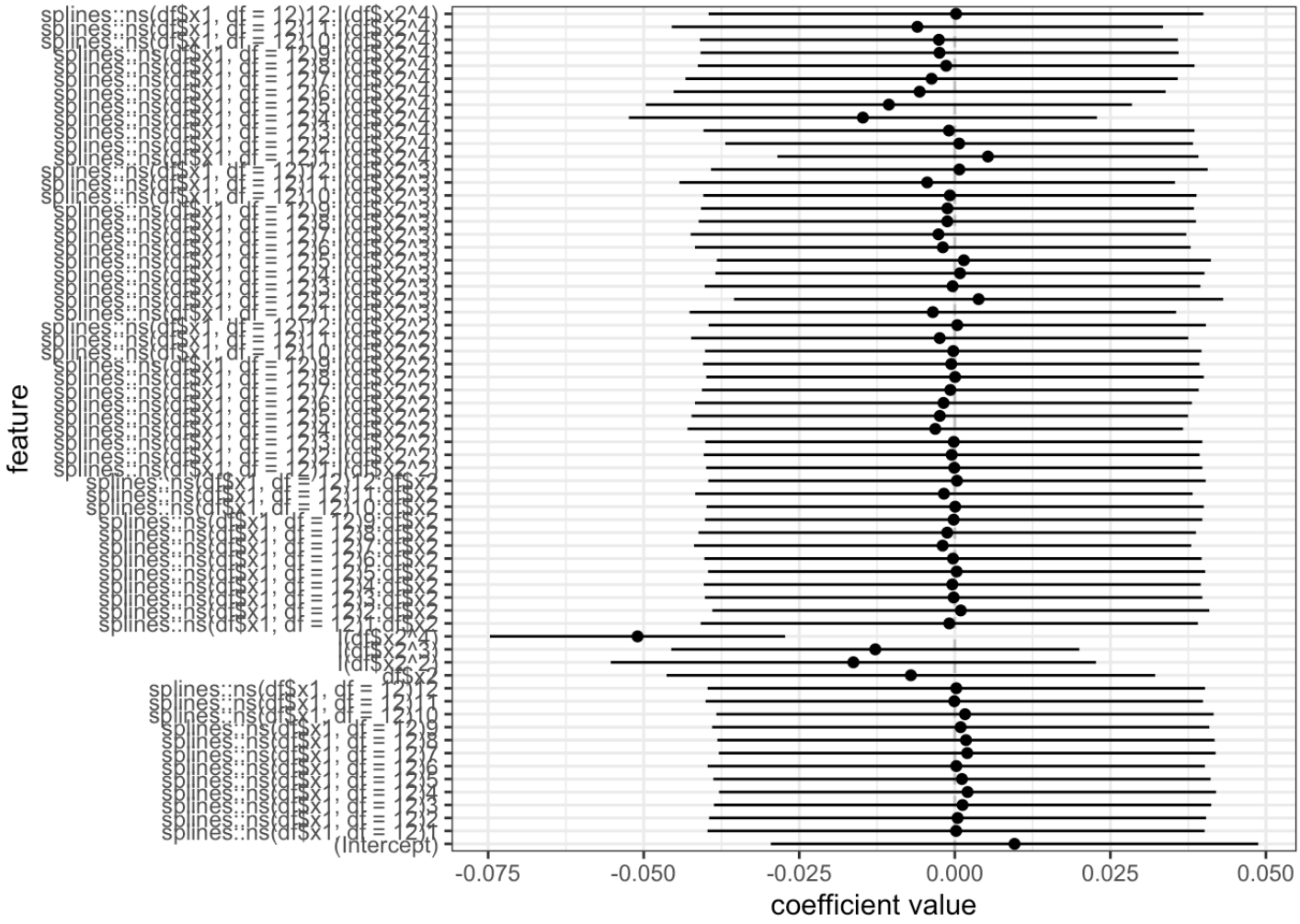
Use the `viz_post_coefs()` function to visualize the posterior coefficient summaries for model 3 and model 6, based on the very strong prior specification.

SOLUTION

```
### add more code chunks if you like
sd <- c(sqrt(diag(laplace_03_very_strong$var_matrix)))
sd <- sd[-length(sd)]
viz_post_coefs(laplace_03_very_strong$mode[-length(laplace_03_very_strong$mode)],sd,
colnames(X03))
```




```
sd <- c(sqrt(diag(laplace_06_very_strong$var_matrix)))
sd <- sd[-length(sd)]
viz_post_coefs(laplace_06_very_strong$mode[-length(laplace_06_very_strong$mode)],sd,
colnames(X06))
```



3j)

Describe the influence of the regression coefficient prior standard deviation on the coefficient posterior distributions.

SOLUTION

What do you think?

When the regression coefficient prior standard deviation on the coefficient posterior distribution is very strong, the posterior distribution follows the prior standard deviation. When the prior is weak, the posterior distribution coefficients follow the trends of the data more.

3k)

You previously compared the two models using the Bayes Factor based on the weak prior specification.

Compare the performance of the two models with Bayes Factors again, but considering the results based on the strong and very strong priors. Does the prior influence which model is considered to be better?

SOLUTIOn

```
### add more code chunks if you like
all_models <- list(laplace_03_strong, laplace_06_strong)
model_evidence <- purrr::map_dbl(all_models, "log_evidence")

model_weights <- model_evidence[1] / model_evidence[2]
model_weights
```

```
## [1] 0.8079022
```

```
all_models <- list(laplace_03_very_strong, laplace_06_very_strong)
model_evidence <- purrr::map_dbl(all_models, "log_evidence")

model_weights <- model_evidence[1] / model_evidence[2]
model_weights
```

```
## [1] 1.063057
```

For the strong prior model 6 appears to be still be doing better but not much better compared to the weak priors. For the very strong priors, model 3 now outperforms model 6.

Problem 04

You examined the behavior of the coefficient posterior based on the influence of the prior. Let's now consider the prior's influence by examining the posterior predictive distributions.

4a)

You will make posterior predictions following the approach from the previous assignment. Posterior samples are generated and those samples are used to calculate the posterior samples of the mean trend and generate random posterior samples of the response around the mean. In the previous assignment, you made posterior predictions in order to calculate errors. In this assignment, you will not calculate errors, instead you will summarize the posterior predictions of the mean and of the random response.

The `generate_lm_post_samples()` function is defined for you below. It uses the `MASS::mvrnorm()` function generate posterior samples from the Laplace Approximation's MVN distribution.

```
generate_lm_post_samples <- function(mvn_result, length_beta, num_samples)
{
  MASS::mvrnorm(n = num_samples,
                mu = mvn_result$mode,
                Sigma = mvn_result$var_matrix) %>%
    as.data.frame() %>% tibble::as_tibble() %>%
    purrr::set_names(c(sprintf("beta_%02d", 0:(length_beta-1)), "varphi")) %>%
    mutate(sigma = exp(varphi))
}
```

The code chunk below starts the `post_lm_pred_samples()` function. This function generates posterior mean trend predictions and posterior predictions of the response. The first argument, `Xnew`, is a potentially new or test design matrix that we wish to make predictions at. The second argument, `Bmat`, is a matrix of posterior samples of the $\boldsymbol{\beta}$-parameters, and the third argument, `sigma_vector`, is a vector of posterior samples of the likelihood noise. The `Xnew` matrix has rows equal to the number of predictions points, `M`, and the `Bmat` matrix has rows equal to the number of posterior samples `S`.

You must complete the function by performing the necessary matrix math to calculate the matrix of posterior mean trend predictions, `Umat`, and the matrix of posterior response predictions, `Ymat`. You must also complete missing arguments to the definition of the `Rmat` and `Zmat` matrices. The `Rmat` matrix replicates the posterior likelihood noise samples the correct number of times. The `Zmat` matrix is the matrix of randomly generated standard normal values. You must correctly specify the required number of rows to the `Rmat` and `Zmat` matrices.

The `post_lm_pred_samples()` returns the `Umat` and `Ymat` matrices contained within a list.

Perform the necessary matrix math to calculate the matrix of posterior predicted mean trends `Umat` and posterior predicted responses `Ymat`. You must specify the number of required rows to create the `Rmat` and `Zmat` matrices.

HINT: The following code chunk should look famaliar…

SOLUTION

```
post_lm_pred_samples <- function(Xnew, Bmat, sigma_vector)
{
  # number of new prediction locations
  M <- nrow(Xnew)
  # number of posterior samples
  S <- nrow(Bmat)

  # matrix of linear predictors
  Umat <- Xnew %*% t(Bmat)

  # assmeble matrix of sigma samples, set the number of rows
  Rmat <- matrix(rep(sigma_vector, M),M , byrow = TRUE)

  # generate standard normal and assemble into matrix
  # set the number of rows
  Zmat <- matrix(rnorm(M*S), M, byrow = TRUE)

  # calculate the random observation predictions
  Ymat <- Umat + Rmat * Zmat

  # package together
  list(Umat = Umat, Ymat = Ymat)
}
```

4b)

Since this assignment is focused on visualizing the predictions, we will summarize the posterior predictions to focus on the posterior means and the middle 95% uncertainty intervals. The code chunk below is defined for you which serves as a useful wrapper function to call `post_lm_pred_samples()`.

```
make_post_lm_pred <- function(Xnew, post)
{
  Bmat <- post %>% select(starts_with("beta_")) %>% as.matrix()

  sigma_vector <- post %>% pull(sigma)

  post_lm_pred_samples(Xnew, Bmat, sigma_vector)
}
```

The code chunk below defines a function `summarize_lm_pred_from_laplace()` which manages the actions necessary to summarize posterior predictions. The first argument, `mvn_result`, is the Laplace Approximation object. The second object is the test design matrix, `Xtest`, and the third argument, `num_samples`, is the number of posterior samples to make.

You must complete the code chunk below which summarizes the posterior predictions. This function takes care of most of the coding for you. You do not have to worry about the generation of the posterior samples OR calculating the posterior quantiles associated with the middle 95% uncertainty interval. You must calculate the posterior average by deciding on whether you should use `colMeans()` or `rowMeans()` to calculate the average across all posterior samples per prediction location.

Follow the comments in the code chunk below to complete the definition of the summarize_lm_pred_from_laplace() function. You must calculate the average posterior mean trend and the average posterior response.

SOLUTION

```
summarize_lm_pred_from_laplace <- function(mvn_result, Xtest, num_samples)
{
  # generate posterior samples of the beta parameters
  post <- generate_lm_post_samples(mvn_result, ncol(Xtest), num_samples)

  # make posterior predictions on the test set
  pred_test <- make_post_lm_pred(Xtest, post)

  # calculate summary statistics on the predicted mean and response
  # summarize over the posterior samples

  # posterior mean, should you summarize along rows (rowMeans) or
  # summarize down columns (colMeans) ???
  mu_avg <- rowMeans(pred_test$Umat)
  y_avg <-  rowMeans(pred_test$Ymat)

  # posterior quantiles for the middle 95% uncertainty intervals
  mu_lwr <- apply(pred_test$Umat, 1, stats::quantile, probs = 0.025)
  mu_upr <- apply(pred_test$Umat, 1, stats::quantile, probs = 0.975)
  y_lwr <- apply(pred_test$Ymat, 1, stats::quantile, probs = 0.025)
  y_upr <- apply(pred_test$Ymat, 1, stats::quantile, probs = 0.975)

  # book keeping
  tibble::tibble(
    mu_avg = mu_avg,
    mu_lwr = mu_lwr,
    mu_upr = mu_upr,
    y_avg = y_avg,
    y_lwr = y_lwr,
    y_upr = y_upr
  ) %>%
    tibble::rowid_to_column("pred_id")
}
```

4c)

When you made predictions in Problem 02, the `lm()` object handled making the test design matrix. However, since we have programmed the Bayesian modeling approach from scratch we need to create the test design matrix manually.

Create the test design matrix based on the visualization grid, `viz_grid`, using the model 3 formulation. Assign the result to the `X03_test` object.

Call the `summarize_lm_pred_from_laplace()` function to summarize the posterior predictions from the model 3 formulation for the weak, strong, and very strong prior specifications. Use 5000 posterior samples for each case. Assign the results from the weak prior to `post_pred_summary_viz_03_weak`, the results from the strong prior to `post_pred_summary_viz_03_strong`, and the results from the very strong prior to `post_pred_summary_viz_03_very_strong`.

SOLUTION

```
### add as many code chunks as you'd like
X03_test <- model.matrix( ~ (x1 + I(x1^2)) * (x2 + I(x2^2)), data = viz_grid)
post_pred_summary_viz_03_weak <- summarize_lm_pred_from_laplace(laplace_03_weak, X03_
test, 5000)
post_pred_summary_viz_03_strong<- summarize_lm_pred_from_laplace(laplace_03_strong, X
03_test, 5000)
post_pred_summary_viz_03_very_strong <- summarize_lm_pred_from_laplace(laplace_03_ver
y_strong, X03_test, 5000)
```

4d)

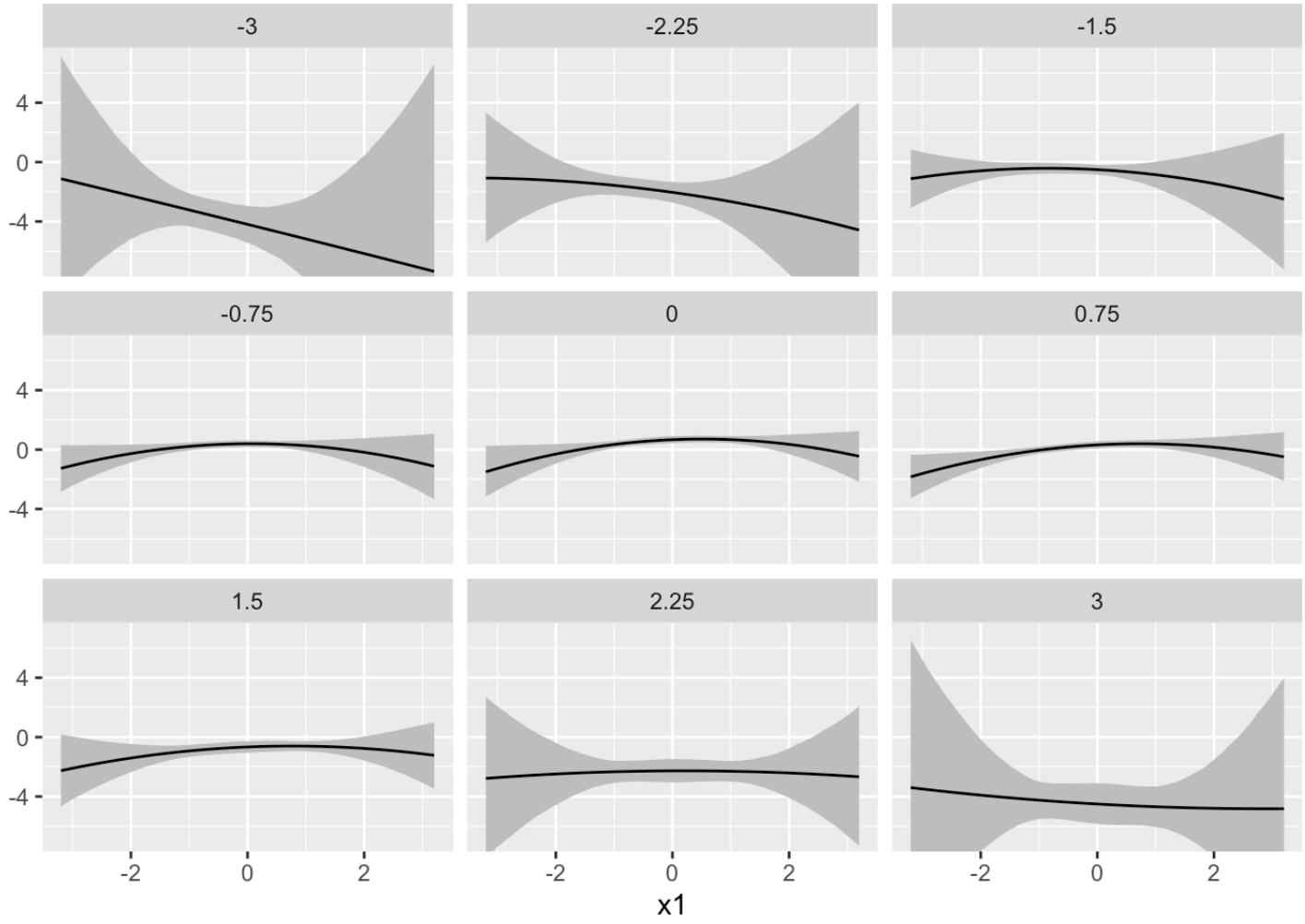
You will now visualize the posterior predictions from the model 3 Bayesian models associated with the weak, strong, and very strong priors. The `viz_grid` object is joined to the prediction dataframes assuming you have used the correct variable names!

Visualize the predicted means, confidence intervals, and prediction intervals in the style of those that you created in Problem 02. The confidence interval bounds are `mu_lwr` and `mu_upr` columns and the prediction interval bounds are the `y_lwr` and `y_upr` columns, respectively. The posterior predicted mean of the mean is `mu_avg`.

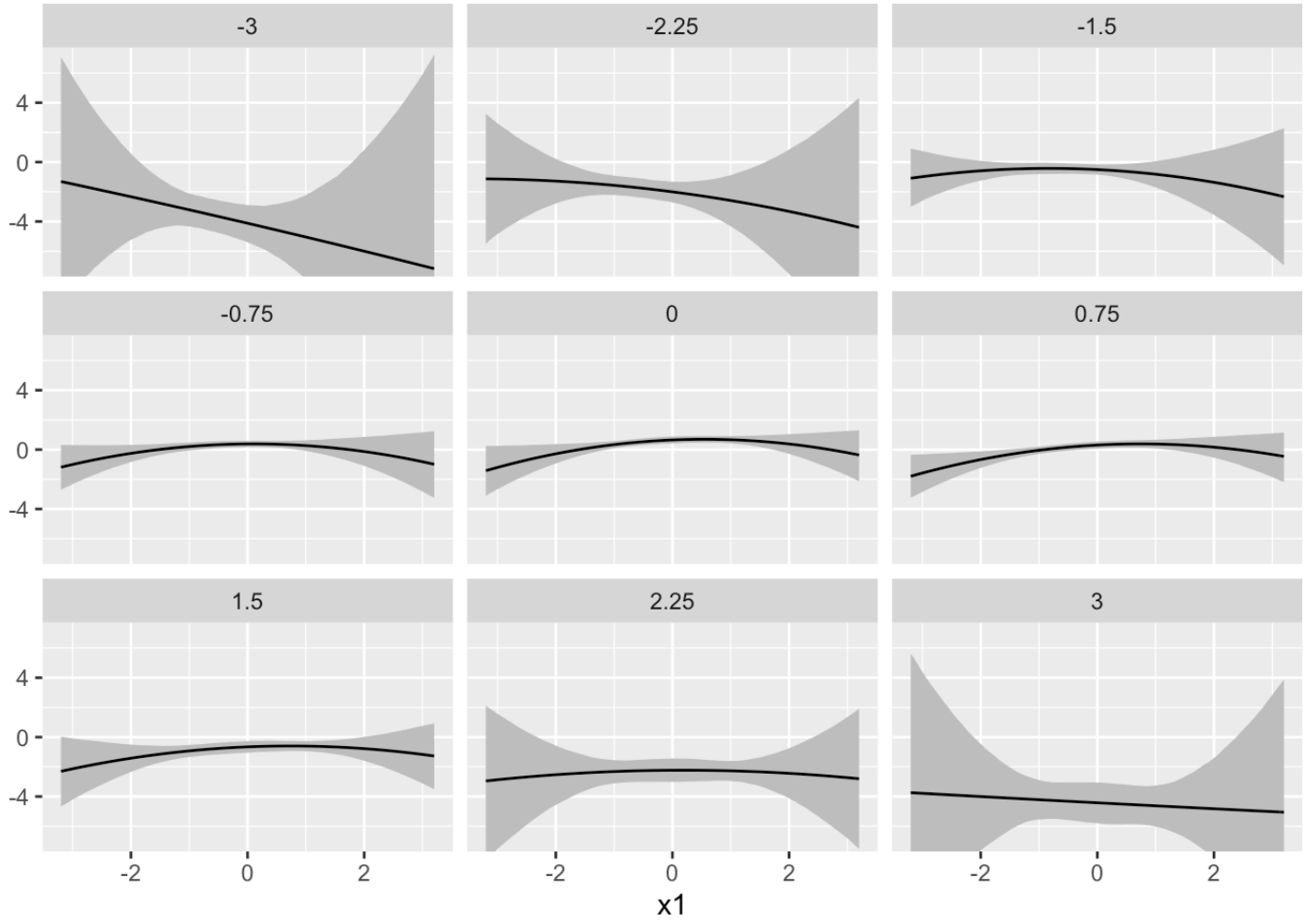
Pipe the result of the joined dataframe into `ggplot()` and make appropriate aesthetics and layers to visualize the predictions with the `x1` variable mapped to the `x` aesthetic and the `x2` variable used as a facet variable.

SOLUTION

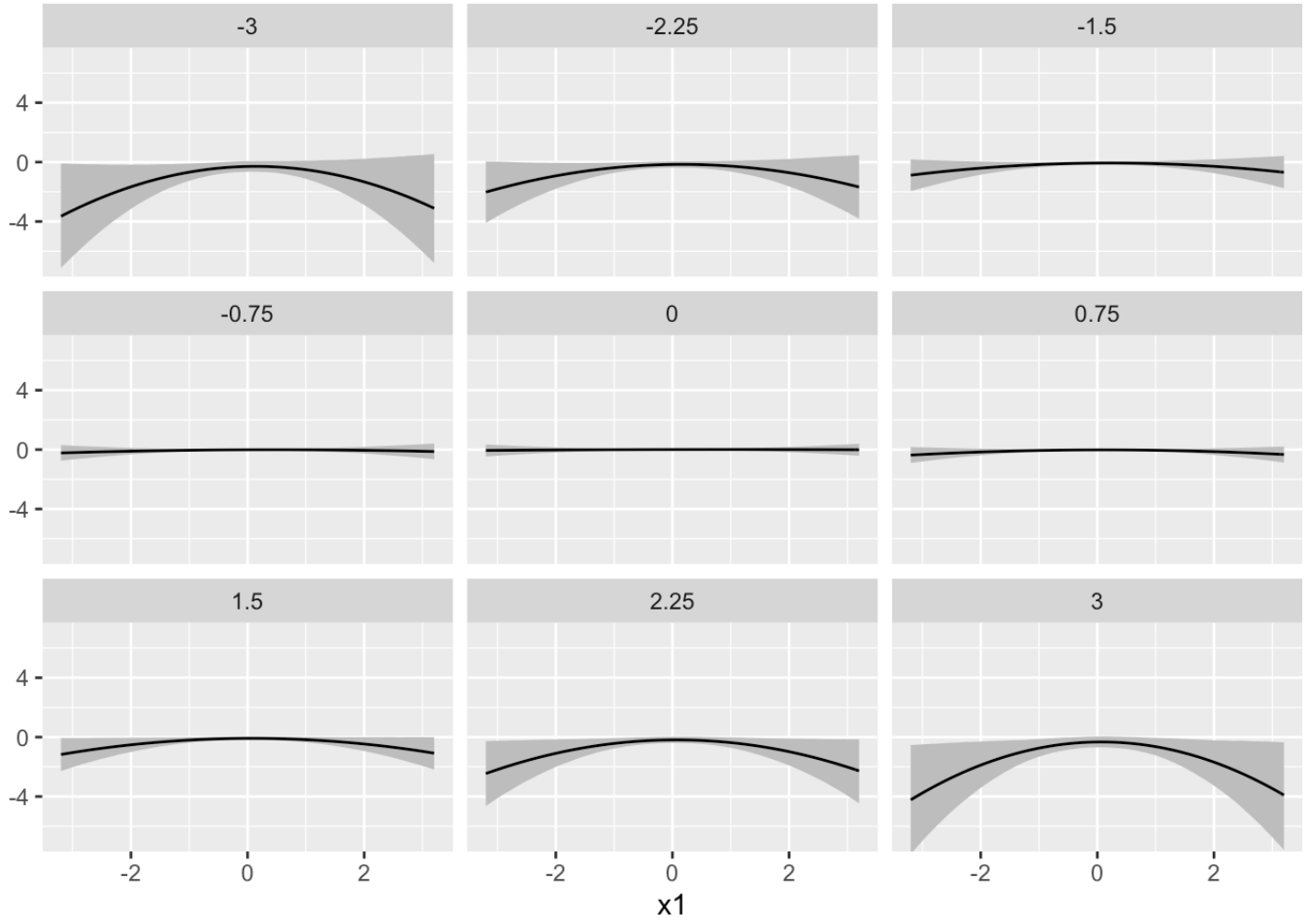
```
post_pred_summary_viz_03_weak %>%
  left_join(viz_grid %>% tibble::rowid_to_column("pred_id"),
            by = 'pred_id') %>%
  ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = y_lwr, ymax = y_lwr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = mu_lwr, ymax = mu_upr), fill = 'grey') +
  geom_line(y = post_pred_summary_viz_03_weak$mu_avg)+
  facet_wrap(facets = viz_grid$x2)+
  coord_cartesian(ylim = c(-7,7))
```



```
post_pred_summary_viz_03_strong %>%
  left_join(viz_grid %>% tibble::rowid_to_column("pred_id"),
            by = 'pred_id')%>%
  ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = y_lwr, ymax = y_lwr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = mu_lwr, ymax = mu_upr), fill = 'grey') +
  geom_line(y = post_pred_summary_viz_03_strong$mu_avg)+
  facet_wrap(facets = viz_grid$x2)+
  coord_cartesian(ylim = c(-7,7))
```



```
post_pred_summary_viz_03_very_strong %>%
  left_join(viz_grid %>% tibble::rowid_to_column("pred_id"),
            by = 'pred_id')%>%
  ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = y_lwr, ymax = y_lwr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = mu_lwr, ymax = mu_upr), fill = 'grey') +
  geom_line(y = post_pred_summary_viz_03_very_strong$mu_avg)+
  facet_wrap(facets = viz_grid$x2)+
  coord_cartesian(ylim = c(-7,7))
```



4e)

In order to make posterior predictions for the model 6 formulation you must create a test design matrix consistent with the training set basis. The code chunk below creates a helper function which extracts the knots of a natural spline associated with the training set for you. The first argument, `J`, is the degrees-of-freedom of the spline, the second argument, `train_data`, is the training data set. The third argument `xname` is the name of the variable you are applying the spline to. The `xname` argument **must** be provided as a character string.

```
make_splines_training_knots <- function(J, train_data, xname)
{
  x <- train_data %>% select(all_of(xname)) %>% pull()

  train_basis <- splines::ns(x, df = J)

  as.vector(attributes(train_basis)$knots)
}
```

Create the test design matrix based on the visualization grid, `viz_grid`, using the model 6 formulation. Assign the result to the `X06_test` object. Use the `make_splines_training_knots()` to get the necessary knots associated with the training set for the `x1` variable to create the test design matrix.

Call the `summarize_lm_pred_from_laplace()` function to summarize the posterior predictions from the model 6 formulation for the weak, strong, and very strong prior specifications. Use 5000 posterior samples for each case. Assign the results from the weak prior to `post_pred_summary_viz_06_weak`, the results from the strong prior to `post_pred_summary_viz_06_strong`, and the results from the very strong prior to `post_pred_summary_viz_06_very_strong`.

SOLUTION

```
### add as many code chunks as you'd like
X06_test <- model.matrix( ~ (splines::ns(x1, df = 12)) * (x2 + I(x2^2) + I(x2^3) + I(
x2^4)), data = viz_grid)
post_pred_summary_viz_06_weak <- summarize_lm_pred_from_laplace(laplace_06_weak, X06_
test, 5000)
post_pred_summary_viz_06_strong<- summarize_lm_pred_from_laplace(laplace_06_strong, X
06_test, 5000)
post_pred_summary_viz_06_very_strong <- summarize_lm_pred_from_laplace(laplace_06_ver
y_strong, X06_test, 5000)
```

4f)

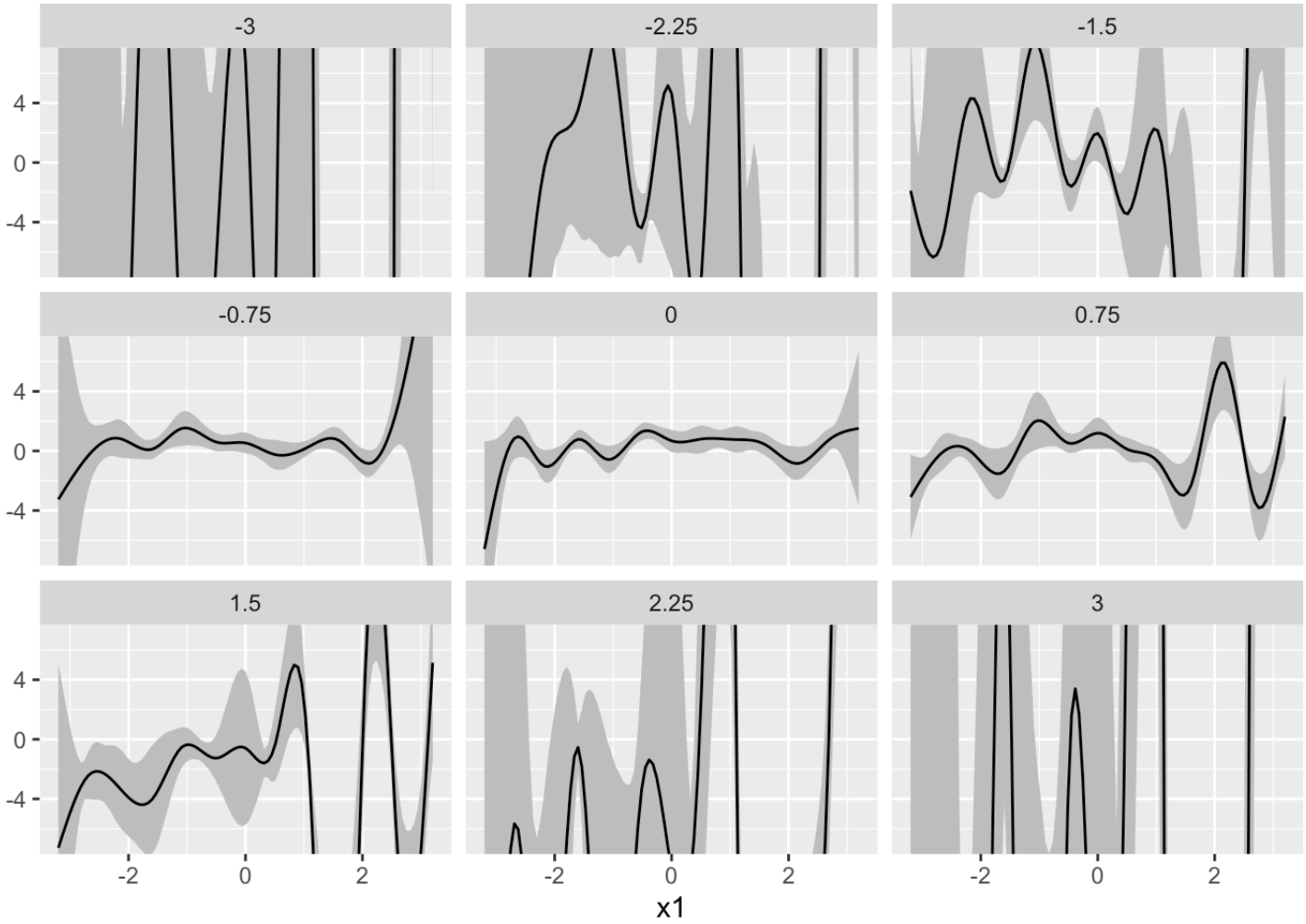
You will now visualize the posterior predictions from the model 6 Bayesian models associated with the weak, strong, and very strong priors. The `viz_grid` object is joined to the prediction dataframes assuming you have used the correct variable names!

Visualize the predicted means, confidence intervals, and prediction intervals in the style of those that you created in Problem 02. The confidence interval bounds are `mu_lwr` and `mu_upr` columns and the prediction interval bounds are the `y_lwr` and `y_upr` columns, respectively. The posterior predicted mean of the mean is `mu_avg`.

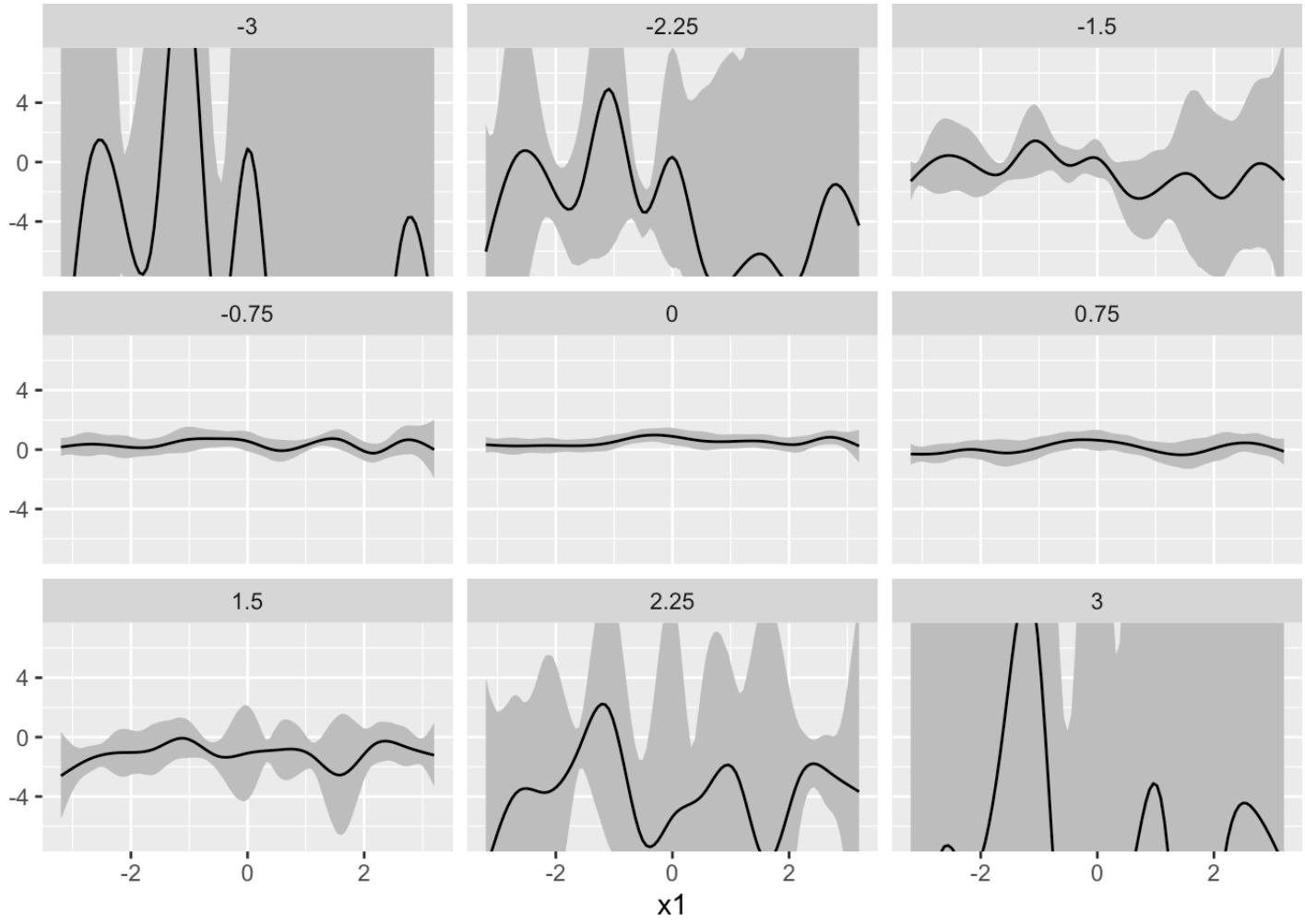
Pipe the result of the joined dataframe into `ggplot()` and make appropriate aesthetics and layers to visualize the predictions with the `x1` variable mapped to the `x` aesthetic and the `x2` variable used as a facet variable.

SOLUTION

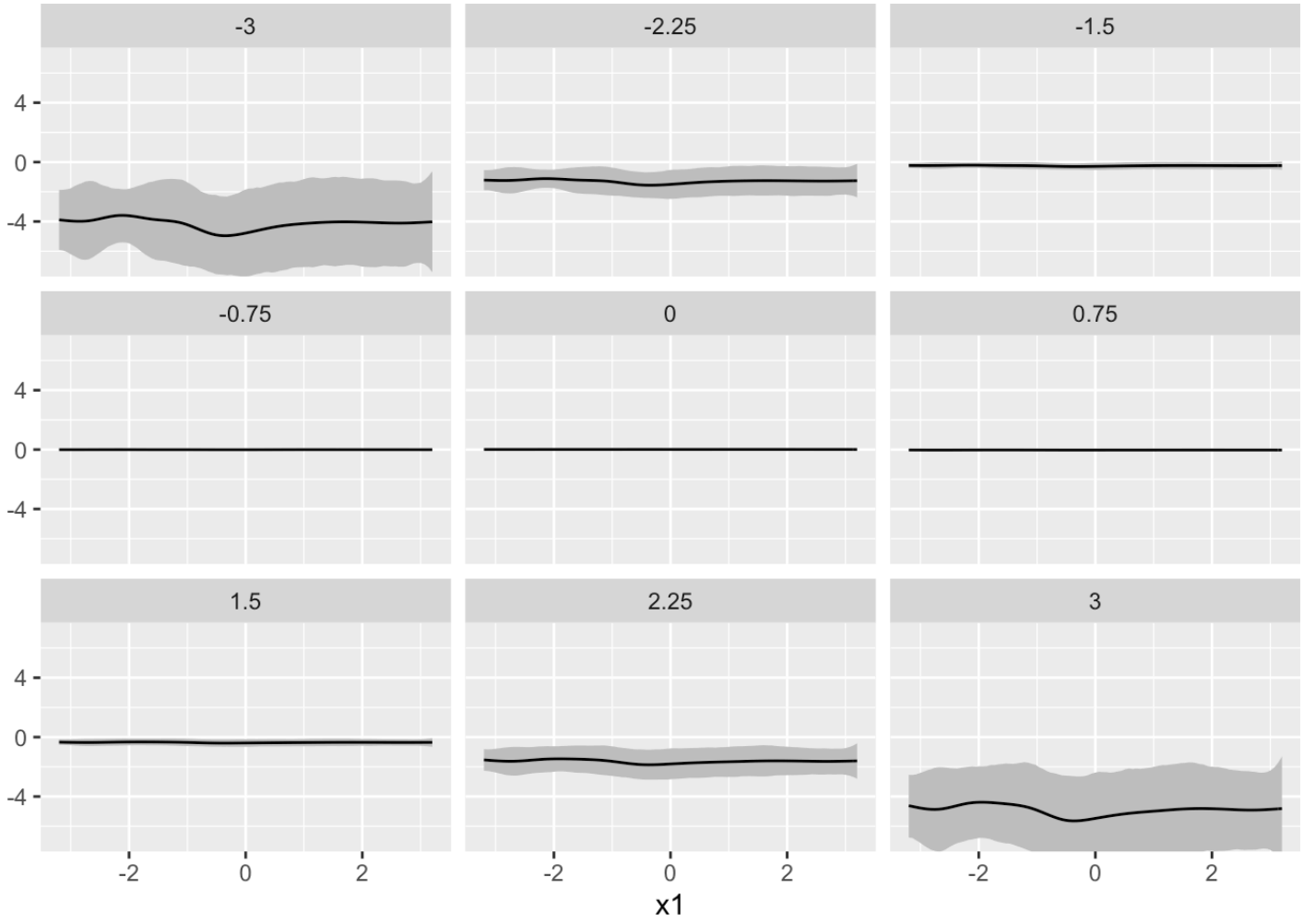
```
post_pred_summary_viz_06_weak %>%
  left_join(viz_grid %>% tibble::rowid_to_column("pred_id"),
            by = 'pred_id')%>%
  ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = y_lwr, ymax = y_lwr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = mu_lwr, ymax = mu_upr), fill = 'grey') +
  geom_line(y = post_pred_summary_viz_06_weak$mu_avg)+
  facet_wrap(facets = viz_grid$x2)+
  coord_cartesian(ylim = c(-7,7))
```



```
post_pred_summary_viz_06_strong %>%
  left_join(viz_grid %>% tibble::rowid_to_column("pred_id"),
            by = 'pred_id')%>%
  ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = y_lwr, ymax = y_lwr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = mu_lwr, ymax = mu_upr), fill = 'grey') +
  geom_line(y = post_pred_summary_viz_06_strong$mu_avg)+
  facet_wrap(facets = viz_grid$x2)+
  coord_cartesian(ylim = c(-7,7))
```



```
post_pred_summary_viz_06_very_strong %>%
  left_join(viz_grid %>% tibble::rowid_to_column("pred_id"),
            by = 'pred_id') %>%
  ggplot(mapping = aes(x = x1)) +
  geom_ribbon(mapping = aes(ymin = y_lwr, ymax = y_lwr), fill = 'orange') +
  geom_ribbon(mapping = aes(ymin = mu_lwr, ymax = mu_upr), fill = 'grey') +
  geom_line(y = post_pred_summary_viz_06_very_strong$mu_avg)+
  facet_wrap(facets = viz_grid$x2)+
  coord_cartesian(ylim = c(-7,7))
```



4g)

Describe the behavior of the predictions as the prior standard deviation decreased. Are the posterior predictions consistent with the behavior of the posterior coefficients?

SOLUTION

What do you think?

As the prior standard deviation decreases, the predictions getting closer to 0. THe posterior predictions are consistent with the behavior of the posterior coefficients.

Problem 05

Now that you have worked with Bayesian models with the prior *regularizing* the coefficients, you will consider non-Bayesian regularization methods. You will work with the `glmnet` package in this problem which takes care of all fitting and visualization for you.

The code chunk below loads in `glmnet` and so you must have `glmnet` installed before running this code chunk. **IMPORANT**: the `eval` flag is set to FALSE below. Once you download `glmnet` set `eval=TRUE`.

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## Loaded glmnet 4.1-3
```

5a)

`glmnet` does not work with the formula interface. And so you must create the training design matrix. However, `glmnet` prefers the the intercept column of ones to **not** be included in the design matrix. To support that you must define new design matrices. These matrices will use the same formulation but you must remove the intercept column. This is easy to do with the formula interface and the `model.matrix()` function. Include `- 1` in the formula and `model.matrix()` will not include the intercept. The code chunk below demonstrates removing the intercept column for a model with linear additive features.

```
model.matrix( y ~ x1 + x2 - 1, data = df) %>% head()
```

```
##           x1         x2
## 1 -0.3092328  0.3087799
## 2  0.6312721 -0.5479198
## 3 -0.6827669  2.1664494
## 4  0.2693056  1.2097037
## 5  0.3725202  0.7854860
## 6  1.2966439 -0.1877231
```

Create the design matrices for `glmnet` for the model 3 and model 6 formulations. Remove the intercept column for both and assign the results to `X03_glmnet` and `X06_glmnet` .

SOLUTION

```
### add more code chunks if you prefer
X03_glmnet <- model.matrix(y ~ (x1 + I(x1^2)) * (x2 + I(x2^2)) - 1, data = df)
X06_glmnet <- model.matrix(y ~ (splines::ns(x1, df = 12)) * (x2 + I(x2^2) + I(x2^3) +
I(x2^4)) - 1, data = df)
```

5b)

By default `glmnet` uses the **lasso** penalty. Fit a Lasso model by calling `glmnet()` . The first argument to `glmnet()` is the design matrix and the second argument is a regular vector for the response.

Train a Lasso model for the model 3 and model 6 formulations, assign the results to `lasso_03` and `lasso_06` , respectively.

SOLUTION

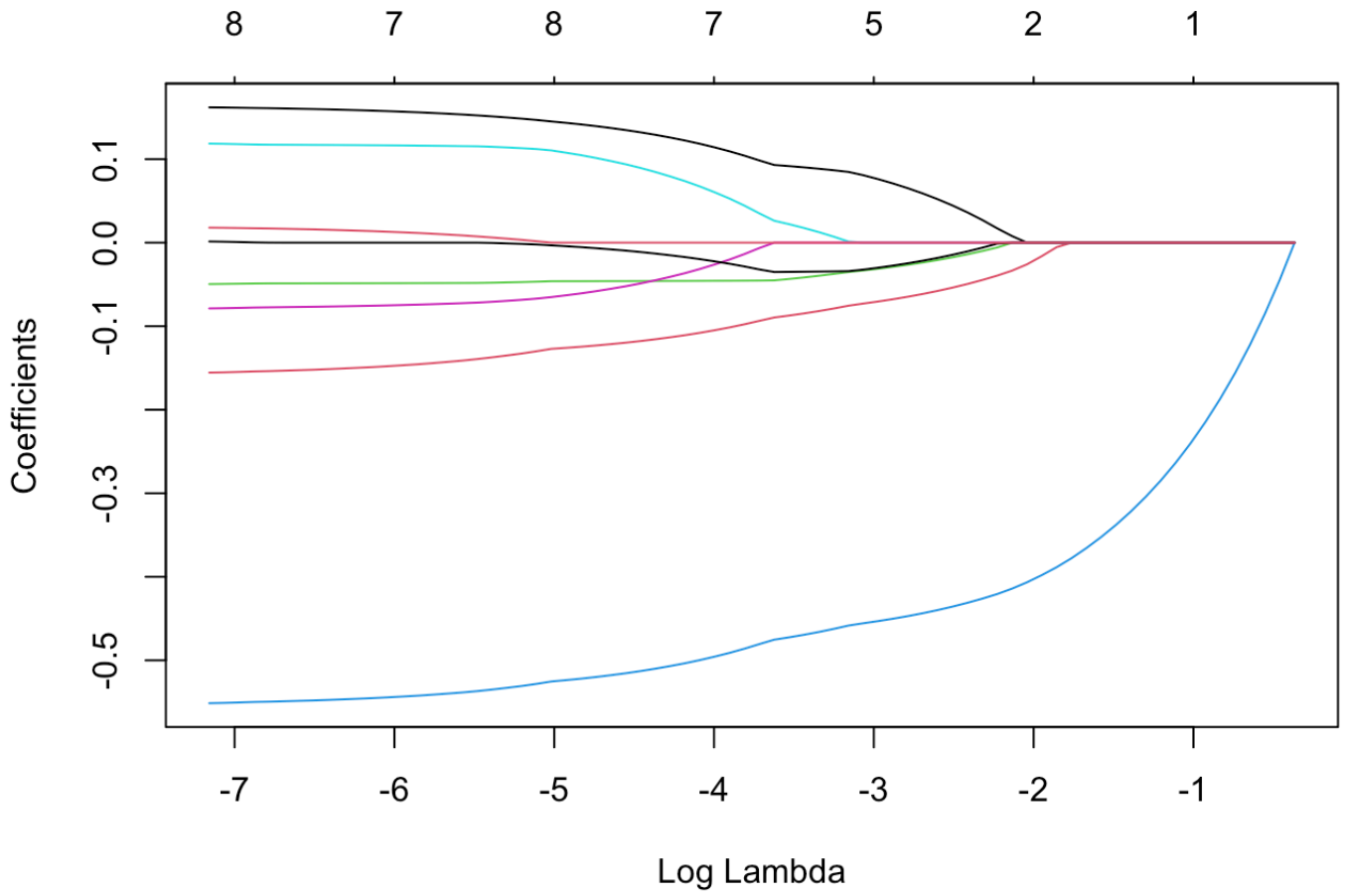
```
### add more code chunks if you like
lasso_03 <- glmnet(X03_glmnet, df$y)
lasso_06 <- glmnet(X06_glmnet, df$y)
```

5c)

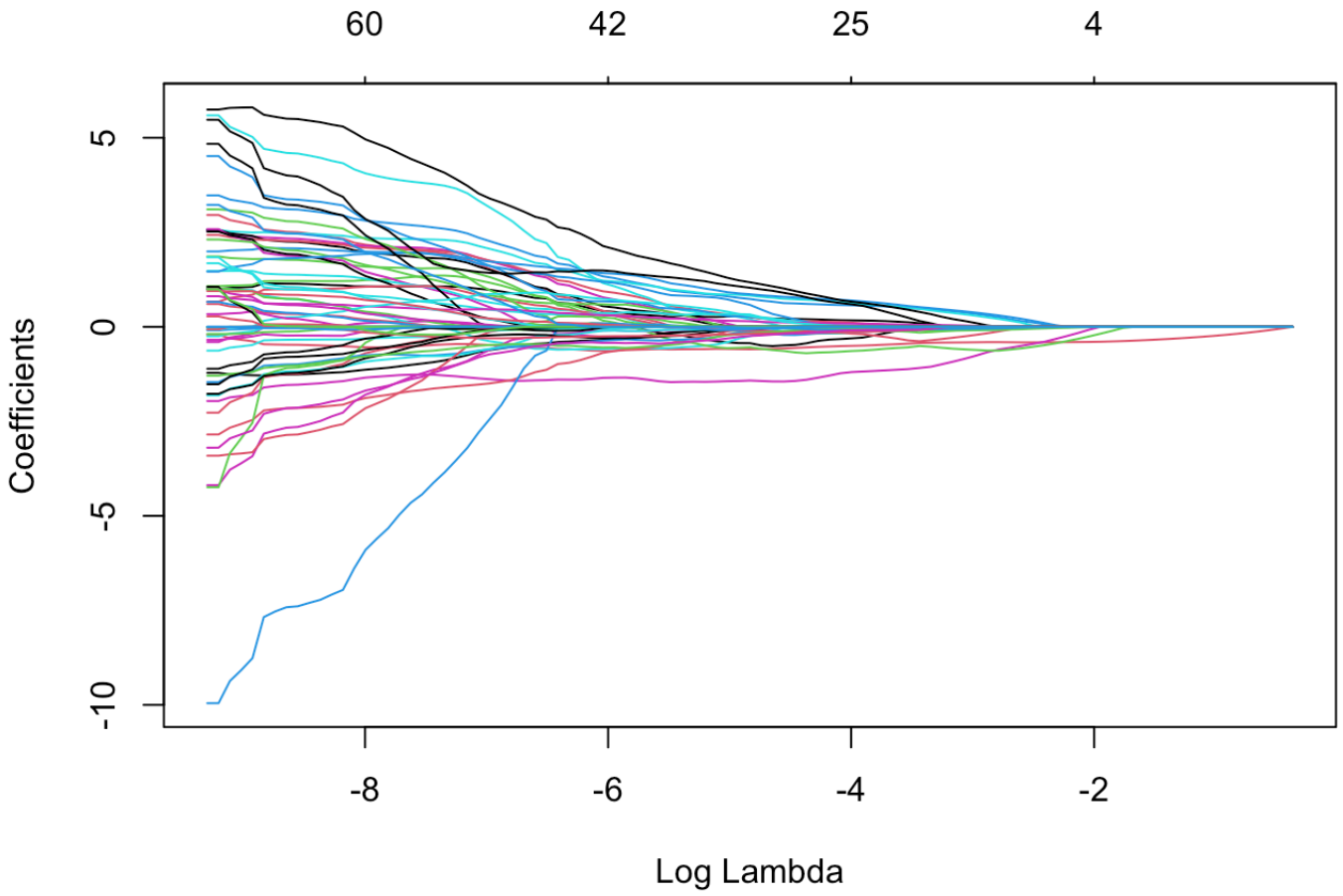
Plot the coefficient path for each Lasso model by calling the `plot()` function on the `glmnet` model object. Specify the `xvar` argument to be `'lambda'` in the `plot()` call.

SOLUTION

```
### add more code chunks if you like
plot(lasso_03, xvar = "lambda")
```




```
plot(lasso_06, xvar = "lambda")
```



5d)

Now that you have visualized the coefficient path, it's time to identify the best `'lambda'` value to use! The `cv.glmnet()` function will by default use 10-fold cross-validation to tune `'lambda'`. The first argument to `cv.glmnet()` is the design matrix and the second argument is the regular vector for the response.

Tune the Lasso regularization strength with cross-validation using the `cv.glmnet()` function for each model formulation. Assign the model 3 result to `lasso_03_cv_tune` and assign the model 6 result to `lasso_06_cv_tune`. Also specify the `alpha` argument to be 1 to make sure the Lasso penalty is applied in the `cv.glmnet()` call.

SOLUTION

```
### add more code chunks if you like
set.seed(123)
lasso_03_tune <- cv.glmnet(X03_glmnet, df$y, nfolds = 10)
lasso_03_tune
```

```
##
## Call:  cv.glmnet(x = X03_glmnet, y = df$y, nfolds = 10)
##
## Measure: Mean-Squared Error
##
##      Lambda Index Measure      SE Nonzero
## min 0.05619    28  0.5241 0.04607       5
## 1se 0.20668    14  0.5684 0.03607       1
```

```
set.seed(123)
lasso_06_tune <- cv.glmnet(X06_glmnet, df$y, nfolds = 10,
                           lambda = exp(seq(log(0.001), log(5), length.out = 101)))
lasso_06_tune
```

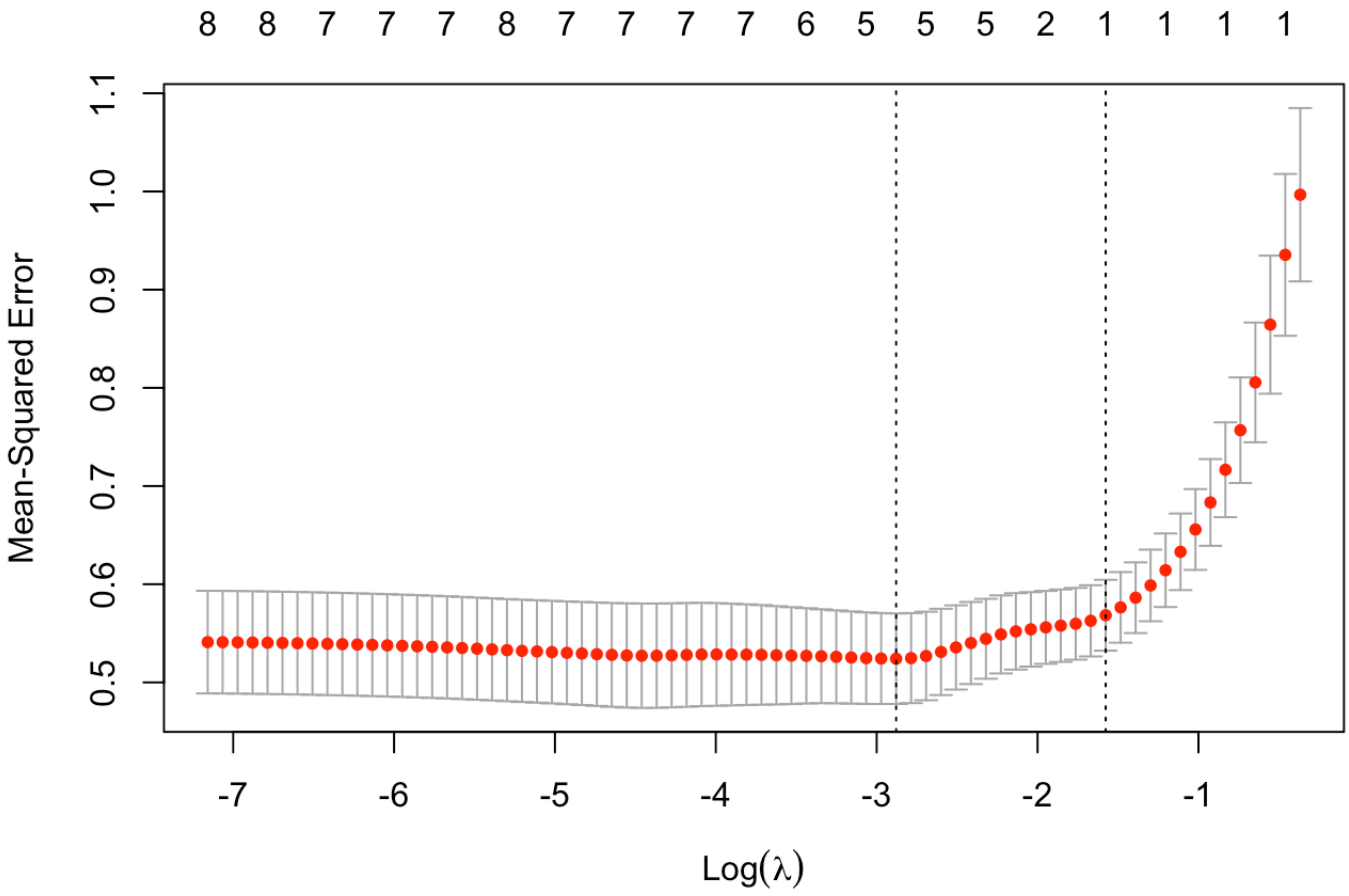
```
##
## Call:  cv.glmnet(x = X06_glmnet, y = df$y, lambda = exp(seq(log(0.001),      log(5
), length.out = 101)), nfolds = 10)
##
## Measure: Mean-Squared Error
##
##     Lambda Index Measure      SE Nonzero
## min 0.1522    42  0.5575 0.03758       2
## 1se 0.2537    36  0.5884 0.03598       1
```

5e)

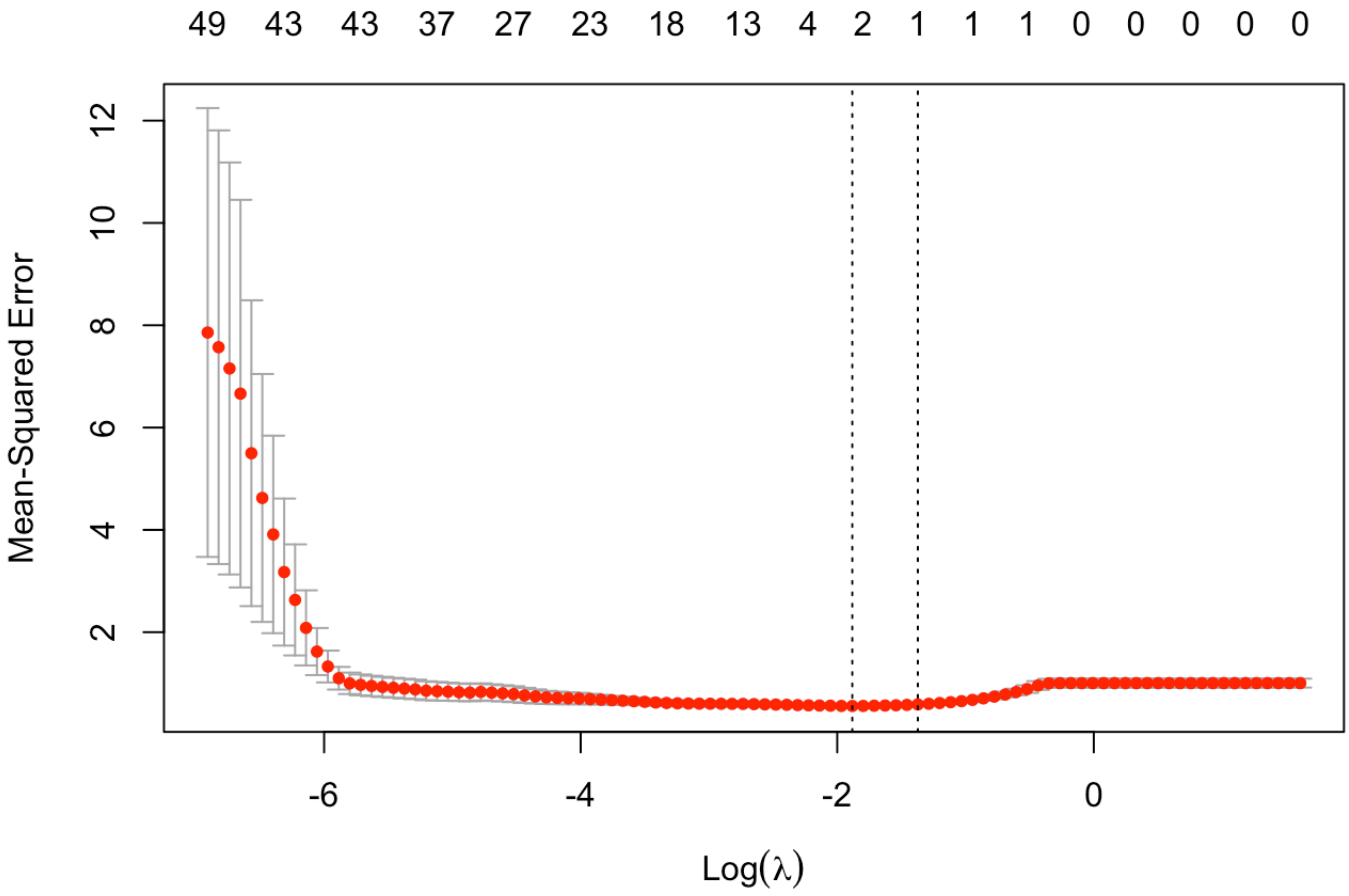
Plot the cross-validation results using the default plot method for each cross-validation result. How many coefficients are remaining after tuning?

SOLUTION

```
### add more code chunks if you like
plot(lasso_03_tune)
```




```
plot(lasso_06_tune)
```



Model 3: 6 coefficients Model 6: 13 coefficients

5f)

Which features have NOT been "turned off" by the Lasso penalty? Use the `coef()` function to display the lasso model cross-validation results to show the tuned penalized regression coefficients for each model.
Are the final tuned models different from each other?

SOLUTION

```
### add more code chunks if you like
coef(lasso_03_tune)
```

```
## 9 x 1 sparse Matrix of class "dgCMatrix"
##                         s1
## (Intercept)      0.3487439
## x1               .
## I(x1^2)          .
## x2               .
## I(x2^2)         -0.3522666
## x1:x2            .
## x1:I(x2^2)       .
## I(x1^2):x2       .
## I(x1^2):I(x2^2)  .
```

```
coef(lasso_06_tune)
```

```
## 65 x 1 sparse Matrix of class "dgCMatrix"
##                                       s1
## (Intercept)                    0.3150025
## splines::ns(x1, df = 12)1      .
## splines::ns(x1, df = 12)2      .
## splines::ns(x1, df = 12)3      .
## splines::ns(x1, df = 12)4      .
## splines::ns(x1, df = 12)5      .
## splines::ns(x1, df = 12)6      .
## splines::ns(x1, df = 12)7      .
## splines::ns(x1, df = 12)8      .
## splines::ns(x1, df = 12)9      .
## splines::ns(x1, df = 12)10     .
## splines::ns(x1, df = 12)11     .
## splines::ns(x1, df = 12)12     .
## x2                             .
## I(x2^2)                       -0.3181844
## I(x2^3)                        .
## I(x2^4)                        .
## splines::ns(x1, df = 12)1:x2   .
## splines::ns(x1, df = 12)2:x2   .
## splines::ns(x1, df = 12)3:x2   .
## splines::ns(x1, df = 12)4:x2   .
## splines::ns(x1, df = 12)5:x2   .
## splines::ns(x1, df = 12)6:x2   .
## splines::ns(x1, df = 12)7:x2   .
## splines::ns(x1, df = 12)8:x2   .
## splines::ns(x1, df = 12)9:x2   .
## splines::ns(x1, df = 12)10:x2  .
## splines::ns(x1, df = 12)11:x2  .
## splines::ns(x1, df = 12)12:x2  .
```

```
## splines::ns(x1, df = 12)1:I(x2^2)   .
## splines::ns(x1, df = 12)2:I(x2^2)   .
## splines::ns(x1, df = 12)3:I(x2^2)   .
## splines::ns(x1, df = 12)4:I(x2^2)   .
## splines::ns(x1, df = 12)5:I(x2^2)   .
## splines::ns(x1, df = 12)6:I(x2^2)   .
## splines::ns(x1, df = 12)7:I(x2^2)   .
## splines::ns(x1, df = 12)8:I(x2^2)   .
## splines::ns(x1, df = 12)9:I(x2^2)   .
## splines::ns(x1, df = 12)10:I(x2^2)  .
## splines::ns(x1, df = 12)11:I(x2^2)  .
## splines::ns(x1, df = 12)12:I(x2^2)  .
## splines::ns(x1, df = 12)1:I(x2^3)   .
## splines::ns(x1, df = 12)2:I(x2^3)   .
## splines::ns(x1, df = 12)3:I(x2^3)   .
## splines::ns(x1, df = 12)4:I(x2^3)   .
## splines::ns(x1, df = 12)5:I(x2^3)   .
## splines::ns(x1, df = 12)6:I(x2^3)   .
## splines::ns(x1, df = 12)7:I(x2^3)   .
## splines::ns(x1, df = 12)8:I(x2^3)   .
## splines::ns(x1, df = 12)9:I(x2^3)   .
## splines::ns(x1, df = 12)10:I(x2^3)  .
## splines::ns(x1, df = 12)11:I(x2^3)  .
## splines::ns(x1, df = 12)12:I(x2^3)  .
## splines::ns(x1, df = 12)1:I(x2^4)   .
## splines::ns(x1, df = 12)2:I(x2^4)   .
## splines::ns(x1, df = 12)3:I(x2^4)   .
## splines::ns(x1, df = 12)4:I(x2^4)   .
## splines::ns(x1, df = 12)5:I(x2^4)   .
## splines::ns(x1, df = 12)6:I(x2^4)   .
## splines::ns(x1, df = 12)7:I(x2^4)   .
## splines::ns(x1, df = 12)8:I(x2^4)   .
## splines::ns(x1, df = 12)9:I(x2^4)   .
## splines::ns(x1, df = 12)10:I(x2^4)  .
## splines::ns(x1, df = 12)11:I(x2^4)  .
## splines::ns(x1, df = 12)12:I(x2^4)  .
```

The only features left in the model are the intercept and x2^2 in both model 3 and model 6.