Homework 2

Adam Shelton

5/13/2019

Loading Data

```
congress_test = read_csv(here("data", "congress_test.csv"))

## Parsed with column specification:
## cols(
##   BillID = col_character(),
##   BillNum = col_double(),
##   Title = col_character(),
##   Major = col_double()
## )
congress_train = read_csv(here("data", "congress_train.csv"))

## Parsed with column specification:
## cols(
##   BillID = col_character(),
##   BillNum = col_double(),
##   Title = col_character(),
##   Major = col_double()
## )
congress_val = read_csv(here("data", "congress_val.csv"))

## Parsed with column specification:
## cols(
##   BillID = col_character(),
##   BillNum = col_double(),
##   Title = col_character(),
##   Major = col_double()
## )
```

Prepare Data

```
max_words = 10000
max_length = 100
x_train = congress_train$Title[1:270000]
y_train = congress_train$Major[1:270000] %>% as.numeric()
tokenizer_train = text_tokenizer(num_words = max_words) %>% fit_text_tokenizer(x_train)
sequences_train = texts_to_sequences(tokenizer_train, x_train)
word_indices_train = tokenizer_train$word_index
word_counts_train = tokenizer_train$word_counts
data_train = pad_sequences(sequences_train, max_length)

x_valid = congress_val$Title
y_valid = congress_val$Major %>% as.numeric()
tokenizer_valid = text_tokenizer(num_words = max_words) %>% fit_text_tokenizer(x_valid)
sequences_valid = texts_to_sequences(tokenizer_valid, x_valid)
```

```
word_indices_valid = tokenizer_valid$word_index
word_counts_valid = tokenizer_valid$word_counts
data_valid = pad_sequences(sequences_valid, max_length)
```

Initial Model

```
init_model = keras_model_sequential() %>%
  layer_embedding(input_dim = max_words, output_dim = length(unique(congress_train$Major)), input_lengt
  layer_dense(units = 1, activation = "sigmoid")

init_model %>% compile(optimizer = "rmsprop",
                       loss = "binary_crossentropy",
                       metrics = c("acc")
)

summary(init_model)
```

```
## _____
## Layer (type)                       Output Shape                  Param #
## ========================================================================
## embedding (Embedding)              (None, 100, 21)               210000
## _____
## flatten (Flatten)                  (None, 2100)                  0
## _____
## dense (Dense)                      (None, 1)                     2101
## ========================================================================
## Total params: 212,101
## Trainable params: 212,101
## Non-trainable params: 0
## _____
```

```
init_history = init_model %>% fit(data_train, y_train, epochs = 20, batch_size = 32, validation_data = 

plot(init_history)
```