Simple Models

Mike Calabro

Set Up

```
library(tidyverse)
library(janitor)
library(tidymodels)
library(rpart.plot)
library(discrim)

set.seed(80)

train <- read_csv('train.csv') %>%
  clean_names() %>%
  mutate(survived = factor(survived),
         pclass = factor(pclass))

folds <- vfold_cv(train, v = 10)
```

Decision Tree Model Workflow

```
tree_spec <-
  decision_tree(min_n = 20) %>%
  set_engine('rpart') %>%
  set_mode('classification')

tree_recipe <-
  recipe(survived ~ pclass + sex + age + embarked, data = train)

tree_workflow <-
  workflow() %>%
  add_model(tree_spec) %>%
```

```
  add_recipe(tree_recipe)

keep_pred <- control_resamples(save_pred = TRUE, save_workflow = TRUE)

tree_res <-
  tree_workflow %>% fit_resamples(resamples = folds, control = keep_pred)
```

Collecting the Metrics

```
collect_metrics(tree_res)
```

```
# A tibble: 2 x 6
  .metric  .estimator  mean     n std_err .config
  <chr>    <chr>      <dbl> <int>   <dbl> <chr>
1 accuracy binary     0.810    10 0.0109  Preprocessor1_Model1
2 roc_auc  binary     0.808    10 0.00996 Preprocessor1_Model1
```

Naive Bayes Model Workflow

```
bayes_spec <-
  naive_Bayes(Laplace = 1) %>%
  set_engine('naivebayes')


bayes_recipe <-
  recipe(survived ~ pclass + sex + age + embarked, data = train) %>%
  step_impute_mode(embarked) %>%
  step_dummy(all_nominal_predictors()) %>%
  step_normalize(all_numeric_predictors())

bayes_workflow <-
  workflow() %>%
  add_model(bayes_spec) %>%
  add_recipe(bayes_recipe)

bayes_res <-
  bayes_workflow %>% fit_resamples(resamples = folds, control = keep_pred)
```

Collecting the Metrics

```
bayes_res %>% collect_metrics()
```

```
# A tibble: 2 x 6
  .metric  .estimator  mean     n std_err .config
  <chr>    <chr>      <dbl> <int>   <dbl> <chr>
1 accuracy binary     0.788    10  0.0129 Preprocessor1_Model1
2 roc_auc  binary     0.838    10  0.0154 Preprocessor1_Model1
```

K-Nearest Neighbors Model Workflow

```
knn_spec <-
  nearest_neighbor(neighbors = 21) %>%
  set_engine('kknn') %>%
  set_mode('classification')


knn_recipe <-
  recipe(survived ~ pclass + sex + age + embarked, data = train) %>%
  step_impute_mean(age) %>%
  step_impute_mode(embarked) %>%
  step_dummy(all_nominal_predictors()) %>%
  step_normalize(all_numeric_predictors())

knn_workflow <-
  workflow() %>%
  add_model(knn_spec) %>%
  add_recipe(knn_recipe)

knn_res <-
  knn_workflow %>% fit_resamples(resamples = folds, control = keep_pred)
```

Collecting the Metrics

```
knn_res %>% collect_metrics()
```

```
# A tibble: 2 x 6
  .metric  .estimator  mean     n std_err .config
  <chr>    <chr>      <dbl> <int>   <dbl> <chr>
1 accuracy binary     0.793    10  0.0110 Preprocessor1_Model1
2 roc_auc  binary     0.859    10  0.0111 Preprocessor1_Model1
```

Logistic Model Workflow

```
glm_spec <-
  logistic_reg() %>%
  set_engine('glm')


glm_recipe <-
  recipe(survived ~ pclass + sex + age + embarked, data = train) %>%
  step_impute_mean(age) %>%
  step_impute_mode(embarked) %>%
  step_dummy(all_nominal_predictors())

glm_workflow <-
  workflow() %>%
  add_model(glm_spec) %>%
  add_recipe(glm_recipe)

glm_res <-
  glm_workflow %>% fit_resamples(resamples = folds, control = keep_pred)
```

Collecting the Metrics

```
glm_res %>% collect_metrics()
```

```
# A tibble: 2 x 6
  .metric  .estimator  mean     n std_err .config
  <chr>    <chr>      <dbl> <int>   <dbl> <chr>
1 accuracy binary     0.788    10  0.0108 Preprocessor1_Model1
2 roc_auc  binary     0.850    10  0.0133 Preprocessor1_Model1
```

Random Forest Model Workflow

```
forest_spec <-
  rand_forest(mtry = 2, min_n = 10) %>%
  set_engine('ranger') %>%
  set_mode('classification')

forest_recipe <-
  recipe(survived ~ pclass + sex + age + embarked, data = train) %>%
  step_impute_mean(age) %>%
  step_impute_mode(embarked)

forest_workflow <-
  workflow() %>%
  add_model(forest_spec) %>%
  add_recipe(forest_recipe)

forest_res <-
  forest_workflow %>% fit_resamples(resamples = folds, control = keep_pred)
```

Collecting the Metrics

```
forest_res %>% collect_metrics()
```

```
# A tibble: 2 x 6
  .metric  .estimator  mean     n std_err .config
  <chr>    <chr>      <dbl> <int>   <dbl> <chr>
1 accuracy binary     0.814    10 0.00836 Preprocessor1_Model1
2 roc_auc  binary     0.863    10 0.0107  Preprocessor1_Model1
```