Getting started with neural networks: Classification and regression

Classifying movie reviews: A binary classification example

The IMDB dataset

Loading the IMDB dataset

```
from tensorflow.keras.datasets import imdb
(train_data, train_labels), (test_data, test_labels) = imdb.load_data(
    num_words=10000)
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb.npz
17464789/17464789 [==============================] - 2s 0us/step
```

```
train_data[0]
```

```
[1,
 14,
 22,
 16,
 43,
 530,
 973,
 1622,
 1385,
 65,
 458,
 4468,
 66,
 3941,
 4,
 173,
 36,
 256,
 5,
 25,
 100,
 43,
 838,
 112,
 50,
 670,
 2,
 9,
 35,
 480,
 284,
 5,
 150,
 4,
 172,
 112,
 167,
 2,
 336,
 385,
 39,
 4,
 172,
 4536,
 1111,
 17,
 546,
 38,
 13,
 447,
 4,
 192,
 50,
 16,
 6,
 147,
 2025,
 19,
```

```
train_labels[0]
```

```
1
```

```
max([max(sequence) for sequence in train_data])
```

```
9999
```

Decoding reviews back to text

```
word_index = imdb.get_word_index()
reverse_word_index = dict(
    [(value, key) for (key, value) in word_index.items()])
decoded_review = " ".join(
    [reverse_word_index.get(i - 3, "?") for i in train_data[0]])
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb_word_index.json
1641221/1641221 [==============================] - 1s 1us/step
```

Preparing the data

Encoding the integer sequences via multi-hot encoding

```
import numpy as np
def vectorize_sequences(sequences, dimension=10000):
    results = np.zeros((len(sequences), dimension))
    for i, sequence in enumerate(sequences):
        for j in sequence:
            results[i, j] = 1.
    return results
x_train = vectorize_sequences(train_data)
x_test = vectorize_sequences(test_data)
```

```
x_train[0]
```

```
array([0., 1., 1., ..., 0., 0., 0.])
```

```
y_train = np.asarray(train_labels).astype("float32")
y_test = np.asarray(test_labels).astype("float32")
```

Building your model

Model definition and compilation

```
from tensorflow import keras
from tensorflow.keras import layers
model = keras.Sequential([
layers.Dense(64, activation="relu"),
layers.Dense(64, activation="relu"),
layers.Dense(64, activation="relu"),
layers.Dropout(0.5),
layers.Dense(1, activation="sigmoid")
])
```

```
model.compile(optimizer="rmsprop",
loss="mean_squared_error",
metrics=["accuracy"])
```

Validating the data set

```
x_val = x_train[:10000]
partial_x_train = x_train[10000:]
y_val = y_train[:10000]
partial_y_train = y_train[10000:]
```

Training model

```
history = model.fit(partial_x_train,
partial_y_train,
epochs=20,
batch_size=512,
validation_data=(x_val, y_val))
```

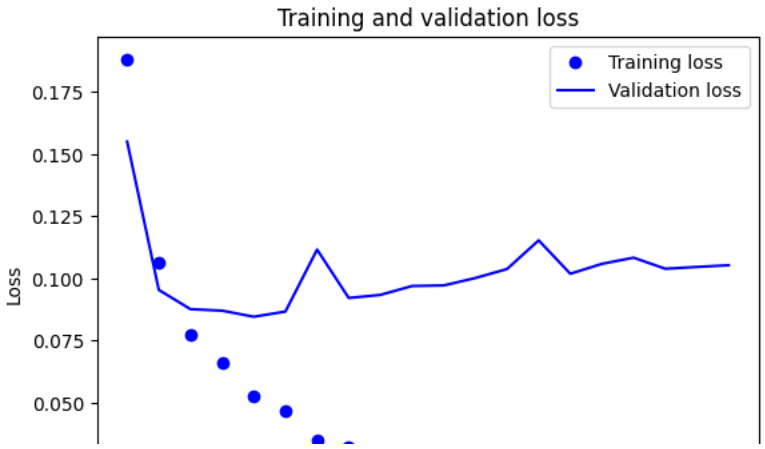
```
Epoch 1/20
30/30 [==============================] - 8s 83ms/step - loss: 0.1881 - accuracy: 0.7223 - val_loss: 0.1550 - val_accuracy: 0.7749
Epoch 2/20
30/30 [==============================] - 1s 33ms/step - loss: 0.1064 - accuracy: 0.8630 - val_loss: 0.0954 - val_accuracy: 0.8697
Epoch 3/20
30/30 [==============================] - 1s 26ms/step - loss: 0.0776 - accuracy: 0.9025 - val_loss: 0.0876 - val_accuracy: 0.8795
Epoch 4/20
30/30 [==============================] - 1s 25ms/step - loss: 0.0660 - accuracy: 0.9177 - val_loss: 0.0870 - val_accuracy: 0.8810
Epoch 5/20
30/30 [==============================] - 1s 26ms/step - loss: 0.0529 - accuracy: 0.9344 - val_loss: 0.0846 - val_accuracy: 0.8860
Epoch 6/20
30/30 [==============================] - 1s 23ms/step - loss: 0.0470 - accuracy: 0.9413 - val_loss: 0.0867 - val_accuracy: 0.8826
Epoch 7/20
30/30 [==============================] - 1s 24ms/step - loss: 0.0350 - accuracy: 0.9601 - val_loss: 0.1116 - val_accuracy: 0.8593
Epoch 8/20
30/30 [==============================] - 1s 25ms/step - loss: 0.0322 - accuracy: 0.9633 - val_loss: 0.0921 - val_accuracy: 0.8805
Epoch 9/20
30/30 [==============================] - 1s 25ms/step - loss: 0.0278 - accuracy: 0.9689 - val_loss: 0.0934 - val_accuracy: 0.8816
Epoch 10/20
30/30 [==============================] - 1s 22ms/step - loss: 0.0236 - accuracy: 0.9737 - val_loss: 0.0970 - val_accuracy: 0.8788
Epoch 11/20
30/30 [==============================] - 1s 23ms/step - loss: 0.0196 - accuracy: 0.9786 - val_loss: 0.0972 - val_accuracy: 0.8815
Epoch 12/20
30/30 [==============================] - 1s 25ms/step - loss: 0.0199 - accuracy: 0.9783 - val_loss: 0.1002 - val_accuracy: 0.8762
Epoch 13/20
30/30 [==============================] - 1s 22ms/step - loss: 0.0183 - accuracy: 0.9797 - val_loss: 0.1038 - val_accuracy: 0.8747
Epoch 14/20
30/30 [==============================] - 1s 23ms/step - loss: 0.0145 - accuracy: 0.9845 - val_loss: 0.1153 - val_accuracy: 0.8634
Epoch 15/20
30/30 [==============================] - 1s 23ms/step - loss: 0.0090 - accuracy: 0.9915 - val_loss: 0.1019 - val_accuracy: 0.8797
Epoch 16/20
30/30 [==============================] - 1s 26ms/step - loss: 0.0166 - accuracy: 0.9815 - val_loss: 0.1059 - val_accuracy: 0.8741
Epoch 17/20
30/30 [==============================] - 1s 29ms/step - loss: 0.0081 - accuracy: 0.9923 - val_loss: 0.1084 - val_accuracy: 0.8732
Epoch 18/20
30/30 [==============================] - 1s 40ms/step - loss: 0.0170 - accuracy: 0.9802 - val_loss: 0.1039 - val_accuracy: 0.8785
Epoch 19/20
30/30 [==============================] - 1s 22ms/step - loss: 0.0075 - accuracy: 0.9926 - val_loss: 0.1046 - val_accuracy: 0.8785
Epoch 20/20
30/30 [==============================] - 1s 21ms/step - loss: 0.0170 - accuracy: 0.9809 - val_loss: 0.1053 - val_accuracy: 0.8772
```

```
history_dict = history.history
history_dict.keys()
```

```
dict_keys(['loss', 'accuracy', 'val_loss', 'val_accuracy'])
```

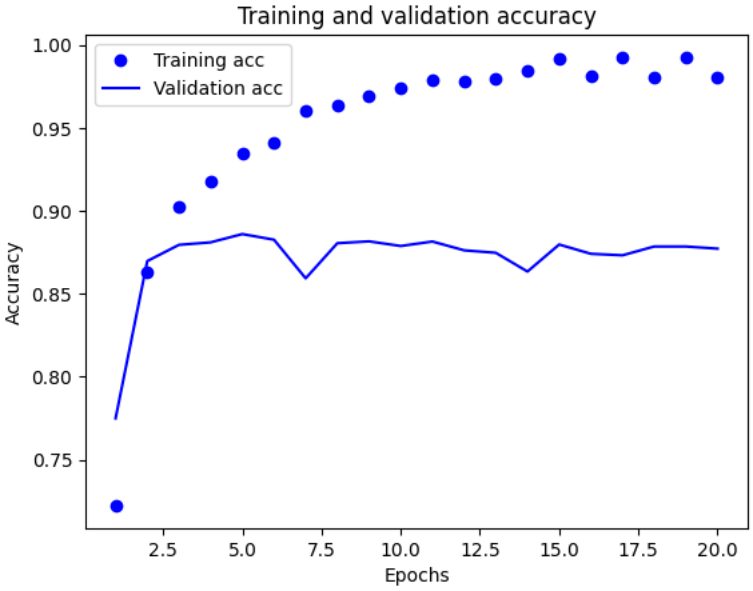
Training and validation loss

```
import matplotlib.pyplot as plt
history_dict = history.history
loss_values = history_dict["loss"]
val_loss_values = history_dict["val_loss"]
epochs = range(1, len(loss_values) + 1)
plt.plot(epochs, loss_values, "bo", label="Training loss")
plt.plot(epochs, val_loss_values, "b", label="Validation loss")
plt.title("Training and validation loss")
plt.xlabel("Epochs")
plt.ylabel("Loss")
plt.legend()
plt.show()
```



Training and validation accuracy

```
plt.clf()
acc = history_dict["accuracy"]
val_acc = history_dict["val_accuracy"]
plt.plot(epochs, acc, "bo", label="Training acc")
plt.plot(epochs, val_acc, "b", label="Validation acc")
plt.title("Training and validation accuracy")
plt.xlabel("Epochs")
plt.ylabel("Accuracy")
plt.legend()
plt.show()
```



Retraining a model from scratch

```
model = keras.Sequential([
layers.Dense(16, activation="relu"),
layers.Dense(16, activation="relu"),
layers.Dense(1, activation="sigmoid")
])
model.compile(optimizer="rmsprop",
loss="mean_squared_error",
metrics=["accuracy"])
model.fit(x_train, y_train, epochs=4, batch_size=512)
results = model.evaluate(x_test, y_test)
```

```
Epoch 1/4
49/49 [==============================] - 2s 14ms/step - loss: 0.1591 - accuracy: 0.7888
Epoch 2/4
49/49 [==============================] - 1s 14ms/step - loss: 0.0861 - accuracy: 0.8980
```

```
Epoch 3/4
49/49 [==============================] - 1s 14ms/step - loss: 0.0663 - accuracy: 0.9209
Epoch 4/4
49/49 [==============================] - 1s 14ms/step - loss: 0.0562 - accuracy: 0.9332
782/782 [==============================] - 2s 3ms/step - loss: 0.0863 - accuracy: 0.8839
```

```
results
```

```
[0.08633721619844437, 0.8839200139045715]
```