

# Assignment 2

Jacob Fabian

2023-03-26

```
library(reticulate)
use_condaenv('r-reticulate')
library(tensorflow)
library(keras)
```

```
original_dataset_dir <- "~/Downloads/dogs-vs-cats"
```

```
base_dir <- "~/Downloads/cats_and_dogs_small"
dir.create(base_dir)
```

```
## Warning in dir.create(base_dir):
## '/Users/jacobfabian/Downloads/cats_and_dogs_small' already exists
```

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

```
## Warning in dir.create(train_dir):
## '/Users/jacobfabian/Downloads/cats_and_dogs_small/train' already exists
```

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

```
## Warning in dir.create(validation_dir):
## '/Users/jacobfabian/Downloads/cats_and_dogs_small/validation' already exists
```

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

```
## Warning in dir.create(test_dir):
## '/Users/jacobfabian/Downloads/cats_and_dogs_small/test' already exists
```

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

```
## Warning in dir.create(train_cats_dir):
## '/Users/jacobfabian/Downloads/cats_and_dogs_small/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/jacobfabian/Downloads/cats_and_dogs_small/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/jacobfabian/Downloads/cats_and_dogs_small/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/jacobfabian/Downloads/cats_and_dogs_small/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/jacobfabian/Downloads/cats_and_dogs_small/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/jacobfabian/Downloads/cats_and_dogs_small/test/dogs' already exists
```

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

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

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

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

```
fnames <- paste0("dog.", 1001:1500, ".jpg")
```

```
file.copy(file.path(original_dataset_dir, fnames),
          file.path(validation_dogs_dir))
```

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

```
library(keras)
library(tensorflow)
```

```
model <- keras_model_sequential() %>%
  layer_conv_2d(filters = 32, 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 = 64, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 128, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 128, kernel_size = c(3, 3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_flatten() %>%
  layer_dense(units = 512, activation = "relu") %>%
  layer_dense(units = 1, activation = "sigmoid")
```

```
summary(model)
```

```
## Model: "sequential"
## -----
## Layer (type)                Output Shape                Param #
## -----
## conv2d_3 (Conv2D)           (None, 148, 148, 32)        896
## max_pooling2d_3 (MaxPooling2D) (None, 74, 74, 32)          0
## conv2d_2 (Conv2D)           (None, 72, 72, 64)          18496
## max_pooling2d_2 (MaxPooling2D) (None, 36, 36, 64)          0
## conv2d_1 (Conv2D)           (None, 34, 34, 128)         73856
## max_pooling2d_1 (MaxPooling2D) (None, 17, 17, 128)         0
## conv2d (Conv2D)             (None, 15, 15, 128)         147584
## max_pooling2d (MaxPooling2D) (None, 7, 7, 128)           0
## flatten (Flatten)           (None, 6272)                0
## dense_1 (Dense)             (None, 512)                 3211776
## dense (Dense)               (None, 1)                   513
## -----
## Total params: 3,453,121
## Trainable params: 3,453,121
## Non-trainable params: 0
## -----
```

```
model %>% compile(
  loss = "binary_crossentropy",
  optimizer = optimizer_rmsprop(learning_rate = 1e-4),
  metrics = c("acc")
)
```

```
train_datagen <- image_data_generator(rescale = 1/255)
validation_datagen <- image_data_generator(rescale = 1/255)
```

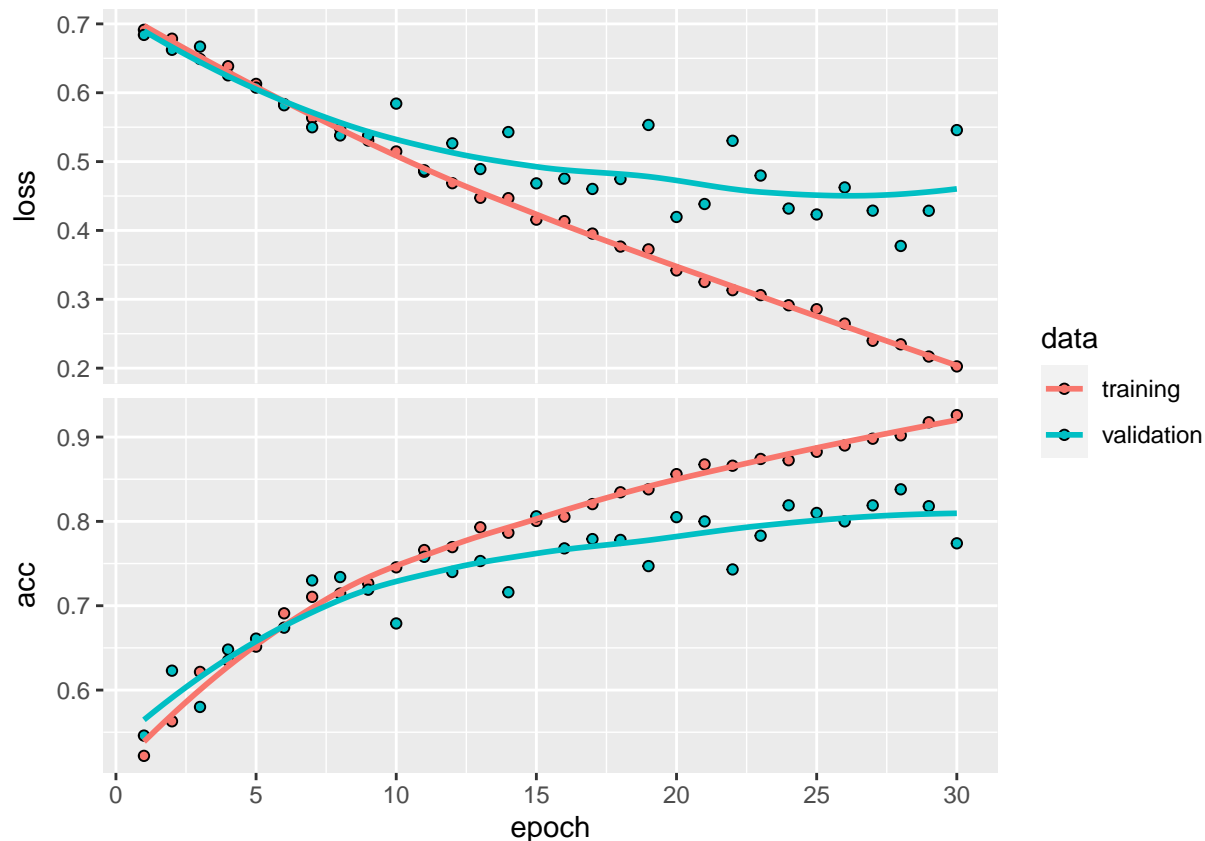
```
train_generator <- flow_images_from_directory(
  train_dir,
  train_datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)

validation_generator <- flow_images_from_directory(
  validation_dir,
  validation_datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
```

```
batch<- generator_next(train_generator)
str(batch)
```

```
## List of 2
## $ : num [1:20, 1:150, 1:150, 1:3] 0.4353 0.0902 0.4314 0.4824 0.1882 ...
## $ : num [1:20(1d)] 1 0 0 0 1 0 0 1 0 1 ...
```

```
history <- model %>% fit(
  train_generator,
  steps_per_epoch = 100,
  epochs = 30,
  validation_data = validation_generator,
  validation_steps = 50
)
plot(history)
```



Our first model is highly overfitted. We will try to increase accuracy and reduce overfitting.

## Question 1:

The following code improves upon the base model by adding a dropout layer. This should help the model to stop overfitting to try to improve performance.

```
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
)
fnames <- list.files(file.path(train_dir,"cats"), full.names = T)
img_path <- fnames[[round(runif(1,1,length(fnames)))]]
img <- image_load(img_path, target_size = c(150,150))
img_array <- image_to_array(img)
img_array <- array_reshape(img_array, c(1,150,150,3))
augmentation_generator <- flow_images_from_data(
  img_array,
  generator = datagen,
  batch_size = 1
)
```

```

)
op <- par(mfrow=c(2,2), pty="s", mar=c(1,0,.1,0))
for (i in 1:4) {
  batch <- generator_next(augmentation_generator)
  plot(as.raster(batch[1,,]))
}

```



```

par(op)

model2 <- keras_model_sequential() %>%
  layer_conv_2d(filters = 32, 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 = 64, kernel_size = c(3,3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 128, kernel_size = c(3,3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 128, 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")
model2 %>% compile(
  loss = "binary_crossentropy",
  optimizer = optimizer_rmsprop(learning_rate = 1e-4),

```

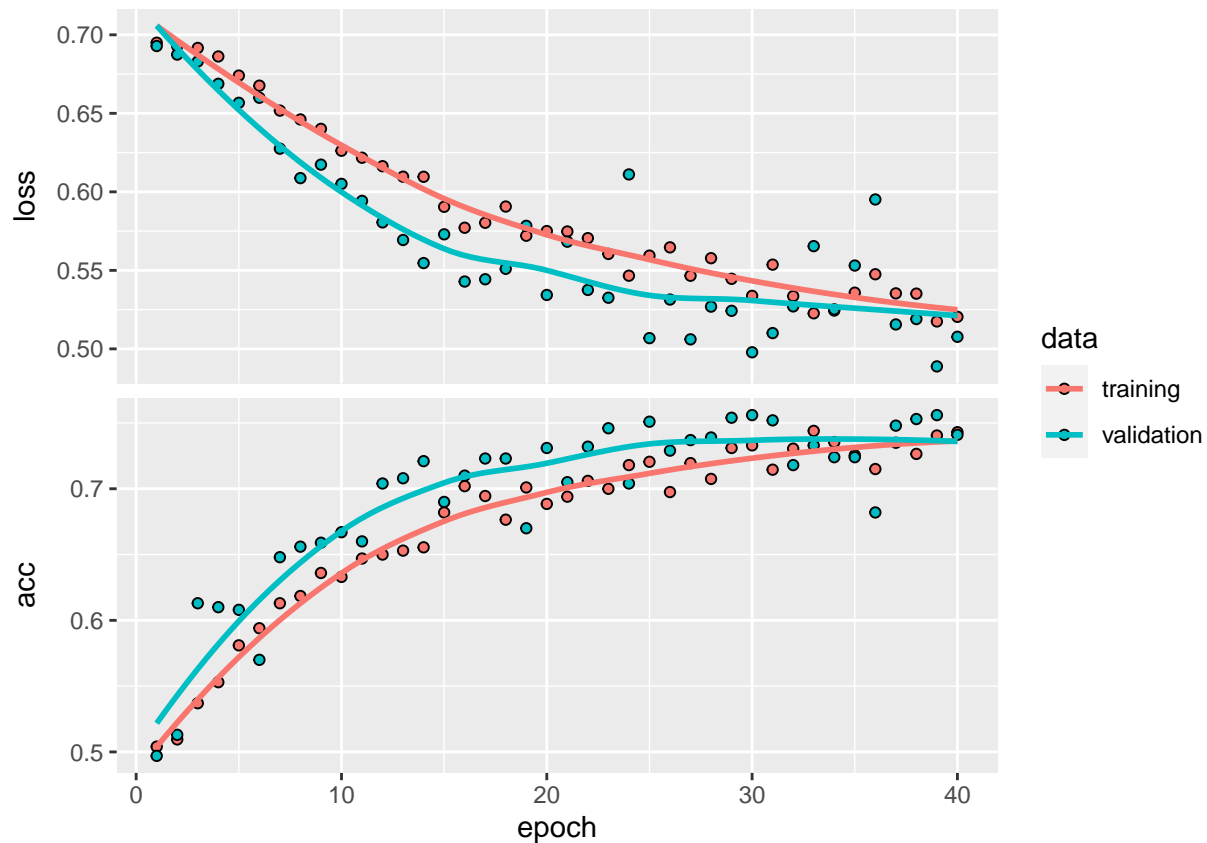
```

    metrics = c("acc")
)
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"
)
history2 <- model2 %>% fit(
  train_generator,
  steps_per_epoch = 100,
  epochs = 40,
  validation_data = validation_generator,
  validation_steps = 50
)
test_datagen <- image_data_generator(rescale = 1/255)

test_generator <- flow_images_from_directory(
  test_dir,
  test_datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
plot(history2)

```



```
model2 %>% evaluate_generator(test_generator, steps = 50)
```

```
##      loss      acc
## 0.5156532 0.7480000
```

40 epochs later, the loss is now .508, and the accuracy is now .77

## Question 2

Adding more pictures to the training data, while keeping the same number of validation and test images

```
dir.create(base_dir)
train_dir <- file.path(base_dir, "train")
dir.create(train_dir)
validation_dir <- file.path(base_dir, "validation")
dir.create(validation_dir)
test_dir <- file.path(base_dir, "test")
dir.create(test_dir)
train_cats_dir <- file.path(train_dir, "cats")
dir.create(train_cats_dir)
train_dogs_dir <- file.path(train_dir, "dogs")
dir.create(train_dogs_dir)
validation_cats_dir <- file.path(validation_dir, "cats")
```



[illegible]

[illegible]







[illegible]

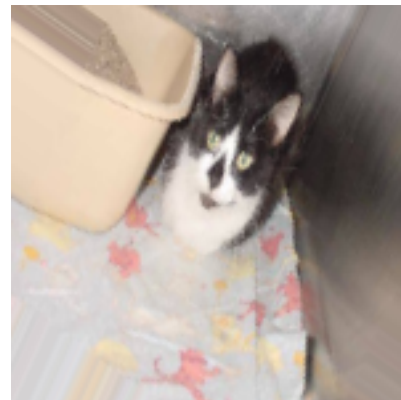






```
## [229] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [277] 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
## [301] 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
## [325] 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
## [349] 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
## [373] 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
## [397] 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
## [421] 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
## [445] 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
## [469] 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
## [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
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
)
fnames <- list.files(file.path(train_dir,"cats"), full.names = T)
img_path <- fnames[[round(runif(1,1,length(fnames))))]
img <- image_load(img_path, target_size = c(150,150))
img_array <- image_to_array(img)
img_array <- array_reshape(img_array, c(1,150,150,3))
augmentation_generator <- flow_images_from_data(
  img_array,
  generator = datagen,
  batch_size = 1
)
op <- par(mfrow=c(2,2), pty="s", mar=c(1,0,.1,0))
for (i in 1:4) {
  batch <- generator_next(augmentation_generator)
  plot(as.raster(batch[1,,]))
}
```



```

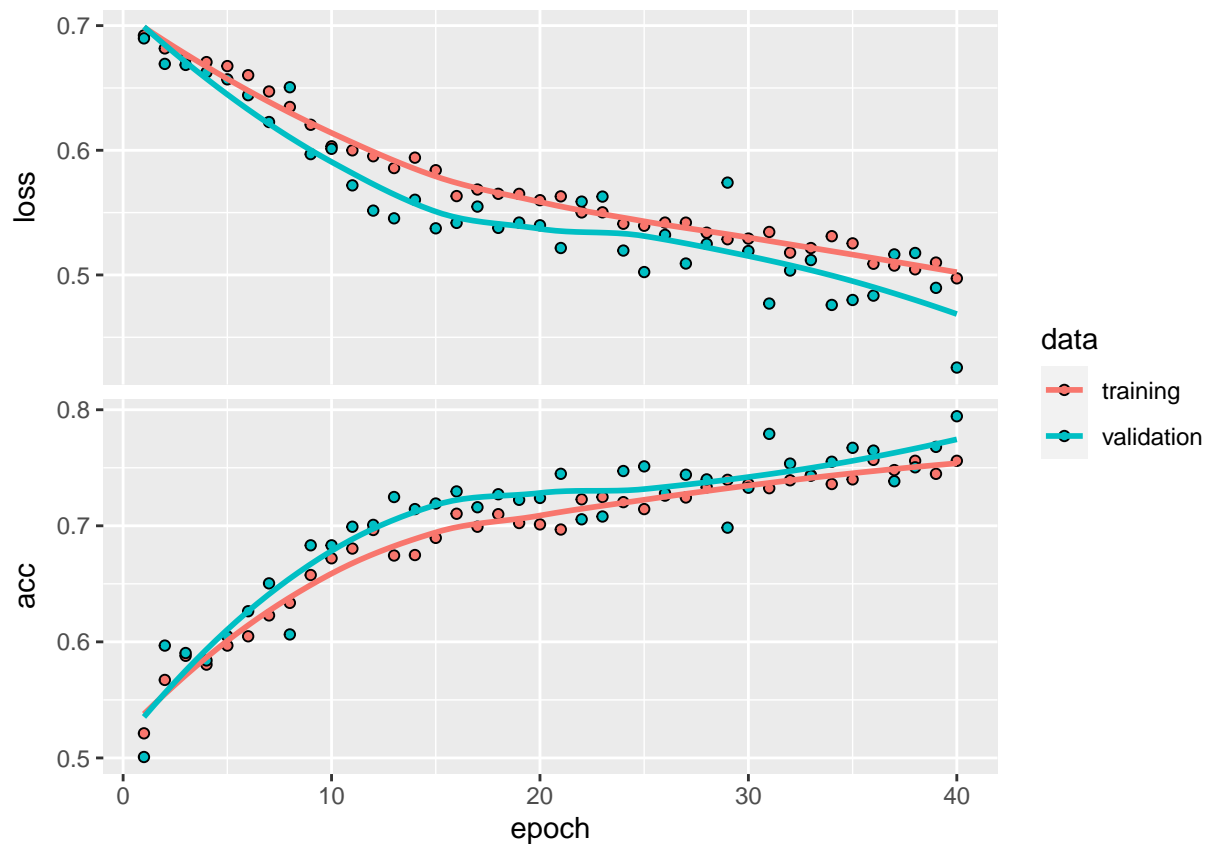
par(op)
model3 <- keras_model_sequential() %>%
  layer_conv_2d(filters = 32, kernel_size = c(3,3), activation = "relu", input_shape = c(150,150,3)) %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 64, kernel_size = c(3,3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 128, kernel_size = c(3,3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 128, 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")
model3 %>% compile(
  loss = "binary_crossentropy",
  optimizer = optimizer_rmsprop(learning_rate = 1e-4),
  metrics = c("acc")
)
test_datagen <- image_data_generator(rescale = 1/255)
train_generator <- flow_images_from_directory(
  train_dir,
  datagen,
  target_size = c(150, 150),
  batch_size = 25,
  class_mode = "binary"
)

```

```

)
validation_generator <- flow_images_from_directory(
  validation_dir,
  test_datagen,
  target_size = c(150, 150),
  batch_size = 25,
  class_mode = "binary"
)
history3 <- model3 %>% fit(
  train_generator,
  steps_per_epoch = 100,
  epochs = 40,
  validation_data = validation_generator,
  validation_steps = 50
)
test_generator <- flow_images_from_directory(
  test_dir,
  test_datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
plot(history3)

```



```
model3 %>% evaluate_generator(test_generator, steps = 50)
```

```
##      loss      acc  
## 0.4895864 0.7570000
```

This model has a loss of .485 and an accuracy of .76 . Small improvement over the model with less training images. Slightly lower loss and accuracy than the previous model.

## Question 3

The following model has the same number of training images, but increases by 500 of validation and testing images.

```
dir.create(base_dir)  
train_dir <- file.path(base_dir, "train")  
dir.create(train_dir)  
validation_dir <- file.path(base_dir, "validation")  
dir.create(validation_dir)  
test_dir <- file.path(base_dir, "test")  
dir.create(test_dir)  
train_cats_dir <- file.path(train_dir, "cats")  
dir.create(train_cats_dir)  
train_dogs_dir <- file.path(train_dir, "dogs")  
dir.create(train_dogs_dir)  
validation_cats_dir <- file.path(validation_dir, "cats")  
dir.create(validation_cats_dir)  
validation_dogs_dir <- file.path(validation_dir, "dogs")  
dir.create(validation_dogs_dir)  
test_cats_dir <- file.path(test_dir, "cats")  
dir.create(test_cats_dir)  
test_dogs_dir <- file.path(test_dir, "dogs")  
dir.create(test_dogs_dir)  
fnames <- paste0("cat.", 1:1500, ".jpg")  
file.copy(file.path(original_dataset_dir, fnames),  
          file.path(train_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
```

[illegible]

[illegible]

```
## [1477] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [1489] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
fnames <- paste0("cat.", 1501:2250, ".jpg")
file.copy(file.path(original_dataset_dir, fnames),
          file.path(validation_cats_dir))
```

[illegible]







[illegible]

[illegible]

```
fnames <- paste0("dog.", 1501:2250, ".jpg")
file.copy(file.path(original_dataset_dir, fnames),
          file.path(validation_dogs_dir))
```

[illegible]



```

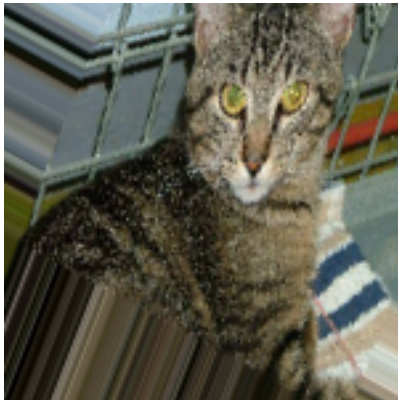
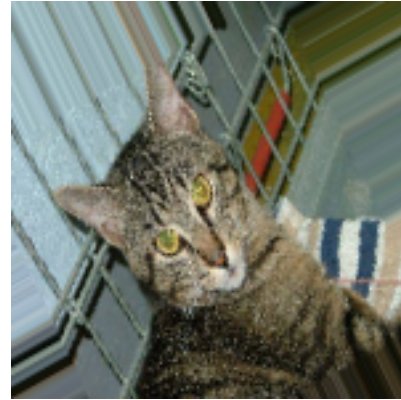
## [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [433] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [445] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [457] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [469] 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
## [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [505] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [517] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [529] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [541] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [553] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [565] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [577] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [589] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [601] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [613] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [625] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [637] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [649] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [661] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [673] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [685] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [697] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [709] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [721] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [733] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [745] FALSE FALSE FALSE FALSE FALSE FALSE

```

```

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
)
fnames <- list.files(file.path(train_dir,"cats"), full.names = T)
img_path <- fnames[[round(runif(1,1,length(fnames)))]
img <- image_load(img_path, target_size = c(150,150))
img_array <- image_to_array(img)
img_array <- array_reshape(img_array, c(1,150,150,3))
augmentation_generator <- flow_images_from_data(
  img_array,
  generator = datagen,
  batch_size = 1
)
op <- par(mfrow=c(2,2), pty="s", mar=c(1,0,.1,0))
for (i in 1:4) {
  batch <- generator_next(augmentation_generator)
  plot(as.raster(batch[1,,]))
}

```



```

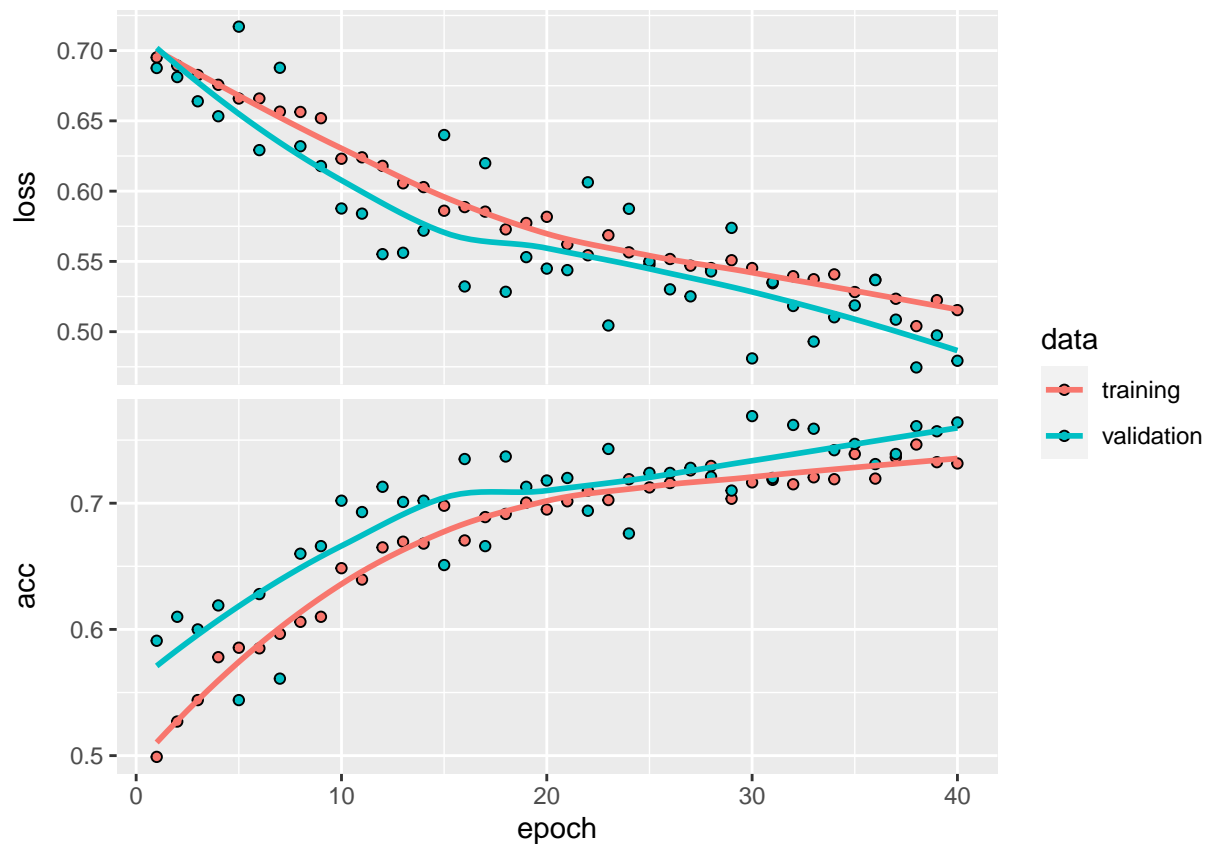
par(op)
model4 <- keras_model_sequential() %>%
  layer_conv_2d(filters = 32, kernel_size = c(3,3), activation = "relu", input_shape = c(150,150,3)) %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 64, kernel_size = c(3,3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 128, kernel_size = c(3,3), activation = "relu") %>%
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_conv_2d(filters = 128, 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")
model4 %>% compile(
  loss = "binary_crossentropy",
  optimizer = optimizer_rmsprop(learning_rate = 1e-4),
  metrics = c("acc")
)
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"
)
history4 <- model4 %>% fit(
  train_generator,
  steps_per_epoch = 100,
  epochs = 40,
  validation_data = validation_generator,
  validation_steps = 50
)
test_datagen <- image_data_generator(rescale = 1/255)
test_generator <- flow_images_from_directory(
  test_dir,
  test_datagen,
  target_size = c(150, 150),
  batch_size = 20,
  class_mode = "binary"
)
plot(history4)

```





```
model4 %>% evaluate_generator(test_generator, steps = 50)
```

```
##      loss      acc  
## 0.5248497 0.7420000
```

This model now has a loss of .515 and an accuracy of .747. This has a higher loss and slightly lower accuracy than the previous model. This is worse than the model with less training and validation data.

## Question 4:

The final model uses feature extraction

```
conv_base <- application_vgg16(  
  weights = "imagenet",  
  include_top = FALSE,  
  input_shape = c(150, 150, 3)  
)  
base_dir <- "~/Downloads/cats_and_dogs_small"  
train_dir <- train_dir <- file.path(base_dir, "train")  
validation_dir <- validation_dir <- file.path(base_dir, "validation")  
test_dir <- test_dir <- file.path(base_dir, "test")  
datagen <- 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  
  )  
}
```

```

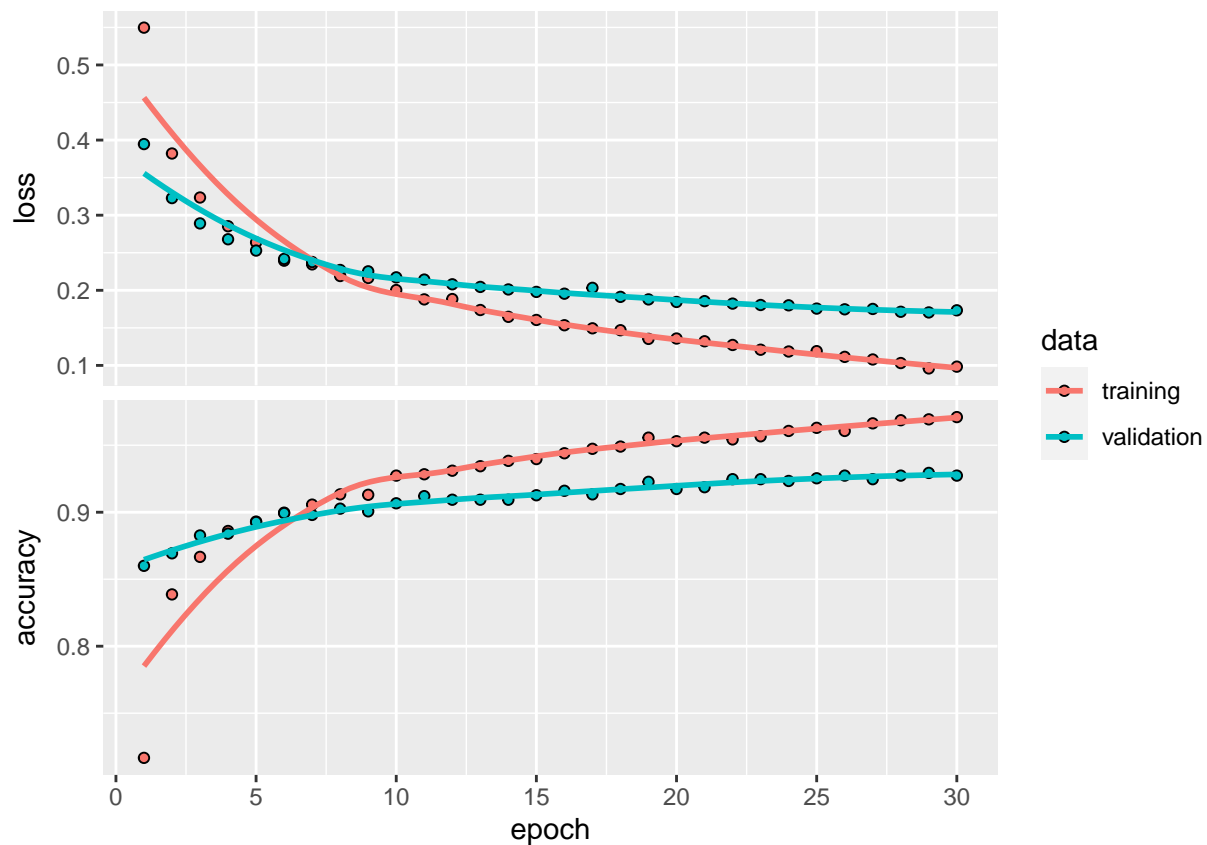
train <- extract_features(train_dir, 3000)
validation <- extract_features(validation_dir, 1500)
test <- extract_features(test_dir, 1500)

reshape_features <- function(features) {
  array_reshape(features, dim = c(nrow(features), 4 * 4 * 512))
}
train$features <- reshape_features(train$features)
validation$features <- reshape_features(validation$features)
test$features <- reshape_features(test$features)

model5 <- keras_model_sequential() %>%
  layer_dense(units = 256, activation = "relu", input_shape = 4 * 4 * 512) %>%
  layer_dropout(rate = 0.5) %>%
  layer_dense(units = 1, activation = "sigmoid")
model5 %>% compile(
  optimizer = optimizer_rmsprop(learning_rate = 2e-5),
  loss = "binary_crossentropy",
  metrics = c("accuracy")
)
History5 <- model5 %>% fit(
  train$features, train$labels,
  epochs = 30,
  batch_size = 20,
  validation_data = list(validation$features, validation$labels)
)

plot(History5)

```



```
model5 %>% evaluate(test$features, test$labels)
```

```
##      loss  accuracy
## 0.2569493 0.8960000
```

This model has a loss of .244 and an accuracy of .898. This has a lower loss and higher accuracy than model 4. This shows a higher level of overfitting than the previous models, but less overfitting than the first model.