AE 7: Exam 2 Review

Add your name here

Packages

```
library(tidyverse)
library(tidymodels)
library(knitr)
library(openintro)

# fix data!
loans_full_schema <- droplevels(loans_full_schema)
```

Goal

Create a model for precicting `interest_rate`.

View data

Note the dimensions of the data and the variable names. Review the data dictionary.

```
# add code here
```

Split data into training and testing

Split your data into testing and training sets.

```
# add code here
```

Write the model

Write the model for predicting interest rate (`interest_rate`) from debt to income ratio (`debt_to_income`), the term of loan (`term`), the number of inquiries (credit checks) into the applicant's credit during the last 12 months (`inquiries_last_12m`), whether there are any bankruptcies listed in the public record for this applicant (`bankrupt`), and the type of application (`application_type`). The model should allow for the effect of to income ratio on interest rate to vary by application type.

Add model here

Exploration

Explore characteristics of the variables you'll use for the model using the training data only.

```
# add code here
```

Specify model

Specify a linear regression model. Call it `office_spec`.

```
# add code here
```

Create recipe

- Predict `interest_rate` from `debt_to_income`, `term`, `inquiries_last_12m`, `public_record_bankrupt`, and `application_type`.
- Mean center `debt_to_income`.
- Make `term` a factor.
- Create a new variable: `bankrupt` that takes on the value "no" if `public_record_bankrupt` is 0 and the value "yes" if `public_record_bankrupt` is 1 or higher. Then, remove `public_record_bankrupt`.
- Interact `application_type` with `debt_to_income`.
- Create dummy variables where needed and drop any zero variance variables.

```
# add code here
```

Create workflow

Create the workflow that brings together the model specification and recipe.

```
# add code here
```

Cross validation

Conduct 10-fold cross validation.

```
# add code here
```

Summarize CV metrics

Summarize metrics from your CV resamples.

```
# add code here
```

Why are we focusing on R-squared and RMSE instead of adjusted R-squared, AIC, BIC?

[Add response here]

Next steps...

Depending on time, either

- Create a workflow for another model with a new recipe (omitting the interaction variable), conduct CV, do model selection between these two, and then interpret the coefficients for the selected model.
- Or interpret the coefficients for the one model you fit.

Make sure to interpret the intercept and slope coefficient for at least one numerical, one categorical, and one interaction predictor.