**Assignment 4: Text and Sequence:**

Steven Kalinoff  
4/07/2024

Adv. Machine Learning 64061

Baseline performance

A graph with blue dots

Description automatically generatedA graph with blue dots

Description automatically generated

Top 150 words 100 samples

from tensorflow.keras.datasets import imdb

(train\_data, train\_labels), (test\_data, test\_labels) = imdb.load\_data(

    num\_words=150)

A graph with blue dots

Description automatically generatedA graph with blue dots

Description automatically generated

Embedding Layer @ 150/100

from tensorflow import keras

from tensorflow.keras import layers

max\_tokens = max([max(sequence) for sequence in train\_data]) + 3

embedding\_layer = layers.Embedding(input\_dim=max\_tokens, output\_dim=256)

inputs = keras.Input(shape=(None,), dtype="int64")

embedded = layers.Embedding(input\_dim=max\_tokens, output\_dim=256)(inputs)

x = layers.Bidirectional(layers.LSTM(32))(embedded)

x = layers.Dropout(0.5)(x)

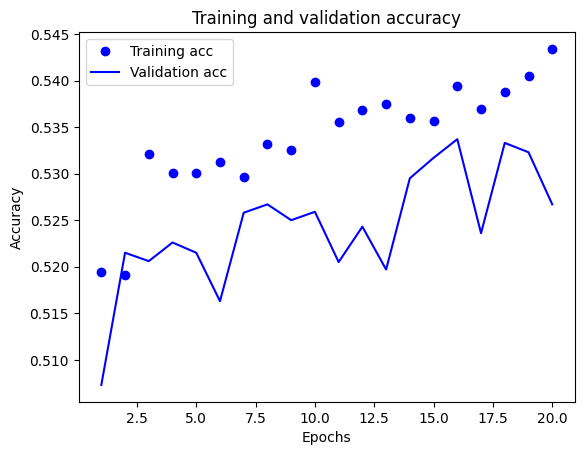
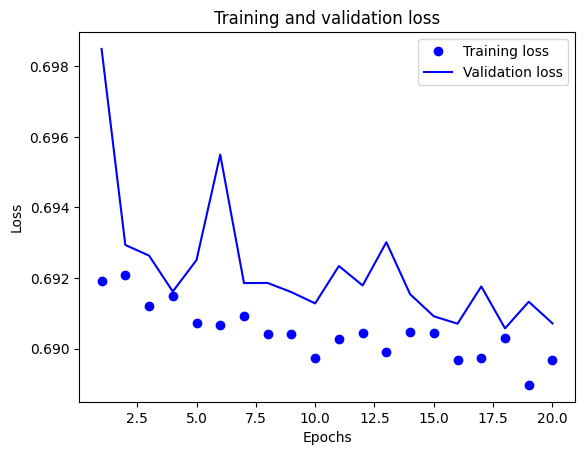
outputs = layers.Dense(1, activation="sigmoid")(x)

model = keras.Model(inputs, outputs)

model.compile(optimizer="rmsprop",

              loss="binary\_crossentropy",

              metrics=["accuracy"])



Pretrained word embedding

#pretrained

import tensorflow\_hub as hub

text\_vectorization = TextVectorization(

    ngrams=2,

    max\_tokens=20000,

    output\_mode="multi\_hot",

)

import numpy as np

path\_to\_glove\_file = "glove.6B.100d.txt"

embeddings\_index = {}

with open(path\_to\_glove\_file) as f:

    for line in f:

        word, coefs = line.split(maxsplit=1)

        coefs = np.fromstring(coefs, "f", sep=" ")

        embeddings\_index[word] = coefs

print(f"Found {len(embeddings\_index)} word vectors.")

from tensorflow import keras

from tensorflow.keras import layers

