ML-Tools-Assignment-2

Gyongyver Kamenar (2103380)

3/27/2022

```
library(tidyverse)
library(kableExtra)
library(keras)
library(ggplot2)
library(imager)

# Set my theme
devtools::source_url('https://raw.githubusercontent.com/gyongyver-droid/ceu-data-analysis/master/Assignmen
#theme_set(theme_gyongyver())
```

Problem 1

A)

What would be an appropriate metric to evaluate your models? Why?

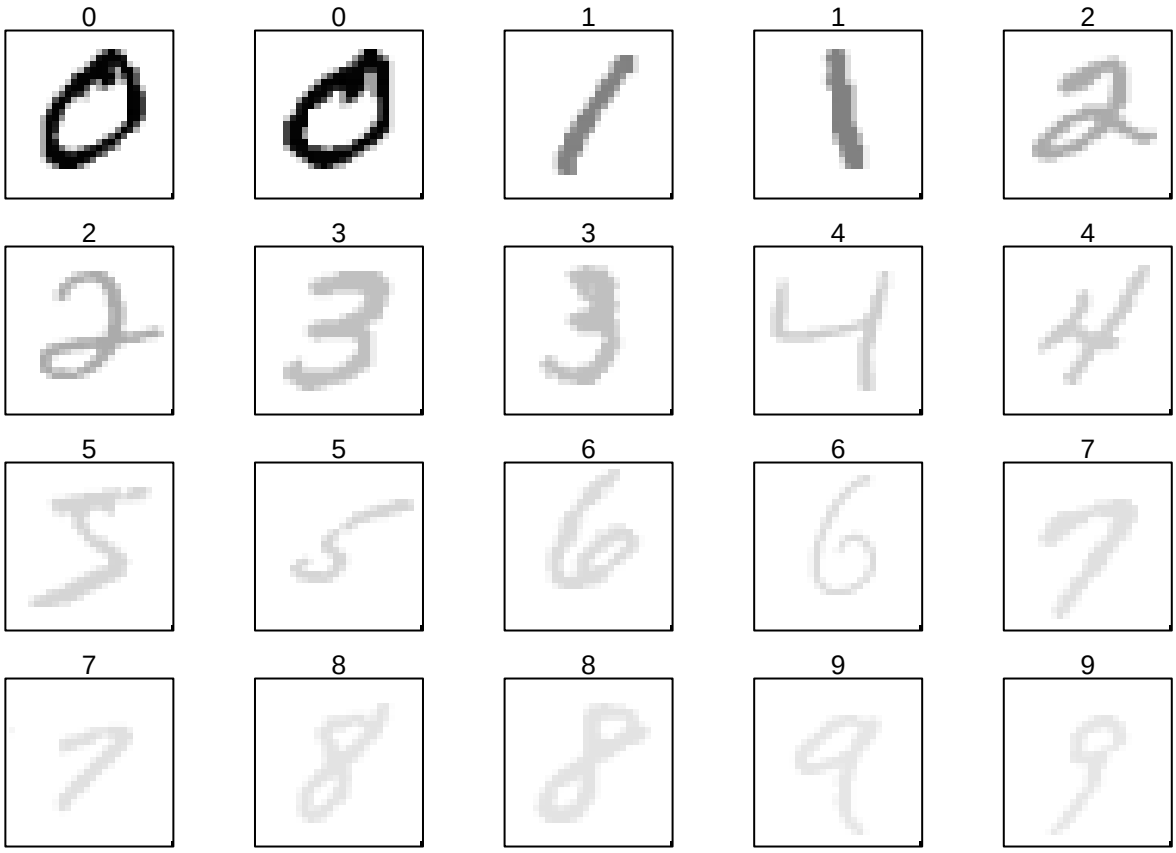
Accuracy would be an appropriate metric here, because it reports the correctly clasified images as the percentage of all and neglects the tradeoff between false classification. We can assume in this digit recognition project, that false classifications have equal loss. For example, we want to classify a digit 1 to label 1, but if it's not successful aka we did not classified it as 1, it does not matter which false label it has. I does not matter whether it's classified as 0,2,3,4 ... so on, anything other than 1. So the only thing matters to us it the rate of correctly classified cases, which is captured by accuracy.

B)

Get the data and show some example images from the data.

```
my_seed<-20200403
set.seed(my_seed)
# Get the data
data <-dataset_mnist()
ind <- sample(nrow(data$train$x),round(nrow(data$train$x)*0.8,0))

# train set
train_x<-as.matrix(as.tibble(data$train$x)[ind,])
train_y<-as.matrix(as.tibble(data$train$y)[ind,])

# validation set
valid_x<-as.matrix(as.tibble(data$train$x)[-ind,])
valid_y<-as.matrix(as.tibble(data$train$y)[-ind,])

# test set
test_x<-as.matrix(as.tibble(data$test$x))
test_y<-as.matrix(as.tibble(data$test$y))
```

```r
# Merge features with labels for the plot
train <-cbind(as.tibble(data$train$x),y=data$train$y)

# Factorizing and scaling for plotting
train <- train %>% mutate(
  y = as.factor(y),
  across(-y, ~./255)
)



# Plot the images
rotate <- function(x) t(apply(x, 2, rev))
par(mfrow=c(4, 5), pty='s', mai=c(0.1, 0, 0.15, 0.1))
for (lab in 0:9) {
  samp <- train %>%
    filter(y == lab)
  for (i in 1:2) {
    img <- matrix(as.numeric(samp[i, -1]), 28, 28, byrow = TRUE)
    image(t(rotate(img))[,28:1], axes = FALSE, col = grey(seq(1, 0, length = 256)),xlim = c(1,0))
    box(lty = 'solid')
    title(main = lab, font.main=22)
  }
}
```



C)

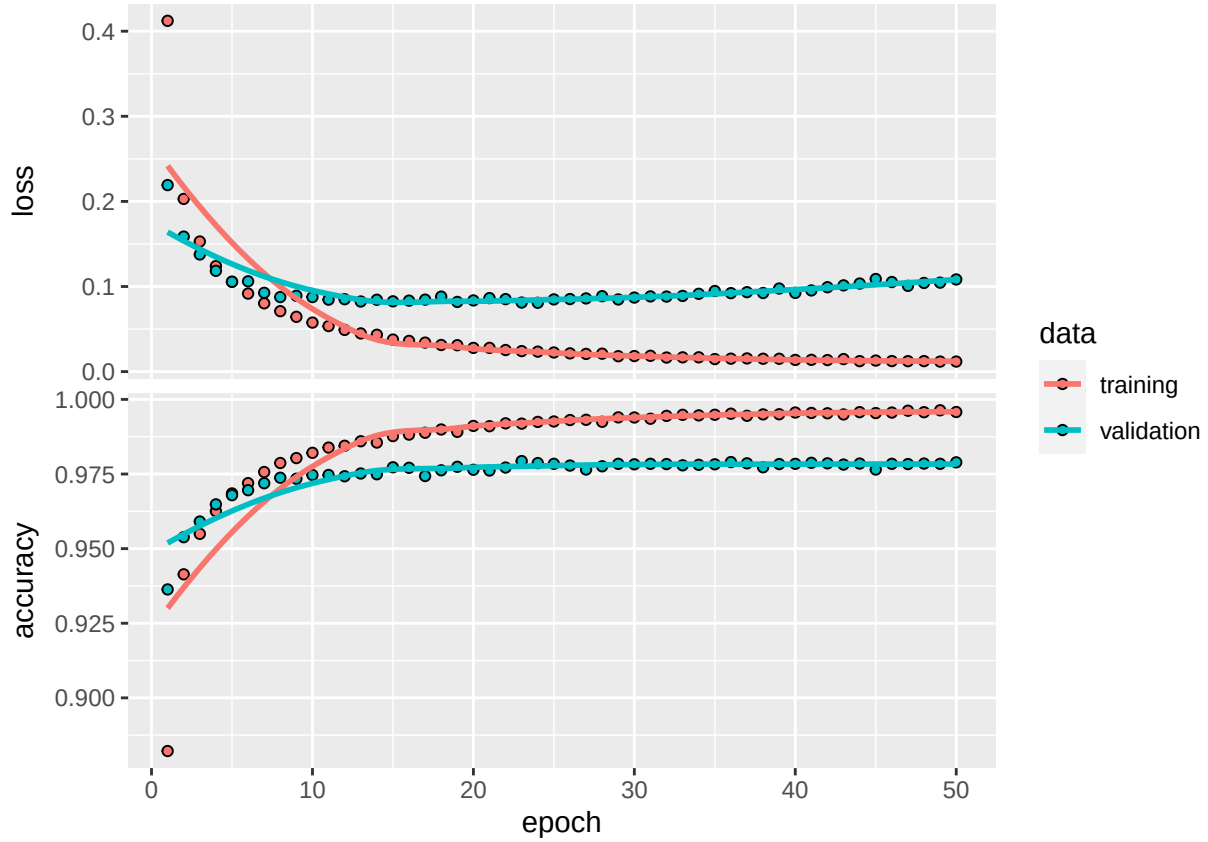
Train a simple fully connected network with a single hidden layer to predict the digits. Do not forget to normalize the data similarly to what we saw with FMNIST in class.

First, I normalized and converted the data into correct format. Then I build a model with 1 hidden layer using 128 neurons and the 'Relu' activation function. I used 0.2 dropout rate and output layer with 10 neurons using the softmax activation function. See the model summary below.

```
train_x<-array_reshape(train_x/255,c(nrow(train_x),784))
valid_x<-array_reshape(valid_x/255,c(nrow(valid_x),784))
test_x<-array_reshape(test_x/255,c(nrow(test_x),784))

train_y <- to_categorical(train_y, num_classes = 10)
valid_y <- to_categorical(valid_y, num_classes = 10)
test_y <- to_categorical(test_y, num_classes = 10)

simple_keras <- keras_model_sequential()
simple_keras |>
    layer_dense(units = 128, activation = 'relu', input_shape = c(784)) |>
    layer_dropout(rate = 0.2) |>
    layer_dense(units = 10, activation = 'softmax')

summary(simple_keras)
```

```
## Model: "sequential"
## ________________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## ================================================================================
##  dense_1 (Dense)                    (None, 128)                     100480
##
##  dropout (Dropout)                  (None, 128)                     0
##
##  dense (Dense)                      (None, 10)                      1290
##
## ================================================================================
## Total params: 101,770
## Trainable params: 101,770
## Non-trainable params: 0
## ________________________________________________________________________________
```

```
set.seed(my_seed)
compile(
    simple_keras,
    loss = 'categorical_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

digits1_history <-fit(
    simple_keras, train_x, train_y,
    epochs = 50, batch_size = 100,
    validation_data = list(valid_x, valid_y)
)

# Plot results
plot(digits1_history)
```




```
digits1_history
```

```
##
## Final epoch (plot to see history):
##         loss: 0.01179
##     accuracy: 0.9958
##     val_loss: 0.1083
## val_accuracy: 0.9789
```

```
res<-evaluate(simple_keras, valid_x, valid_y)
evaluate(simple_keras, valid_x, valid_y)
```

```
##      loss  accuracy
## 0.1083095 0.9789166
```

```
# Checking predictions and confusioon matrix
pred<-predict(simple_keras,valid_x)
#pred <- format(round(pred, 2), nsamll = 4)

#
# pred <- data.frame("0"=pred[,1], "1"=pred[,2], "2"=pred[,3], "3"=pred[,4], "4"=pred[,5], "5"=pred[,6], "
#
#
# which(pred[1,]==max(pred[1,]))
# result<-c()
# for(i in 1:nrow(pred)){
#   result[i]<-which(pred[i,]>0.5)-1
# }
#
#
```

```
#
# y<-c()
# for(i in 1:nrow(valid_y)){
#   y[i]<-which(valid_y[i,]>0.5)-1
# }
#
# cm<-caret::confusionMatrix(factor(result),factor(y))
# cm$table
```

The accuracy on the test set is 97.89 %. You can see the confusion matrix above.

D)

Experiment with different network architectures and settings (number of hidden layers, number of nodes, type and extent of regularization, etc.). Train at least 5 models. Explain what you have tried, what worked and what did not. Make sure that you use enough epochs so that the validation error starts flattening out - provide a plot about the training history.

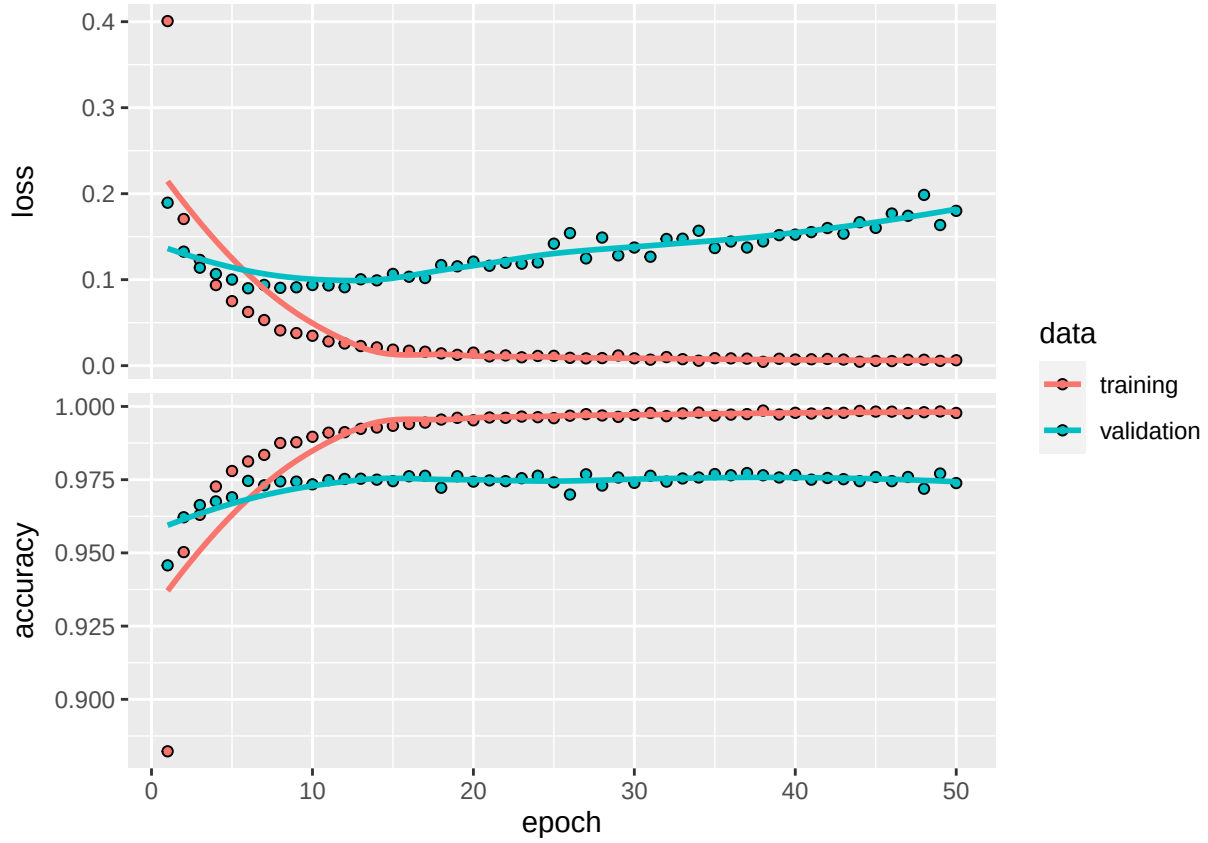
In my second model, I added one more hidden layer with 64 neurons and relu activation function. I expect that due to the plus layer, the model will perform better.

```
extended_keras1 <- keras_model_sequential()
extended_keras1 |>
    layer_dense(units = 128, activation = 'relu', input_shape = c(784)) |>
    layer_dense(units = 64, activation = 'relu', input_shape = c(784)) |>
    layer_dropout(rate = 0.2) |>
    layer_dense(units = 10, activation = 'softmax')

summary(extended_keras1)
```

```
## Model: "sequential_1"
## ________________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## ================================================================================
##  dense_4 (Dense)                    (None, 128)                     100480
##
##  dense_3 (Dense)                    (None, 64)                      8256
##
##  dropout_1 (Dropout)                (None, 64)                      0
##
##  dense_2 (Dense)                    (None, 10)                      650
##
## ================================================================================
## Total params: 109,386
## Trainable params: 109,386
## Non-trainable params: 0
## ________________________________________________________________________________
```

```
set.seed(my_seed)
compile(
    extended_keras1,
    loss = 'categorical_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

digits2_history <-fit(
    extended_keras1, train_x, train_y,
```

```
    epochs = 50, batch_size = 100,
    validation_data = list(valid_x, valid_y)
)

# Plot results

plot(digits2_history)
```




```
digits2_history
```

```
##
## Final epoch (plot to see history):
##         loss: 0.006307
##     accuracy: 0.9978
##     val_loss: 0.18
## val_accuracy: 0.9738
```

We can see, that a valuation accuracy increased from 0.9789 to 0.9738 in the second model. Besides, we can also see on the training plot, that the valuation accuracy reached this much earlier, it it steeper in the first few epochs.

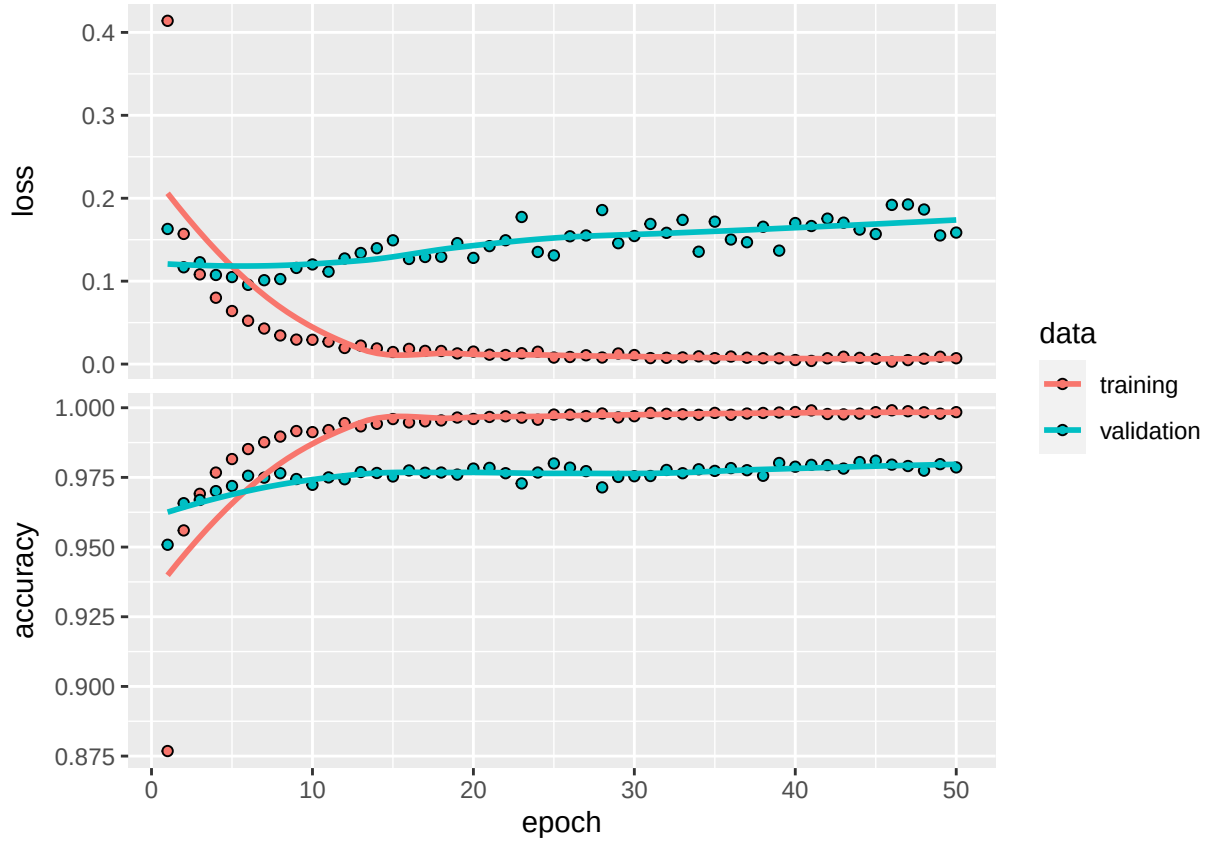
In my 3rd model, I add two more layers and more neurons, to test, whether it will similarly increase accuracy.

```
extended_keras2 <- keras_model_sequential()
extended_keras2 |>
    layer_dense(units = 256, activation = 'relu', input_shape = c(784)) |>
    layer_dense(units = 128, activation = 'relu', input_shape = c(784)) |>
    layer_dense(units = 64, activation = 'relu', input_shape = c(784)) |>
  layer_dense(units = 32, activation = 'relu', input_shape = c(784)) |>
    layer_dropout(rate = 0.2) |>
    layer_dense(units = 10, activation = 'softmax')
```

```
summary(extended_keras2)
```

```
## Model: "sequential_2"
## _________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## =========================================================================
##  dense_9 (Dense)                    (None, 256)                     200960
##
##  dense_8 (Dense)                    (None, 128)                     32896
##
##  dense_7 (Dense)                    (None, 64)                      8256
##
##  dense_6 (Dense)                    (None, 32)                      2080
##
##  dropout_2 (Dropout)                (None, 32)                      0
##
##  dense_5 (Dense)                    (None, 10)                      330
##
## =========================================================================
## Total params: 244,522
## Trainable params: 244,522
## Non-trainable params: 0
## _________________________________________________________________________
```

```
set.seed(my_seed)
compile(
    extended_keras2,
    loss = 'categorical_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

digits3_history <-fit(
    extended_keras2, train_x, train_y,
    epochs = 50, batch_size = 100,
    validation_data = list(valid_x, valid_y)
)

# Plot results

plot(digits3_history)
```




```
digits3_history
```

```
##
## Final epoch (plot to see history):
##         loss: 0.007052
##     accuracy: 0.9984
##     val_loss: 0.1586
## val_accuracy: 0.9786
```

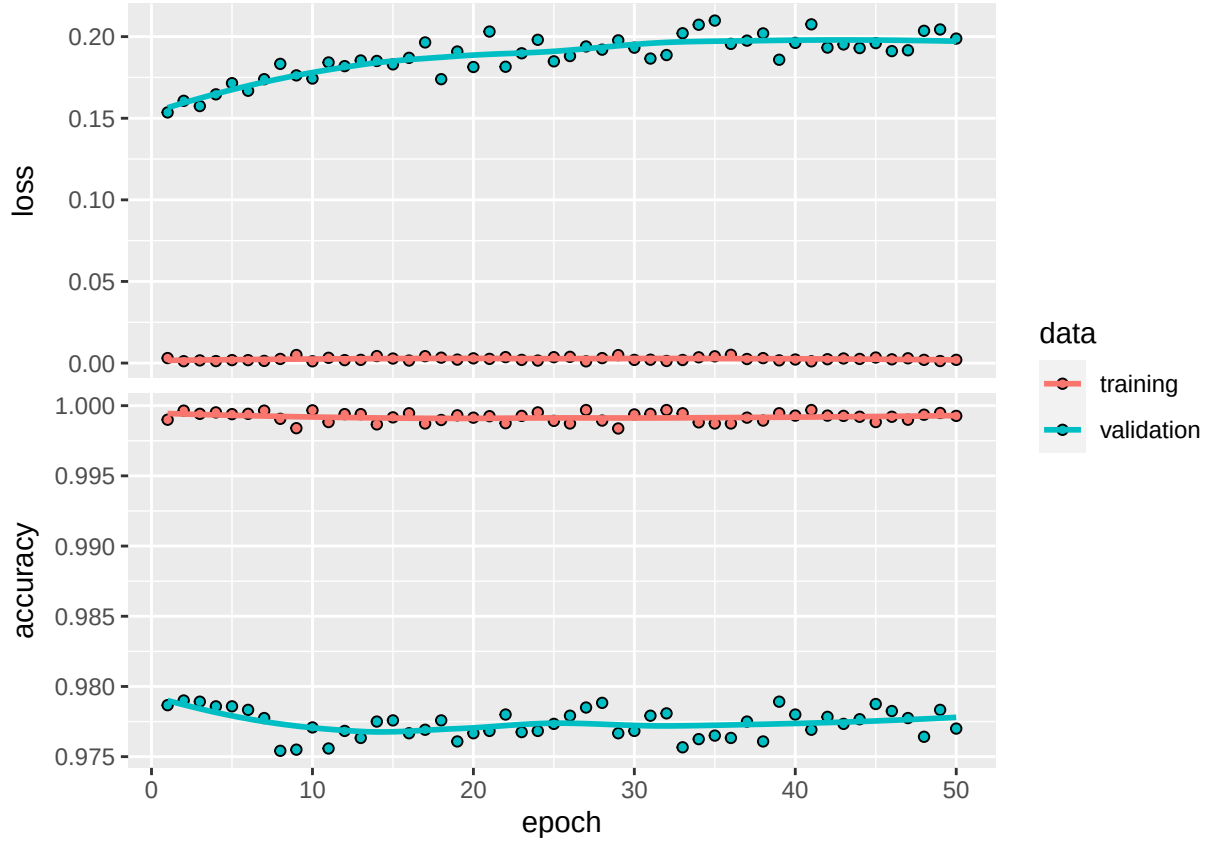
In the next model, I increase batch size from 100 to 200, and I use 2 hidden layers just like in the second model.

```
extended_keras3 <- keras_model_sequential()
extended_keras3 |>
    layer_dense(units = 128, activation = 'relu', input_shape = c(784)) |>
    layer_dense(units = 64, activation = 'relu', input_shape = c(784)) |>
    layer_dropout(rate = 0.2) |>
    layer_dense(units = 10, activation = 'softmax')

summary(extended_keras3)
```

```
## Model: "sequential_3"
## ________________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## ================================================================================
##  dense_12 (Dense)                   (None, 128)                     100480
##
##  dense_11 (Dense)                   (None, 64)                      8256
##
##  dropout_3 (Dropout)                (None, 64)                      0
##
```

```
##  dense_10 (Dense)                    (None, 10)                        650
##
## ================================================================================
## Total params: 109,386
## Trainable params: 109,386
## Non-trainable params: 0
## _________________________________________________________________________
```

```
set.seed(my_seed)
compile(
    extended_keras3,
    loss = 'categorical_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

digits4_history <-fit(
    extended_keras1, train_x, train_y,
    epochs = 50, batch_size = 200,
    validation_data = list(valid_x, valid_y)
)

# Plot results

plot(digits4_history)
```




```
digits4_history
```

```
##
## Final epoch (plot to see history):
```

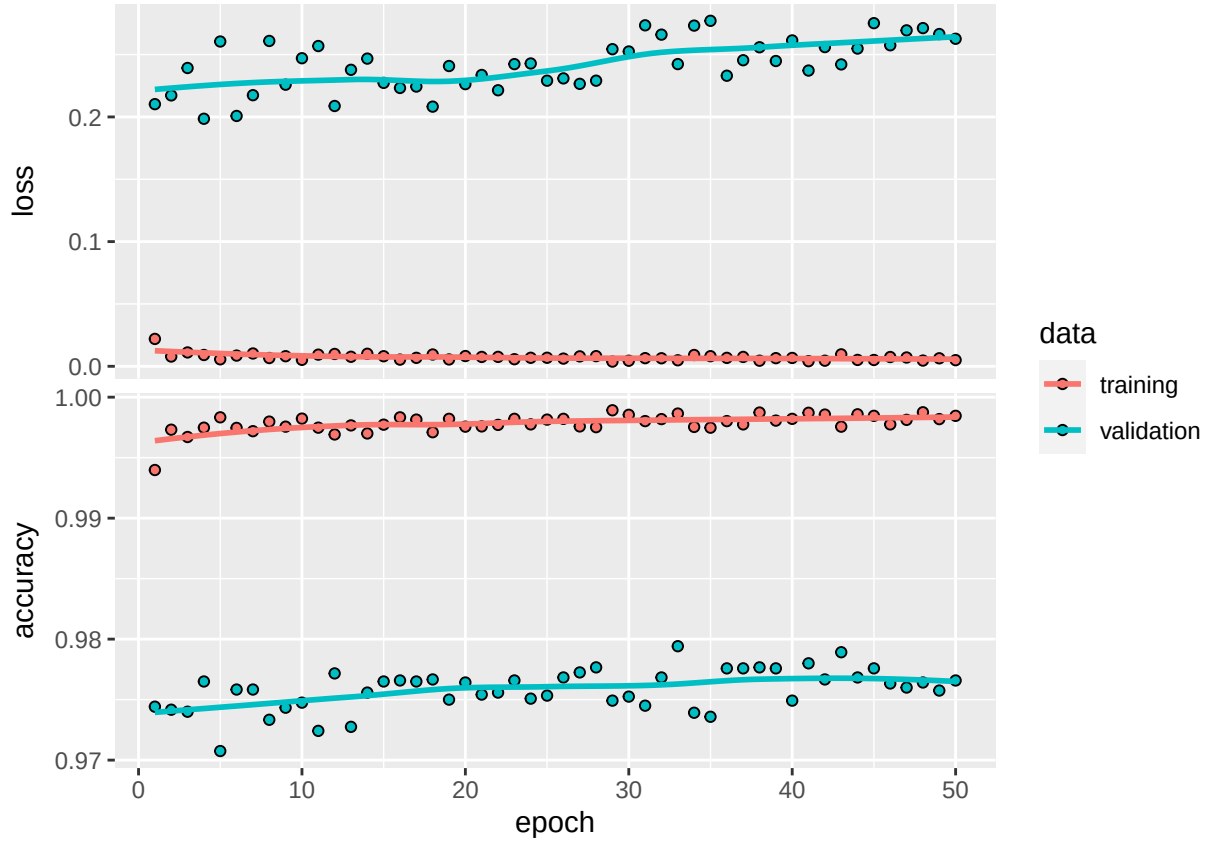
```
##         loss: 0.001981
##     accuracy: 0.9993
##     val_loss: 0.1988
## val_accuracy: 0.977
```

The bigger batch size clearly not helped the model. The valuation accuracy is continously decreasing. Therefore, I try the same with smaller batch size = 50.

```
extended_keras4 <- keras_model_sequential()
extended_keras4 |>
    layer_dense(units = 128, activation = 'relu', input_shape = c(784)) |>
    layer_dense(units = 64, activation = 'relu', input_shape = c(784)) |>
    layer_dropout(rate = 0.2) |>
    layer_dense(units = 10, activation = 'softmax')

summary(extended_keras4)
```

```
## Model: "sequential_4"
## _________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## =========================================================================
##  dense_15 (Dense)                   (None, 128)                     100480
##
##  dense_14 (Dense)                   (None, 64)                      8256
##
##  dropout_4 (Dropout)                (None, 64)                      0
##
##  dense_13 (Dense)                   (None, 10)                      650
##
## =========================================================================
## Total params: 109,386
## Trainable params: 109,386
## Non-trainable params: 0
## _________________________________________________________________________
```

```
set.seed(my_seed)
compile(
    extended_keras4,
    loss = 'categorical_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

digits5_history <-fit(
    extended_keras1, train_x, train_y,
    epochs = 50, batch_size = 50,
    validation_data = list(valid_x, valid_y)
)

# Plot results

plot(digits5_history)
```




```
digits5_history
```

```
##
## Final epoch (plot to see history):
##         loss: 0.0049
##     accuracy: 0.9985
##     val_loss: 0.2627
## val_accuracy: 0.9766
```

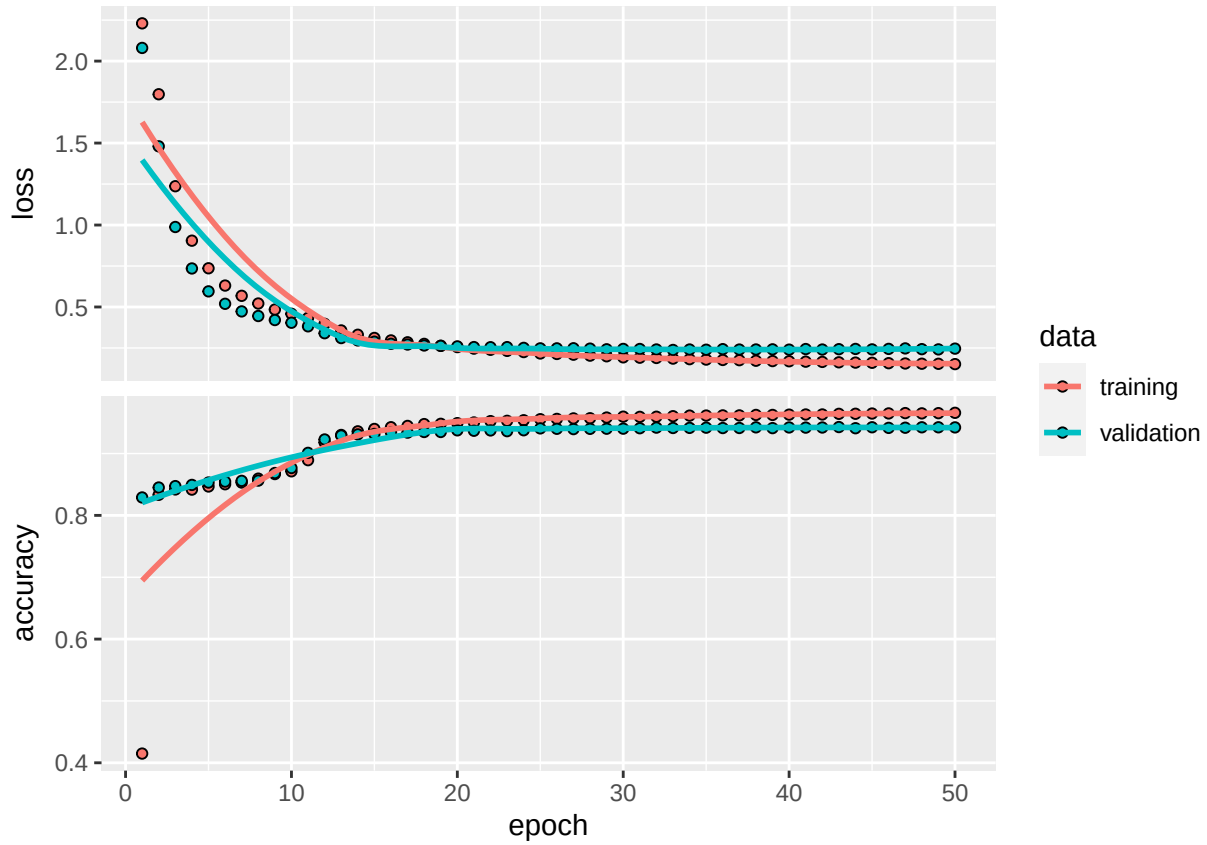
It seems like the smaller batch size also performed worse than model 2 wiith batch size=100, so it seems like 100 is about the appropriate value in this setting. In the 5th extended model, I try using other activation function, softmax instead of the previously used relu in the hidden layers.

```
extended_keras5 <- keras_model_sequential()
extended_keras5 |>
    layer_dense(units = 128, activation = 'softmax', input_shape = c(784)) |>
    layer_dense(units = 64, activation = 'softmax', input_shape = c(784)) |>
    layer_dropout(rate = 0.2) |>
    layer_dense(units = 10, activation = 'softmax')

summary(extended_keras5)
```

```
## Model: "sequential_5"
## ________________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## ================================================================================
##  dense_18 (Dense)                   (None, 128)                     100480
##
##  dense_17 (Dense)                   (None, 64)                      8256
##
```

```
##  dropout_5 (Dropout)                (None, 64)                      0
##
##  dense_16 (Dense)                   (None, 10)                      650
##
## ================================================================================
## Total params: 109,386
## Trainable params: 109,386
## Non-trainable params: 0
## --------------------------------------------------------------------------------
```

```
set.seed(my_seed)
compile(
    extended_keras5,
    loss = 'categorical_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

digits6_history <-fit(
    extended_keras5, train_x, train_y,
    epochs = 50, batch_size = 100,
    validation_data = list(valid_x, valid_y)
)

# Plot results

plot(digits6_history)
```



```
digits6_history
```

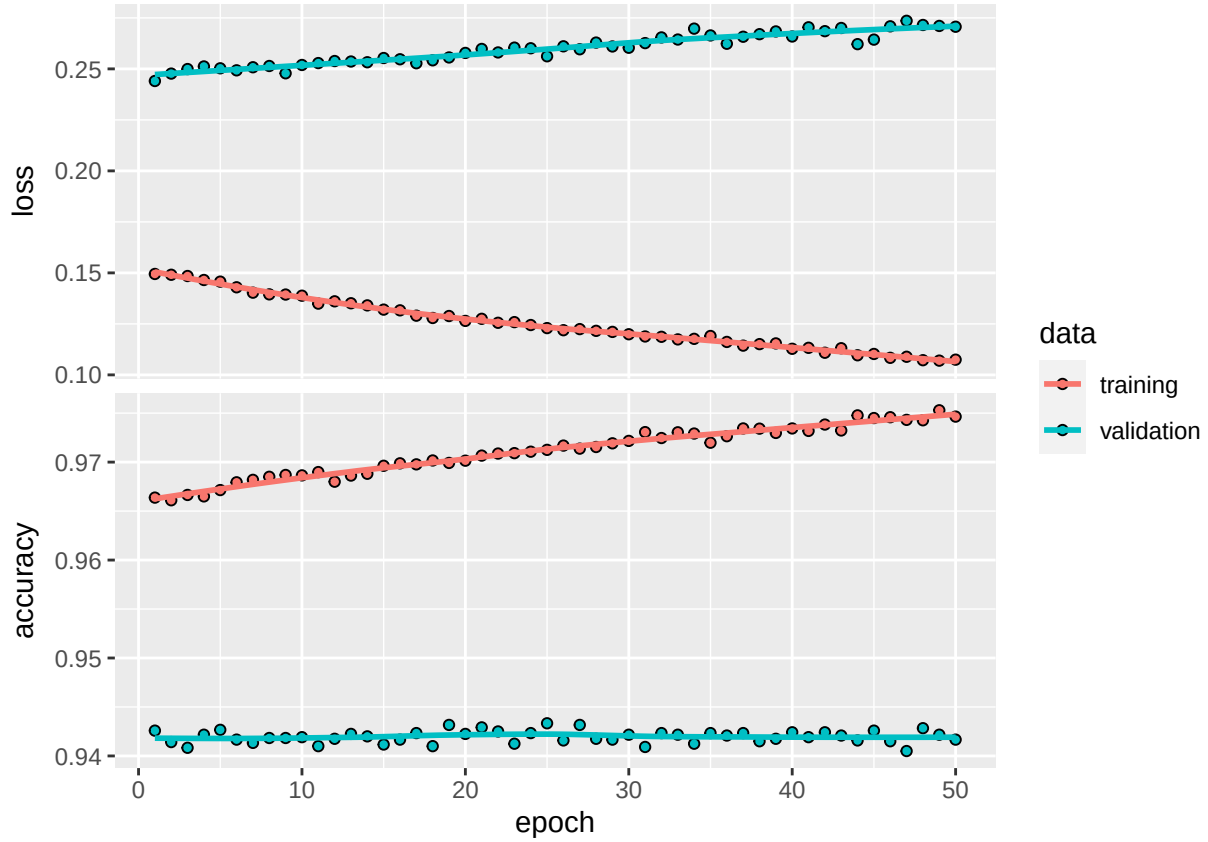
```
##
## Final epoch (plot to see history):
##         loss: 0.1515
##     accuracy: 0.9659
##     val_loss: 0.246
## val_accuracy: 0.9423
```

We can see from the accuracy measures, that the softmax activation function performes worse than the relu. So in my last model, I set all of them to relu and also change the optimaser from adam to rmsprop.

```
extended_keras6 <- keras_model_sequential()
extended_keras6 |>
    layer_dense(units = 128, activation = 'relu', input_shape = c(784)) |>
    layer_dense(units = 64, activation = 'relu', input_shape = c(784)) |>
    layer_dropout(rate = 0.2) |>
    layer_dense(units = 10, activation = 'relu')

summary(extended_keras6)
```

```
## Model: "sequential_6"
## ________________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## ================================================================================
##  dense_21 (Dense)                   (None, 128)                     100480
##
##  dense_20 (Dense)                   (None, 64)                      8256
##
##  dropout_6 (Dropout)                (None, 64)                      0
##
##  dense_19 (Dense)                   (None, 10)                      650
##
## ================================================================================
## Total params: 109,386
## Trainable params: 109,386
## Non-trainable params: 0
## ________________________________________________________________________________
```

```
set.seed(my_seed)
compile(
    extended_keras6,
    loss = 'categorical_crossentropy',
    optimizer = optimizer_rmsprop(),
    metrics = c('accuracy')
)

digits7_history <-fit(
    extended_keras5, train_x, train_y,
    epochs = 50, batch_size = 100,
    validation_data = list(valid_x, valid_y)
)

# Plot results

plot(digits7_history)
```




```
digits7_history
```

```
##
## Final epoch (plot to see history):
##         loss: 0.1075
##     accuracy: 0.9746
##     val_loss: 0.2707
## val_accuracy: 0.9417
```

E)

Choose a final model and evaluate it on the test set. How does test error compare to validation error?

My chose model is the first one, the simple model with 1 hidden layer. See the summary below.

```
summary(simple_keras)
```

```
## Model: "sequential"
## _________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## =========================================================================
##  dense_1 (Dense)                    (None, 128)                     100480
##
##  dropout (Dropout)                  (None, 128)                     0
##
##  dense (Dense)                      (None, 10)                      1290
##
## =========================================================================
## Total params: 101,770
## Trainable params: 101,770
```

```
## Non-trainable params: 0
## _________________________________________________________________________
```

Evaluation on the test set.

```
evaluate(simple_keras,test_x,test_y)
```

```
##      loss  accuracy
## 0.1062619 0.9792000
```

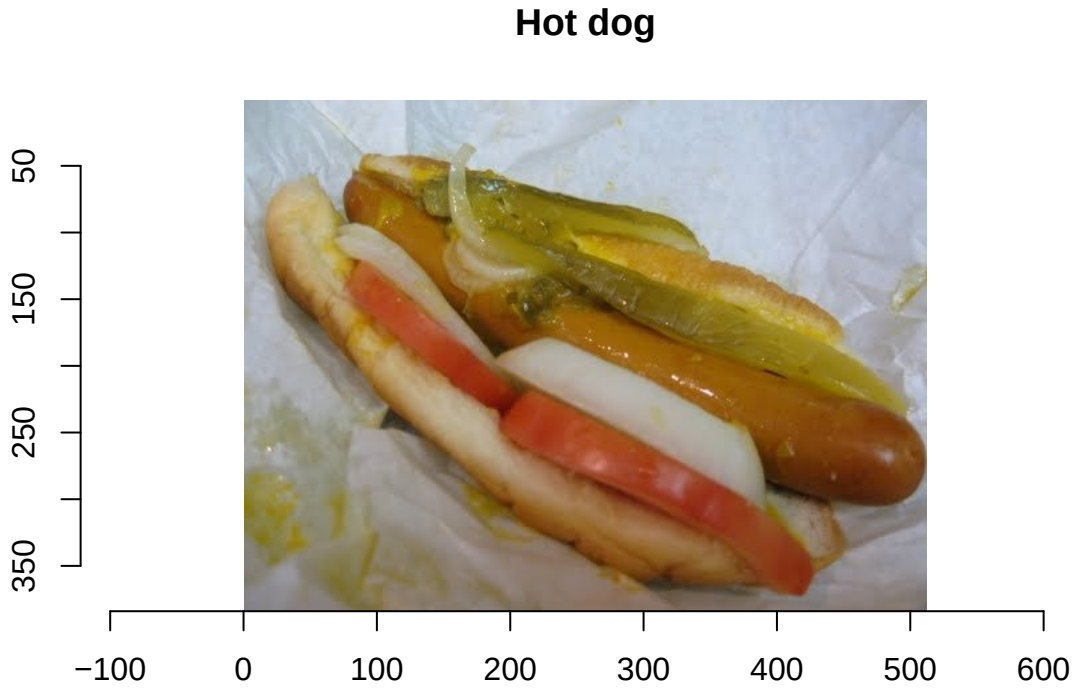
The test accuracy is higher than the validation accuracy on the model. It's still a bit lower than the training accuracy, because that is around 100%. It's a good sign that the test is higher than the validation accuracy, it can be partly due lack, partly due to the huge smaller sample size of the test. This model is certainly not overfitted on the validation set, so I would be confident using it.

Problem 2

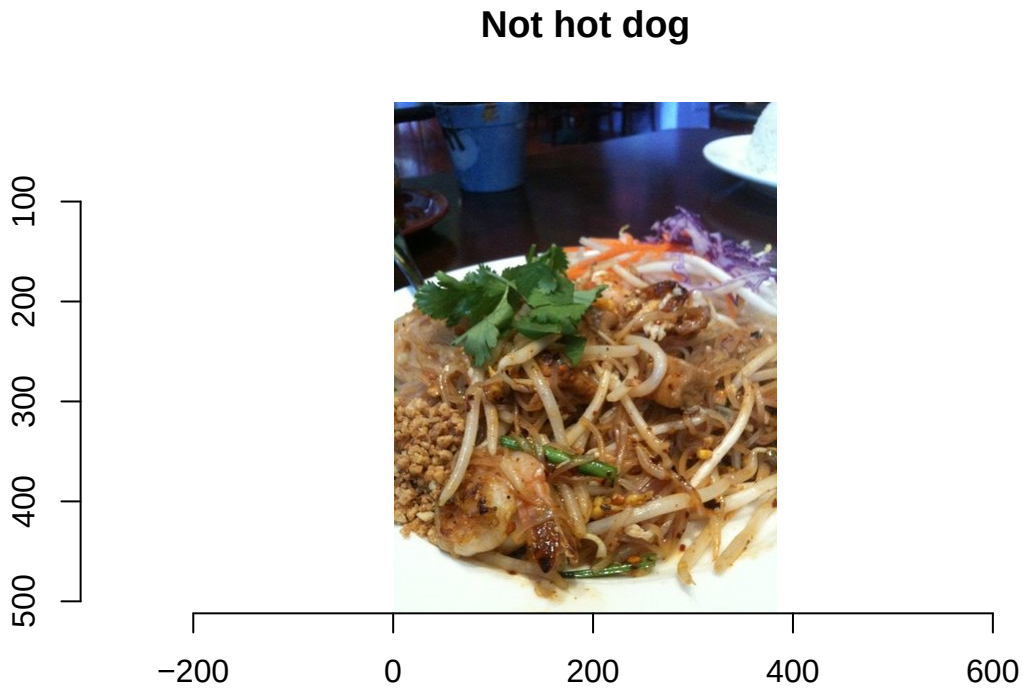
A)

Show two images: one hot dog and one not hot dog. (Hint: You may use knitr::include_graphics() or install the imager package to easily accomplish this.)

```
#Show hot dog
path<-"C:/CEU/DS1/ceu-ds1/data/train/"
hot_dog<-load.image(paste0(path,"hot_dog/7896.jpg"))
plot(hot_dog, main="Hot dog")
```




```
# Show not hot dog
not_hot_dog<-load.image(paste0(path,"not_hot_dog/4770.jpg"))
plot(not_hot_dog, main="Not hot dog")
```



B)

What would be a good metric to evaluate such a prediction?

C)

To be able to train a neural network for prediction, let's first build data batch generator functions as we did in class for data augmentation (train_generator and valid_generator).

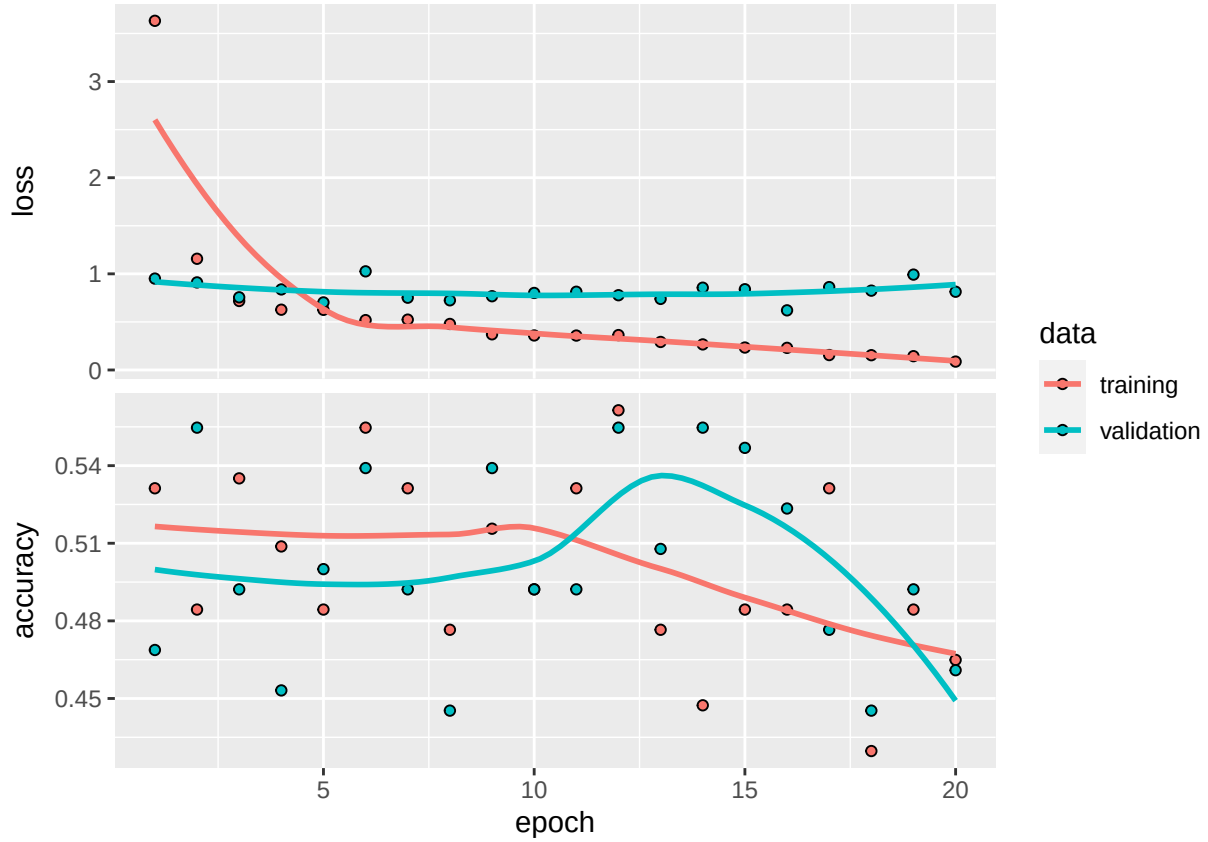
```
batch_size <- 16

train_datagen <- image_data_generator(rescale = 1/255)

valid_datagen <- image_data_generator(rescale = 1/255)

train_generator <- flow_images_from_directory(
    directory="C:/CEU/DS1/ceu-ds1/data/train/",
    class_mode = "binary",
    generator = train_datagen,
    batch_size = batch_size,
    target_size = c(128, 128)
)

valid_generator <- flow_images_from_directory(
    directory="C:/CEU/DS1/ceu-ds1/data/test/",
    class_mode = "binary",
    generator = valid_datagen,
    batch_size = batch_size,
    target_size = c(128, 128)
```

```
)
```

D)

Build a simple convolutional neural network (CNN) to predict if an image is a hot dog or not. Evaluate your model on the test set. (Hint: Account for the fact that you work with colored images in the input_shape parameter: set the third dimension to 3 instead of 1.)

```
cnn_model_wo_augmentation <- keras_model_sequential()
cnn_model_wo_augmentation |>
    layer_conv_2d(
        filters = 32,
        kernel_size = c(3, 3),
        activation = 'relu',
        input_shape = c(128, 128, 3)
    ) |>
    layer_max_pooling_2d(pool_size = c(2, 2)) |>
    layer_dropout(rate = 0.2) |>
    layer_flatten() |>
    layer_dense(units = 32, activation = 'relu') |>
    layer_dense(units = 1, activation = 'softmax')

compile(
    cnn_model_wo_augmentation,
    loss = 'binary_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

history1 <-fit(
    cnn_model_wo_augmentation,
    train_generator,
    epochs = 20,
    steps_per_epoch = 8, #nrow(data_train_x) / batch_size,  # this does not make a difference here -- batc
    validation_data = valid_generator,
    validation_steps = 8#nrow(data_valid_x) / batch_size
)

cnn_model_wo_augmentation %>% save_model_hdf5("hotdog_cnn1.h5")

plot(history1)
```




```
history1
```

```
##
## Final epoch (plot to see history):
##         loss: 0.08732
##     accuracy: 0.4649
##     val_loss: 0.8139
## val_accuracy: 0.4609
```

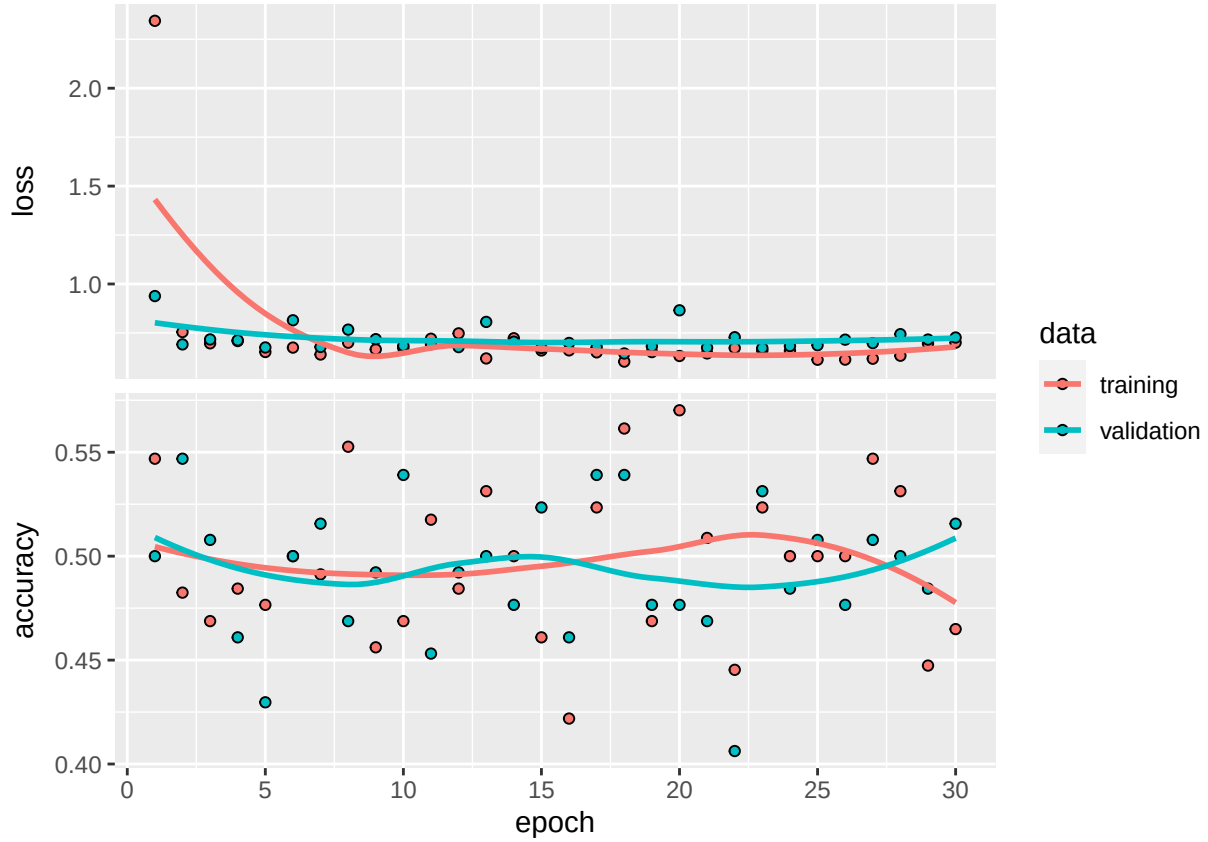
The train and test accuracy are both about 50% which means, that this binary classification model is not better than random :(.

E)

Could data augmentation techniques help with achieving higher predictive accuracy? Try some augmentations that you think make sense and compare to your previous model.

```
train_datagen <- image_data_generator(
  rescale = 1/255,
  rotation_range = 50,
  width_shift_range = 0.1,
  height_shift_range = 0.1,
  shear_range = 0.1,
  zoom_range = 0.1,
  horizontal_flip = TRUE,
  fill_mode = "nearest"
)
```

```
train_generator <- flow_images_from_directory(
    directory="C:/CEU/DS1/ceu-ds1/data/train/",
    class_mode = "binary",
    generator = train_datagen,
    batch_size = batch_size,
    target_size = c(128, 128)
)

valid_generator <- flow_images_from_directory(
    directory="C:/CEU/DS1/ceu-ds1/data/test/",
    class_mode = "binary",
    generator = valid_datagen,
    batch_size = batch_size,
    target_size = c(128, 128)
)

# try the same model, just with augmentation
cnn_model_w_augmentation <- keras_model_sequential()
cnn_model_w_augmentation |>
    layer_conv_2d(
        filters = 32,
        kernel_size = c(3, 3),
        activation = 'relu',
        input_shape = c(128, 128, 3)
    ) |>
    layer_average_pooling_2d(pool_size = c(2, 2)) |>
    layer_dropout(rate = 0.1) |>
    layer_flatten() |>
    layer_dense(units = 64, activation = 'relu') |>
    layer_dense(units = 1, activation = 'softmax')

compile(
    cnn_model_w_augmentation,
    loss = 'binary_crossentropy',
    optimizer = optimizer_adam(),
    metrics = c('accuracy')
)

history2 <-fit(
    cnn_model_w_augmentation,
    train_generator,
    epochs = 30,
    steps_per_epoch = 8, #nrow(data_train_x) / batch_size,  # this does not make a difference here -- batc
    validation_data = valid_generator,
    validation_steps = 8#nrow(data_valid_x) / batch_size
)

cnn_model_w_augmentation %>% save_model_hdf5("hotdog_cnn2.h5")

plot(history2)
```



Try the same model

```
history2
```

```
##
## Final epoch (plot to see history):
##         loss: 0.7004
##     accuracy: 0.4649
##     val_loss: 0.7266
## val_accuracy: 0.5156
```

optional

Try to rely on some pre-built neural networks to aid prediction. Can you achieve a better performance using transfer learning for this problem?

Using Resnet50 pre-trained model from keras

```
library(keras)

# instantiate the model
model <- application_resnet50(weights = 'imagenet')

# load the image
img_path <- "../data/train/hot_dog/7896.jpg"
img <- image_load(img_path, target_size = c(224,224))
x <- image_to_array(img)

# ensure we have a 4d tensor with single element in the batch dimension,
# the preprocess the input for prediction using resnet50
x <- array_reshape(x, c(1, dim(x)))
```

```
x <- imagenet_preprocess_input(x)

# make predictions then decode and print them
preds <- model %>% predict(x)
imagenet_decode_predictions(preds, top = 3)[[1]]
```

```
##   class_name class_description       score
## 1  n07697537            hotdog 9.999820e-01
## 2  n07697313      cheeseburger 1.745761e-05
## 3  n07693725             bagel 3.798117e-07
```