

Assignment2

Hannah Cronin

2023-03-26

1. Separating the data like in the book.

```
library(keras)
original_dataset_dir <- "/Users/hannahcronin/Desktop/kaggle_original_data"
base_dir <- "/Users/hannahcronin/Desktop/kaggle_original_data"
dir.create(base_dir)
```

```
## Warning in dir.create(base_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data' already exists
```

```
train_dir <- file.path(base_dir, "train")
dir.create(train_dir)
```

```
## Warning in dir.create(train_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/train' already exists
```

```
validation_dir <- file.path(base_dir, "validation")
dir.create(validation_dir)
```

```
## Warning in dir.create(validation_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/validation' already exists
```

```
test_dir <- file.path(base_dir, "test")
dir.create(test_dir)
```

```
## Warning in dir.create(test_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/test' already exists
```

```
train_cats_dir <- file.path(train_dir, "cats")
dir.create(train_cats_dir)
```

```
## Warning in dir.create(train_cats_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/train/cats' already exists
```

```
train_dogs_dir <- file.path(train_dir, "dogs")
dir.create(train_dogs_dir)
```

```
## Warning in dir.create(train_dogs_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/train/dogs' already exists
```

```
validation_cats_dir <- file.path(validation_dir, "cats")
dir.create(validation_cats_dir)
```

```
## Warning in dir.create(validation_cats_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/validation/cats' already exists
```

```
validation_dogs_dir <- file.path(validation_dir, "dogs")
dir.create(validation_dogs_dir)
```

```
## Warning in dir.create(validation_dogs_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/validation/dogs' already exists
```

```
test_cats_dir <- file.path(test_dir, "cats")
dir.create(test_cats_dir)
```

```
## Warning in dir.create(test_cats_dir): '/Users/hannahcronin/Desktop/
## kaggle_original_data/test/cats' already exists
```

```
test_dogs_dir <- file.path(test_dir, "dogs")  
dir.create(test_dogs_dir)
```

```
## Warning in dir.create(test_dogs_dir): '/Users/hannahcronin/Desktop/  
## kaggle_original_data/test/dogs' already exists
```

```
fnames <- paste0("cat.", 1:1000, ".jpg")  
file.copy(file.path(train_dir, fnames),  
          file.path(train_cats_dir))
```

file:///Users/hannahcronin/Desktop/GITHUB/HCRONIN_MACHINELEARNING/Assignment 2/Assignment2.html

```
## [937] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [949] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [961] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [973] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [985] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [997] FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("cat.", 1001:1500, ".jpg")
file.copy(file.path(train_dir, fnames),
          file.path(validation_cats_dir))
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [217] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [229] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [277] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [301] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [325] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [337] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [349] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [373] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [385] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [433] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [445] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [457] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [469] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [481] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("cat.", 1501:2000, ".jpg")
file.copy(file.path(train_dir, fnames),
          file.path(test_cats_dir))
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [217] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [229] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [277] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [301] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [325] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [337] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [349] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [373] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [385] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [433] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [445] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [457] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [469] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [481] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("dog.", 1:1000, ".jpg")
file.copy(file.path(train_dir, fnames),
          file.path(train_dogs_dir))
```

file:///Users/hannahcronin/Desktop/GITHUB/HCRONIN_MACHINELEARNING/Assignment 2/Assignment2.html

```
## [937] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [949] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [961] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [973] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [985] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [997] FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("dog.", 1001:1500, ".jpg")
file.copy(file.path(train_dir, fnames),
          file.path(validation_dogs_dir))
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [217] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [229] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [277] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [301] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [325] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [337] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [349] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [373] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [385] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [433] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [445] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [457] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [469] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [481] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("dog.", 1501:2000, ".jpg")
file.copy(file.path(train_dir, fnames),
          file.path(test_dogs_dir))
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [217] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [229] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [277] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [301] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [325] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [337] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [349] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [373] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [385] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [433] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [445] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [457] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [469] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [481] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
cat("total training cat images:", length(list.files(train_cats_dir)), "\n")
```

```
## total training cat images: 1000
```

```
## total training cat images: 1000
cat("total training dog images:", length(list.files(train_dogs_dir)), "\n")
```

```
## total training dog images: 1000
```

```
## total training dog images: 1000
cat("total validation cat images:",length(list.files(validation_cats_dir)), "\n")
```

```
## total validation cat images: 500
```

```
## total validation cat images: 500
cat("total validation dog images:",length(list.files(validation_dogs_dir)), "\n")
```

```
## total validation dog images: 500
```

```
## total validation dog images: 500
cat("total test cat images:", length(list.files(test_cats_dir)), "\n")
```

```
## total test cat images: 500
```

```
## total test cat images: 500
cat("total test dog images:", length(list.files(test_dogs_dir)), "\n")
```



```
## total test dog images: 500
```

```
## total test dog images: 500
```

```
## Using data augmentation as model optimizer
datagen <- image_data_generator(
  rescale = 1/255,
  rotation_range = 40,
  width_shift_range = 0.2,
  height_shift_range = 0.2,
  shear_range = 0.2,
  zoom_range = 0.2,
  horizontal_flip = TRUE,
  fill_mode = "nearest"
)
```

```
model <- keras_model_sequential() %>%
  layer_conv_2d(filters = 16, kernel_size = c(3, 3), activation = "relu",
    input_shape = c(150, 150, 3)) %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 32, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_flatten() %>%
  layer_dropout(rate = 0.5) %>%
  layer_dense(units = 512, activation = "relu") %>%
  layer_dense(units = 1, activation = "sigmoid")
model %>% compile(
  loss = "binary_crossentropy",
  optimizer = optimizer_rmsprop(learning_rate = 0.001),
  metrics = c("acc")
)
```

```
##training the data
set.seed(100)
datagen <- image_data_generator(
  rescale = 1/255,
  rotation_range = 40,
  width_shift_range = 0.2,
  height_shift_range = 0.2,
  shear_range = 0.2,
  zoom_range = 0.2,
  horizontal_flip = TRUE
)
test_datagen <- image_data_generator(rescale = 1/255)
train_generator <- flow_images_from_directory(
  train_dir,
  datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
validation_generator <- flow_images_from_directory(
  validation_dir,
  test_datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
history <- model %>% fit(
  train_generator,
  steps_per_epoch = 50,
  epochs = 10,
  validation_data = validation_generator,
  validation_steps = 5
)
```

```
history
```

```
##
## Final epoch (plot to see history):
##     loss: 0.6411
##     acc: 0.638
## val_loss: 0.5913
## val_acc: 0.68
```

```
## This model had a validation accuracy of 0.7 and an accuracy of 0.687.
```

Adding more samples to training data.

```
r train_dir2a <- file.path(base_dir, "train2a") dir.create(train_dir2a)

## Warning in dir.create(train_dir2a): '/Users/hannahcronin/Desktop/ ## kaggle_original_data/train2a' already exists

r train_cats_dir2a <- file.path(train_dir2a, "cats2a") dir.create(train_cats_dir2a)

## Warning in dir.create(train_cats_dir2a): '/Users/hannahcronin/Desktop/ ## kaggle_original_data/train2a/cats2a' already exists

r train_dogs_dir2a <- file.path(train_dir2a, "dogs2a") dir.create(train_dogs_dir2a)

## Warning in dir.create(train_dogs_dir2a): '/Users/hannahcronin/Desktop/ ## kaggle_original_data/train2a/dogs2a' already exists

r fnames <- paste0("cat.", 1:1000, ".jpg") file.copy(file.path(train_dir2a, fnames), file.path(train_cats_dir2a))

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE ## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE

r fnames <- paste0("cat.", 2001:2500, ".jpg") file.copy(file.path(train_dir2a, fnames), file.path(train_cats_dir2a))

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE ## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE

r fnames <- paste0("dog.", 1:1000, ".jpg") file.copy(file.path(train_dir2a, fnames), file.path(train_dogs_dir2a))

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE ## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE

r fnames <- paste0("dog.", 2001:2500, ".jpg") file.copy(file.path(train_dir2a, fnames), file.path(train_dogs_dir2a))

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE ## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE

r cat("total training cat images:", length(list.files(train_cats_dir2a)), "\n")

## total training cat images: 1500

r ## total training cat images: 1500 cat("total training dog images:", length(list.files(train_dogs_dir2a)), "\n")

## total training dog images: 1500

r ## total training dog images: 1500

r ## Using data augmentation as model optimizer datagen <- image_data_generator( rescale = 1/255, rotation_range = 40, width_shift_re

r library(keras) model2a <- keras_model_sequential() %>% layer_conv_2d(filters = 16, kernel_size = c(3, 3), activation = "relu", input_shape = (180, 180, 3))

r ##training the data ## scaling training images set.seed(100) datagen <- image_data_generator( rescale = 1/255, rotation_range = 40, width_shift_re

r history2a

## ## Final epoch (plot to see history): ##     loss: 0.645 ##     acc: 0.631 ## val_loss: 0.7637 ## val_acc: 0.5

r ## Accuracy decreased overall when I increased the number of training images.
```

```
train_dir3a <- file.path(base_dir, "train3a")
dir.create(train_dir3a)
```

```
## Warning in dir.create(train_dir3a): '/Users/hannahcronin/Desktop/
## kaggle_original_data/train3a' already exists
```

```
train_cats_dir3a <- file.path(train_dir3a, "cats3a")
dir.create(train_cats_dir3a)
```

```
## Warning in dir.create(train_cats_dir3a): '/Users/hannahcronin/Desktop/
## kaggle_original_data/train3a/cats3a' already exists
```

```
train_dogs_dir3a <- file.path(train_dir3a, "dogs3a")
dir.create(train_dogs_dir3a)
```

```
## Warning in dir.create(train_dogs_dir3a): '/Users/hannahcronin/Desktop/
## kaggle_original_data/train3a/dogs3a' already exists
```

```
fnames <- paste0("cat.", 1:1000, ".jpg")  
file.copy(file.path(train_dir3a, fnames),  
          file.path(train_cats_dir3a))
```

12/21

```
## [937] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [949] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [961] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [973] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [985] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [997] FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("cat.", 2001:2200, ".jpg")
file.copy(file.path(train_dir3a, fnames),
          file.path(train_cats_dir3a))
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("dog.", 1:1000, ".jpg")
file.copy(file.path(train_dir3a, fnames),
          file.path(train_dogs_dir3a))
```

14/21

```
## [937] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [949] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [961] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [973] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [985] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [997] FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("dog.", 2001:2200, ".jpg")
file.copy(file.path(train_dir3a, fnames),
          file.path(train_dogs_dir3a))
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
cat("total training cat images:", length(list.files(train_cats_dir3a)), "\n")
```

```
## total training cat images: 1300
```

```
## total training cat images: 1300
cat("total training dog images:", length(list.files(train_dogs_dir3a)), "\n")
```

```
## total training dog images: 1300
```

```
## total training dog images: 1300
```

```
## Using data augmentation as model optimizer
datagen <- image_data_generator(
  rescale = 1/255,
  rotation_range = 40,
  width_shift_range = 0.2,
  height_shift_range = 0.2,
  shear_range = 0.2,
  zoom_range = 0.2,
  horizontal_flip = TRUE,
  fill_mode = "nearest"
)
```

```
library(keras)
model3a <- keras_model_sequential() %>%
  layer_conv_2d(filters = 16, kernel_size = c(3, 3), activation = "relu",
               input_shape = c(150, 150, 3)) %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 32, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_flatten() %>%
  layer_dropout(rate = 0.5) %>%
  layer_dense(units = 512, activation = "relu") %>%
  layer_dense(units = 1, activation = "sigmoid")
model3a %>% compile(
  loss = "binary_crossentropy",
  optimizer = optimizer_rmsprop(learning_rate = 0.001),
  metrics = c("acc")
)
```

```
##training the data
## scaling training images
set.seed(100)
datagen <- image_data_generator(
  rescale = 1/255,
  rotation_range = 40,
  width_shift_range = 0.2,
  height_shift_range = 0.2,
  shear_range = 0.2,
  zoom_range = 0.2,
  horizontal_flip = TRUE
)
## scaling testing images
test_datagen <- image_data_generator(rescale = 1/255)
train_generator3a <- flow_images_from_directory(
  train_dir3a,
  datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
validation_generator <- flow_images_from_directory(
  validation_dir,
  test_datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
history3a <- model3a %>% fit(
  train_generator3a,
  steps_per_epoch = 50,
  epochs = 10,
  validation_data = validation_generator,
  validation_steps = 5
)
```

```
history3a
```

```
##
## Final epoch (plot to see history):
##   loss: 0.6678
##   acc: 0.604
## val_loss: 0.7189
## val_acc: 0.54
```

Validation is 0.75 and training accuracy is 0.69. I think with more epochs and more steps per epoch, this number could rise. I've had to lower the number of steps and epochs for performance reasons. Initially, each run was taking ~15 minutes for my R browser to complete. Given the size of the file, I did not want the document to take over an hour to knit/export (as it may overheat my computer).

Using a pre-trained network 1)

Larger training dataset

```
r train_dir2b <- file.path(base_dir2, "train2b") datagen1b <- image_data_generator(rescale = 1/255) batch_size <- 20 extract_features
r model2b <- keras_model_sequential() %>% conv_base %>% layer_flatten() %>% layer_dense(units = 256, activation = "relu") %>% layer_c
r set.seed(100) train_datagen2b = image_data_generator( rescale = 1/255, rotation_range = 40, width_shift_range = 0.2, height_shift_r
r history2b

## ## Final epoch (plot to see history): ##   loss: 0.5735 ##   acc: 0.678 ## val_loss: 0.5682 ##   val_acc: 0.74

r ## Performance is slightly poorer with training data samples of 1500 each for cats & dogs. The model does seem to have some more va
```

Trying to improve model

```
train_dir3b <- file.path(base_dir2, "train3b")
dir.create(train_dir3b)
```

```
## Warning in dir.create(train_dir3b): '/Users/hannahcronin/Desktop/
## cats_and_dogs_small/train3b' already exists
```



```
train_cats_dir3b <- file.path(train_dir3b, "cats3b")  
dir.create(train_cats_dir3b)
```

```
## Warning in dir.create(train_cats_dir3b): '/Users/hannahcronin/Desktop/  
## cats_and_dogs_small/train3b/cats3b' already exists
```

```
train_dogs_dir3b <- file.path(train_dir3b, "dogs3b")  
dir.create(train_dogs_dir3b)
```

```
## Warning in dir.create(train_dogs_dir3b): '/Users/hannahcronin/Desktop/  
## cats_and_dogs_small/train3b/dogs3b' already exists
```

```
fnames <- paste0("cat.", 1:800, ".jpg")  
file.copy(file.path(train_dir3b, fnames),  
          file.path(train_cats_dir3b))
```

```
fnames <- paste0("dog.", 1:800, ".jpg")
file.copy(file.path(train_dir3b, fnames),
          file.path(train_dogs_dir3b))
```

[illegible]

```

train_dir3b <- file.path(base_dir2, "train3b")
datagen1b <- image_data_generator(rescale = 1/255)
batch_size <- 20
extract_features <- function(directory, sample_count) {
  features <- array(0, dim = c(sample_count, 4, 4, 512))
  labels <- array(0, dim = c(sample_count))
  generator <- flow_images_from_directory(
    directory = directory,
    generator = datagen,
    target_size = c(150, 150),
    batch_size = batch_size,
    class_mode = "binary"
  )
  i <- 0
  while(TRUE) {
    batch <- generator_next(generator)
    inputs_batch <- batch[[1]]
    labels_batch <- batch[[2]]
    features_batch <- conv_base %>% predict(inputs_batch)
    index_range <- ((i * batch_size)+1):((i + 1) * batch_size)
    features[index_range,,] <- features_batch
    labels[index_range] <- labels_batch
    i <- i + 1
    if (i * batch_size >= sample_count)
      break
  }
  list(
    features = features,
    labels = labels
  )
}
train3b <- extract_features(train_dir3b, 1600)
train3b$features <- reshape_features(train$features)

```

```

model3b <- keras_model_sequential() %>%
  conv_base %>%
  layer_flatten() %>%
  layer_dense(units = 256, activation = "relu") %>%
  layer_dense(units = 1, activation = "sigmoid")

```

```

set.seed(100)
train_datagen3b = image_data_generator(
  rescale = 1/255,
  rotation_range = 40,
  width_shift_range = 0.2,
  height_shift_range = 0.2,
  shear_range = 0.2,
  zoom_range = 0.2,
  horizontal_flip = TRUE,
  fill_mode = "nearest"
)
test_datagen3b <- image_data_generator(rescale = 1/255)
train_generator3b <- flow_images_from_directory(
  train_dir3b,
  train_datagen3b,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
validation_generator1b <- flow_images_from_directory(
  validation_dir1b,
  test_datagen1b,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
model3b %>% compile(
  loss = "binary_crossentropy",
  optimizer = optimizer_rmsprop(learning_rate = 0.001),
  metrics = c("accuracy")
)
history3b <- model %>% fit(
  train_generator3b,
  steps_per_epoch = 50,
  epochs = 10,
  validation_data = validation_generator1b,
  validation_steps = 5
)

```

```
history3b
```

```

##
## Final epoch (plot to see history):
##   loss: 0.56
##   acc: 0.724
## val_loss: 0.4988
## val_acc: 0.77

```

Performance increased with less data in the training samples. Accuracy was 0.79 and validation accuracy was 0.76. This suggests there may be some overfitting due to the smaller size of the sample data, but the accuracies were not consistently higher than validation accuracy.

Conclusion, with the pretrained covnet, the models had overall better performance. For the model's built from scratch, I tried to use less training data and got worse results, used more training data and also got worse results from the initial run. However, slightly more data (not as much as run 2) procured the best results of the runs.

When it came to the pretrained covnet, there was better results immediately. Again, more training data resulted in worse results. However, less training data gave the model better accuracy.

I am wondering if hyperparameters or if more epochs/steps per epoch were done in my "built" from scratch model that these results would be replicated.

Overall, less training data with pretrained worked better and a data samples from 1000-1300 worked best for my built from scratch model.