

Wayne Luo HW2

```
library(ggthemes)
library(ggplot2)
library(tidymodels)
```

```
## -- Attaching packages ----- tidymodels 0.2.0 --
```

```
## v broom      0.7.12    v rsample      0.1.1
## v dials      0.1.0     v tibble       3.1.6
## v dplyr      1.0.8     v tidyr        1.2.0
## v infer      1.0.0     v tune         0.2.0
## v modeldata  0.1.1     v workflows    0.2.6
## v parsnip    0.2.1     v workflowsets 0.2.1
## v purrr      0.3.4     v yardstick    0.0.9
## v recipes    0.2.0
```

```
## -- Conflicts ----- tidymodels_conflicts() --
```

```
## x purrr::discard() masks scales::discard()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x recipes::step()  masks stats::step()
## * Use tidymodels_prefer() to resolve common conflicts.
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v readr      2.1.2     v forcats      0.5.1
## v stringr    1.4.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x readr::col_factor() masks scales::col_factor()
## x purrr::discard()    masks scales::discard()
## x dplyr::filter()     masks stats::filter()
## x stringr::fixed()    masks recipes::fixed()
## x dplyr::lag()        masks stats::lag()
## x readr::spec()       masks yardstick::spec()
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(yardstick)
tidymodels_prefer()
```

```
abalone <- read.csv(file = 'abalone.csv')
abalone %>% head()
```

```
##   type longest_shell diameter height whole_weight shucked_weight viscera_weight
## 1    M           0.455   0.365  0.095     0.5140         0.2245         0.1010
## 2    M           0.350   0.265  0.090     0.2255         0.0995         0.0485
## 3    F           0.530   0.420  0.135     0.6770         0.2565         0.1415
## 4    M           0.440   0.365  0.125     0.5160         0.2155         0.1140
## 5    I           0.330   0.255  0.080     0.2050         0.0895         0.0395
## 6    I           0.425   0.300  0.095     0.3515         0.1410         0.0775
##   shell_weight rings
## 1         0.150    15
## 2         0.070     7
## 3         0.210     9
## 4         0.155    10
## 5         0.055     7
## 6         0.120     8
```

QUESTION 1:

```
abalone["age"] <- abalone$rings + 1.5
head(abalone)
```

```
##   type longest_shell diameter height whole_weight shucked_weight viscera_weight
## 1    M           0.455   0.365  0.095     0.5140         0.2245         0.1010
## 2    M           0.350   0.265  0.090     0.2255         0.0995         0.0485
## 3    F           0.530   0.420  0.135     0.6770         0.2565         0.1415
## 4    M           0.440   0.365  0.125     0.5160         0.2155         0.1140
## 5    I           0.330   0.255  0.080     0.2050         0.0895         0.0395
## 6    I           0.425   0.300  0.095     0.3515         0.1410         0.0775
##   shell_weight rings  age
## 1         0.150    15 16.5
## 2         0.070     7  8.5
## 3         0.210     9 10.5
## 4         0.155    10 11.5
## 5         0.055     7  8.5
## 6         0.120     8  9.5
```

QUESTION 2:

```
set.seed(3435)
abalone_split <- initial_split(abalone, prop = 0.80, strata = age)
abalone_train <- training(abalone_split)
abalone_test  <- testing(abalone_split)
```

QUESTION 3:

```
?tidymodels
abalone_train_new <- abalone_train %>% select(-rings)
abalone_train_new %>% head()
```

```
##   type longest_shell diameter height whole_weight shucked_weight
## 5    I           0.330   0.255  0.080     0.2050         0.0895
## 17   I           0.355   0.280  0.085     0.2905         0.0950
## 19   M           0.365   0.295  0.080     0.2555         0.0970
```

```
## 36      M      0.465    0.355  0.105      0.4795      0.2270
## 38      F      0.450    0.355  0.105      0.5225      0.2370
## 43      I      0.240    0.175  0.045      0.0700      0.0315
##      viscera_weight shell_weight age
## 5          0.0395          0.055 8.5
## 17         0.0395          0.115 8.5
## 19         0.0430          0.100 8.5
## 36         0.1240          0.125 9.5
## 38         0.1165          0.145 9.5
## 43         0.0235          0.020 6.5
```

```
ad_rec <- recipe(age~ ., data = abalone_train_new) %>%
  step_dummy(all_nominal_predictors()) %>%
  step_interact(terms = ~ type:shucked_weight) %>%
  step_interact(terms = ~ longest_shell:diameter) %>%
  step_interact(terms = ~ shucked_weight:shell_weight) %>%
  step_normalize(all_nominal_predictors())
```

```
ad_rec
```

```
## Recipe
##
## Inputs:
##
##      role #variables
##      outcome      1
##      predictor      8
##
## Operations:
##
## Dummy variables from all_nominal_predictors()
## Interactions with type:shucked_weight
## Interactions with longest_shell:diameter
## Interactions with shucked_weight:shell_weight
## Centering and scaling for all_nominal_predictors()
```

QUESTION 4:

```
lm_model_abalone <- linear_reg() %>%
  set_engine("lm")
```

QUESTION 5:

```
lm_wflow_abalone <- workflow() %>%
  add_model(lm_model_abalone) %>%
  add_recipe(ad_rec)
```

QUESTION 6:

```
hypo_abalone <- data.frame(type = "F", longest_shell = 0.50, diameter = 0.10, height = 0.30, whole_weight = 0.27,
                           shucked_weight = 1, viscera_weight = 2, shell_weight = 1)

lm_fit_abalone <- fit(lm_wflow_abalone, abalone_train_new)
```

```
## Warning: Interaction specification failed for: ~type:shucked_weight. No
## interactions will be created.
```

```
lm_fit_abalone %>% extract_fit_parsnip() %>%
  tidy()
```

```
## # A tibble: 12 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)         2.16      0.590      3.66 2.52e- 4
## 2 longest_shell       7.15      2.36      3.03 2.46e- 3
## 3 diameter           23.9      3.13      7.66 2.51e-14
## 4 height              5.88      1.64      3.59 3.39e- 4
## 5 whole_weight        8.65      0.792     10.9 2.80e-27
## 6 shucked_weight     -16.4      1.08     -15.3 7.28e-51
## 7 viscera_weight     -7.41      1.45      -5.11 3.44e- 7
## 8 shell_weight       13.6      1.52      8.92 7.51e-19
## 9 type_I             -0.708     0.116     -6.11 1.13e- 9
## 10 type_M              0.0566    0.0930      0.608 5.43e- 1
## 11 longest_shell_x_diameter -34.7      4.13     -8.40 6.64e-17
## 12 shucked_weight_x_shell_weight -1.93      1.69     -1.14 2.55e- 1
```

```
res <- predict(lm_fit_abalone, new_data = hypo_abalone)
res %>% head()
```

```
## # A tibble: 1 x 1
##   .pred
##   <dbl>
## 1  23.2
```

QUESTION 7:

```
abalone_train_res <- predict(lm_fit_abalone, new_data = abalone_train_new)
abalone_train_res
```

```
## # A tibble: 3,340 x 1
##   .pred
##   <dbl>
## 1  8.22
## 2  9.95
## 3 10.3
## 4 10.1
## 5 10.6
## 6  6.35
## 7  5.76
## 8  5.94
## 9  8.87
## 10 11.4
## # ... with 3,330 more rows
```

```
abalone_train_res <- bind_cols(abalone_train_res, abalone_train_new %>% select(age))
abalone_train_res %>% head()
```

```
## # A tibble: 6 x 2
##   .pred age
##   <dbl> <dbl>
## 1  8.22  8.5
## 2  9.95  8.5
## 3 10.3   8.5
## 4 10.1   9.5
## 5 10.6   9.5
## 6  6.35  6.5
```

```
rmse(abalone_train_res, truth = age, estimate = .pred)
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>         <dbl>
## 1 rmse    standard        2.18
```

```
abalone_metrics <- metric_set(rmse, rsq, mae)
abalone_metrics(abalone_train_res, truth = age, estimate = .pred)
```

```
## # A tibble: 3 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>         <dbl>
## 1 rmse    standard        2.18
## 2 rsq     standard        0.546
## 3 mae     standard        1.57
```