Exercise 7

Instructor of PSYC 573

Note: before submission, please change the YAML option to eval: true

In this exercise, you will practice model comparison and stacking, using the example described in the note.

```
library(haven)
library(dplyr)
library(brms)
options(brms.backend = "cmdstanr")
library(bayesplot)
```

We'll use a data set kidiq that is used in the textbook by Gelman et al. (2021), which can be downloaded and imported with the direct link:

```
kidiq <- haven::read_dta(
    "http://www.stat.columbia.edu/~gelman/arm/examples/child.iq/kidiq.dta")
head(kidiq)</pre>
```

Let's run four models. First rescale some of the variables:

```
kidiq100 <- kidiq |>
mutate(mom_iq = mom_iq / 100, # divid mom_iq by 100
kid_score = kid_score / 100, # divide kid_score by 100
mom_iq_c = mom_iq - 1,
mom_hs = factor(mom_hs, labels = c("no", "yes")),
mom_age_c = (mom_age - 18) / 10)
```

The first two models are:

```
m1 <- brm(kid_score ~ mom_iq_c,
    data = kidiq100,
    prior = c(
        prior(normal(0, 1), class = "Intercept"),
        prior(normal(0, 1), class = "b"),
        prior(student_t(4, 0, 1), class = "sigma")
    ),
    file = "ex7_m1"
)
# Use `update` will sometimes avoid recompiling
m2 <- update(m1, kid_score ~ mom_iq_c + mom_hs,
        newdata = kidiq100,
        file = "ex7_m2"
)
```

Q1: Fit two additional models:

- m3: interaction between mom_iq_c and mom_hs
- m4: interaction between mom_iq_c and mom_hs and mom_age_c

 $\ensuremath{\texttt{\#}}$ Fit m3 and m4

Q2: Obtain the posterior predictive distribution of kid_score for someone with $mom_iq = 120$, $mom_hs = no$, and $mom_age = 35$, from each model. How different are the predictions across models?

```
new_data <- data.frame(
    mom_iq_c = 120 / 100 - 1,
    mom_hs = "no",
    mom_age_c = (35 - 18) / 10
)
m1_pp <- posterior_predict(m1, newdata = new_data)</pre>
```

Q3: Use add_criterion() to add loo and waic to each model, and use loo_compare() to compare the models. Which model is the best according to LOO-IC?

`add_criterion()` to each model

Compare the models

Q4: The following obtain weights for stacking. Is there a relationship between the weights and the model LOO-IC?

Weights based on Stacking (based on the posterior predictive distribution)
(stack_wts <- loo_model_weights(m1, m2, m3, m4))</pre>

Q5: Use pp_average() to obtain the prediction of the new observation in new_data. Show a density plot of the predictive distribution.

Compute prediction based on stacking

Density plot of the predictive distribution based on stacking

Q6: How does the prediction based on stacking compare to the predictions based on the individual models?