HW1

February 7, 2020

0.1 # Assignment1, Deep Learning

- Seyedeh Mina Mousavifar
- 11279515
- sem311

0.2 1. TensorFlow

[0]:
```
try:
  # %tensorflow_version only exists in Colab.
  %tensorflow_version 2.x
except Exception:
  pass
```

0.2.1 Loading Data

[0]:
```
import tensorflow as tf

fashion_mnist = tf.keras.datasets.fashion_mnist
(train_images, train_labels), (test_images, test_labels) = fashion_mnist.
 ↪load_data()

# convert 0-255 pixels to 0-1 pixels for nn input
train_images, test_images = train_images / 255.0, test_images / 255.0
```

0.2.2 Image Classifier with Keras

[0]:
```
###############################
# Image Classifier with Keras - given

import os, datetime

def create_model_keras():
  return tf.keras.models.Sequential([
    tf.keras.layers.Flatten(input_shape=(28, 28)),
    tf.keras.layers.Dense(512, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(10, activation='softmax')
  ])


def train_model_keras():
  model = create_model_keras()
  model.compile(optimizer='adam',
                loss='sparse_categorical_crossentropy',
                metrics=['accuracy'])

  logdir = os.path.join("logs/fit/", datetime.datetime.now().
 ↪strftime("%Y%m%d-%H%M%S"))
  tensorboard_callback = tf.keras.callbacks.TensorBoard(logdir,␣
 ↪histogram_freq=1)

  model.fit(x=train_images,
            y=train_labels,
            epochs=70,
            validation_data=(test_images, test_labels),
            callbacks=[tensorboard_callback])

train_model_keras()
```

0.2.3 Image Classifier without keras

Defining Neural Network Operations

[0]:
```
import tensorflow as tf

DROPOUT_RATE = 0.2
OUTPUT_CLASSES = 10

def dense_relu(inputs , weights):
  """
  This function fully connect inputs to weights with relu activation function
  :param inputs: inputs of neural network
  :param weights: weights of neural network
  :return tensor: tensor network
  """
  return tf.nn.relu(tf.matmul(inputs , weights))


def dense_softmax(inputs , weights):
  """
  This function fully connect inputs to weights with softmax activation␣
 ↪function
  :param inputs: inputs of neural network
```

```
    :param weights: weights of neural network
    :return tensor: tensor network
    """
    return tf.nn.softmax(tf.matmul(inputs , weights))


def drop_out(inputs, dropout_rate):
    """
    This function fully connect inputs to weights with softmax activation␣
 ↪function
    :param inputs: inputs of neural network
    :param dropout_rate: dropout rate of neural network
    :return tensor: tensor network
    """
    return tf.nn.dropout(inputs, rate=dropout_rate)


def flatten(inputs):
    """
    This function transforms the format of the images from
    a three-dimensional array (of batch size by 28 by 28 pixels) to
    a two-dimensional array (of batch size by 28 * 28 = 784 pixels)
    :param inputs: three-dimensional array of 28*28 pixels of number of batches
    :return tensor: reformatted tensor
    """
    # obtaining batch size
    data_batch_size = tf.shape(inputs)[0]
    return tf.reshape(inputs, shape=[data_batch_size,784])
```

Initializing weights

[0]:
```
# initializing weights
initializer = tf.initializers.glorot_uniform()

def get_weight(shape, name):
    """
    This function initializes weights for neural network
    :param shape: shape of weights to initialize
    :param name: name of layer
    :return weights_out: list of weight tensors
    """
    return tf.Variable(initializer(shape), name=name, trainable=True, dtype=tf.
 ↪float32)


# first layer with 512 nodes
# second layer with 10 nodes
```

```
shapes = [[784, 512],
          [512, 10]]

weights = []
# obtatining weights
for i in range(len(shapes)):
  weights.append(get_weight(shapes[i], 'weight{}'.format(i)))
```

Creating Model

[0]:
```
def model(x):
  """
  This function creates three layer neural network
  :param x: image data
  :return x: image classifier neural network
  """
  x = tf.cast(x, dtype=tf.float32)

  # transformation layer, flatting 28*28 matrix to 1*784
  x = flatten(x)

  # first layer, dense layer with relu activation
  x = dense_relu(inputs=x, weights=weights[0])
  # second layer, prunning data
  x = drop_out(inputs=x, dropout_rate=DROPOUT_RATE)
  # third layer, dense layer with softmax activation
  x = dense_softmax(inputs=x, weights=weights[1])

  return x
```

Defining the loss function and optimization

[0]:
```
LEARNING_RATE = 0.001

def loss(pred, target):
  """
  This function calculates loss between predicted value and original value␣
 ↪based on sparse
  categorical cross entropy method
  :param pred: predicted value
  :param target: original value
  :return tensor: reduced tensor
  """
  return tf.losses.sparse_categorical_crossentropy(target, pred)

# adding optimizer
optimizer = tf.optimizers.Adam(LEARNING_RATE)
```

Creating training and test functions

[0]:
```
import datetime

# Define our metrics

# training loss, which is mean of train loss tensor
train_loss = tf.metrics.Mean('train_loss', dtype=tf.float32)
# train accuracy, which calculates accuracy on sparse categorical data
train_accuracy = tf.metrics.SparseCategoricalAccuracy('train_accuracy')
# test loss, which is mean of test loss tensor
test_loss = tf.metrics.Mean('test_loss', dtype=tf.float32)
test_accuracy = tf.metrics.SparseCategoricalAccuracy('test_accuracy')

def train_step(model, optimizer, inputs, outputs):
  """
  This function train the model by calculating gradient and optimizing weights
  :param model: image classification model
  :param inputs: train images
  :param outputs: original labels of train set
  """
  # calculating gradient
  with tf.GradientTape() as tape:
    prediction = model(inputs)
    current_loss = loss(prediction, outputs)
  grads = tape.gradient(current_loss, weights)
  optimizer.apply_gradients(zip(grads, weights))

  # saving loss and accuracy
  train_loss(current_loss)
  train_accuracy(outputs, prediction)


def test_step(model, inputs, outputs):
  """
  This function test the model by its predictions
  :param model: image classification model
  :param inputs: test images
  :param outputs: original labels of test set
  """
  # predict output
  prediction = model(inputs)
  # calculate loss
  current_loss = loss(prediction, outputs)

  # saving loss and accuracy
  test_loss(current_loss)
  test_accuracy(outputs, prediction)
```

```
# logging info for summarizing data and visualizing on tensorboard
current_time = datetime.datetime.now().strftime("%Y%m%d-%H%M%S")
train_log_dir = 'logs/fit/' + current_time + '/my_train'
test_log_dir = 'logs/fit/' + current_time + '/my_test'
train_summary_writer = tf.summary.create_file_writer(train_log_dir)
test_summary_writer = tf.summary.create_file_writer(test_log_dir)
```

Training

[0]:
```
BATCH_SIZE = 64
NUM_EPOCHS = 70

# Using batching capabilities
train_dataset = tf.data.Dataset.from_tensor_slices((train_images, train_labels))
test_dataset = tf.data.Dataset.from_tensor_slices((test_images, test_labels))

# converting to batches
train_dataset = train_dataset.shuffle(60000).batch(BATCH_SIZE)
test_dataset = test_dataset.batch(BATCH_SIZE)


for epoch in range(NUM_EPOCHS):
  # model computation starting time
  start_timer = datetime.datetime.now()

  # training for each batch
  for (image, label) in train_dataset:
    train_step(model, optimizer, image, label)

  # logging train loss and accuracy after one epoch for training set
  with train_summary_writer.as_default():
    tf.summary.scalar('loss', train_loss.result(), step=epoch)
    tf.summary.scalar('accuracy', train_accuracy.result(), step=epoch)

  # predicting for each batch
  for (x_test, y_test) in test_dataset:
    test_step(model, x_test, y_test)

  # logging test loss and accuracy after one epoch for test set
  with test_summary_writer.as_default():
    tf.summary.scalar('loss', test_loss.result(), step=epoch)
    tf.summary.scalar('accuracy', test_accuracy.result(), step=epoch)

  # modeling end time after one epoch and logging
  taken_time = datetime.datetime.now() - start_timer
```

```
# printing summary of data for each epoch
template = 'Epoch {} - time: {}s, {}s per sample - loss: {} - accuracy: {} -␣
↪test loss: {} - test accuracy: {}'
print (template.format(epoch+1,
                      round(taken_time.total_seconds(),2),
                      round(taken_time.total_seconds()*100/6,0),
                      train_loss.result(),
                      train_accuracy.result()*100,
                      test_loss.result(),
                      test_accuracy.result()*100))

# Reset metrics every epoch
train_loss.reset_states()
test_loss.reset_states()
train_accuracy.reset_states()
test_accuracy.reset_states()
```

0.2.4 Tensorboard

[0]:
```
# running tensorboard

%load_ext tensorboard
#%reload_ext tensorboard

%tensorboard --logdir logs
```

[0]:
```
#!rm -rf ./logs/
```

This code is a mixture of tensorflow website, cityscape image segmentation tutorial, and image classification from scratch tutorial.

0.2.5 Comparison

The primary usage of Keras is facilitating modelling, training and validating procedures, in which I needed to implement various functions as mentioned above.

In the tensorboard above, the result of image classification without Keras is described as my_train suffix for the training set(light blue line) and my_test suffix for the test set(red line). The metrics for Keras image classification is described as epoch_accuracy(dark blue line) and epoch_loss(dark blue line).

Accuracy

As we can see in the tensorboard above the first plot, our model gain accuracy of 96% on the training data after 70 epochs, which shows obvious overfitting, as after 31 epochs, the test accuracy won't increase from 88%, but our model continues to improve its train accuracy. However, our final test accuracy is 88.82%.

Keras model gain 95.59% accuracy on the training data and final accuracy of test set after 70 epochs is 89.54%, which shows Keras is better in the modelling procedure.

Loss

In the above tensorboard, the second plots show the loss for Keras classification and the third plot shows the loss for our non-Keras model. We can see that our model has less loss compared to Keras. Our model test loss after 70 Epochs is 49.11%, and Keras' loss on test set is 49.92%. Consequently, our model and Keras doesn't have significant different in final loss on the test set.

Keras' loss is 11.44% on the training data and our model loss is 10.16% after 70 epochs. I think this is caused because of my learning rate, which is 0.001 for the model.

Moreover, we can see that after 14 epochs, both models manage to get to the optimal point during learning, which is shown by the least loss, and after 20 epochs,testing accuracy fluctuates only in 1% range.

Time

In the prints above, we can see that our model epoch time is between 8 to 9 seconds and fluctuates slightly. Keras epoch time is fixed for 7 seconds, which results in 2 minutes difference between Keras and without Keras modelling.

In conclusion, Keras is easier, and faster in modelling neural networks.

Summary

Here is a short comparison between these two models:

Metric	Without Keras	Keras
Accuracy - Test	88.82%	89.54%
Loss - Test	49.11%	49.92%
Time per Epoch	8.7s	7s
Time per Sample	145s	113s
Total Time	9m 58s	7m 54s
Accuracy - Train	96.11%	95.59%
Loss - Train	10.16%	11.44%

0.3 2. CPU vs. GPU computation

[0]:
```
import tensorflow as tf
import numpy as np
import pandas as pd
import time

def create_random_matrix(size_in):
    """
    This function creates two square matrices based on given size of the matrix
    :param size_in: size of the matrix
    :return first_matrix, second_matrix: Returns two random square matrices with␣
  ↪the same size
    """

    # setting seed for having different random numbers for the other matrix
    first_matrix = tf.random.uniform([size_in, size_in])

    tf.random.set_seed(time.time())
    second_matrix = tf.random.uniform([size_in, size_in])
    return first_matrix, second_matrix

def time_product(first_matrix, second_matrix):
    """
    This function multiplicates two square matrices and calculates execution time
    :param first_matrix: first random square matrix
    :param second_matrix: second random square matrix
    :return elapsed: Returns elapsed time for multiplicating matrices
    :return result_product: Returns product of two matrices
    """

    # starting time
    start = time.perf_counter()

    result_product = tf.matmul(first_matrix, second_matrix)

    # calculating elapsed time from execution
    elapsed = time.perf_counter() - start

    return elapsed, result_product

def compute_cpu(first_matrix, second_matrix):
    """
    This function uses CPU to multiplicate two square matrices and calculate␣
 ↪execution time
    :param first_matrix: first random square matrix
    :param second_matrix: second random square matrix
    :return elapsed_time: Returns elapsed time for multiplicating matrices on CPU
    """

    # Force execution on CPU
    with tf.device("CPU:0"):
        elapsed_time, product = time_product(first_matrix, second_matrix)
        # throws exception if the code is not executed on CPU
        assert product.device.endswith("CPU:0")

    return elapsed_time

def compute_gpu(first_matrix, second_matrix):
    """
    This function uses GPU to multiplicate two square matrices and calculate␣
 ↪execution time
    :param first_matrix: first random square matrix
    :param second_matrix: second random square matrix
    :return elapsed_time: Returns elapsed time for multiplicating matrices on GPU
    """
```

```
  # Force execution on GPU #0 if available
  if tf.config.experimental.list_physical_devices("GPU"):
    with tf.device("GPU:0"):
      elapsed_time, product = time_product(matrix1, matrix2)
      # throws exception if the code is not executed on GPU
      assert product.device.endswith("GPU:0")

  return elapsed_time


# main program
rounds = 10
square_size = [500, 1000, 5000, 10000]

answer_timer = {500:{}, 1000:{}, 5000:{}, 10000:{}}

for size in square_size:
  # calculating for each size

  elapsed_time_cpu = 0
  elapsed_time_gpu = 0

  for i in range(rounds):
    # calculating computation for a specific rounds for each size
    matrix1, matrix2 = create_random_matrix(size)

    elapsed_time_cpu += compute_cpu(matrix1, matrix2)

    elapsed_time_gpu += compute_gpu(matrix1, matrix2)

  # calculating average for computation in ms
  answer_timer[size]['CPU'] = elapsed_time_cpu*1000/rounds
  answer_timer[size]['GPU'] = elapsed_time_gpu*1000/rounds


# printing result in tabular format
print("Computing computation time for 10 round of multiplication")
print("{:<8} {:<15} {:<10}".format('size','processor', 'execution time(ms)'))
for size, item in answer_timer.items():
    for proc, timer in item.items():
      print("{:<8} {:<15} {:<10}".format(size, proc, timer))
```

```
Computing computation time for 10 round of multiplication
size     processor       execution time(ms)
```

```
500       CPU               8.90463819996512
500       GPU               0.1385243000186165
1000      CPU               31.14645260029647
1000      GPU               0.2542460002587177
5000      CPU               3473.367635300201
5000      GPU               0.3389669003809104
10000     CPU               28565.83050890004
10000     GPU               0.3881196995280334
```

Sample Output for multiplication computation for average in 10 round of multiplication in m second:

Size	Processor	Execution time(ms)
500	CPU	9.760756099922219
500	GPU	0.15373470009762968
1000	CPU	32.54002580001725
1000	GPU	0.24147199987964996
5000	CPU	3449.983324600135
5000	GPU	0.3366110000115441
10000	CPU	28675.065360599954
10000	GPU	0.38258300000961754

Because the matrices were sampled from uniform distribution, for each size the computation time was averaged between specific number of rounds.

As we can see in this table, the computation time for GPU is significantly less than CPU. But one important result is that as size of matrices increase dramatically, the computation time for CPU increases significantly but for GPU there is a slight rise in computation time.

This code was obtained from tensorflow website.

0.4 3. Differentiation

Jaconbian Matrix **J**:

$$\mathbf{f} : R^m \to R^n \,, \mathbf{J} \in R^{n*m} \ : J_{i,j} = \frac{\partial}{\partial x_j} f(\mathbf{x})_i$$

$$\mathbf{y} = 3\mathbf{x}^2 + 2\mathbf{x} + 3 = 3(\mathbf{x} \odot \mathbf{x}) + 2\mathbf{x} + 3 \Rightarrow \frac{\partial y}{\partial x} = 3(\frac{\partial x}{\partial x} \odot x + x \odot \frac{\partial x}{\partial x}) + 2\frac{\partial x}{\partial x} + 3\frac{\partial}{\partial x}$$

$$\mathbf{J} = \begin{pmatrix} 3(x+x) + 2 + 0 \end{pmatrix} = \begin{pmatrix} \underline{6x+2} \end{pmatrix}$$

for $x = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \to \mathbf{J} = \begin{pmatrix} 8 \\ 20 \end{pmatrix}$

This following code was obtained from tensorflow website.

[0]:
```
import tensorflow as tf
import numpy as np

# creating x matrix
```

```
x = tf.constant([[1.0], [3.0]])

with tf.GradientTape(persistent=True) as t:
  t.watch(x)
  y = 3* x * x + 2 * x + 3

dy_dx = t.gradient(y, x)

print("Jacobian of y = 3x^2 + 2x + 3 for x = [[1][3]]:")
with tf.Session() as sess:  print("J = ", dy_dx.eval())
```

```
Jacobian of y = 3x^2 + 2x + 3 for x = [[1][3]]:
J =  [[ 8.]
 [20.]]
```

As we can see above, by hand and by code the same result is obtained.

0.5 4. Eigenvectors

$$Av = \lambda v \quad where\ \lambda : eigenvalue\ ,\ v : eigenvector$$

$$Av - \lambda v = 0 \rightarrow (A - \lambda I)v = 0\ , v \neq 0 \rightarrow A - \lambda I = 0$$

Calculating eigenvalues)

$$A - \lambda I = 0\ , ker(A - \lambda I) \neq 0 \rightarrow A - \lambda I\ \ : not - invertible \Rightarrow \underline{\underline{det(A - \lambda I) = 0}}$$

$$det(\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} - \lambda I) = det(\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) = det(\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}) = det(\begin{pmatrix} 3-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix}) \rightarrow$$

$$det(A - \lambda I) = \lambda^2 - 6\lambda + 5 = 0 \rightarrow (\lambda - 5)(\lambda - 1) = 0 \rightarrow \underline{\lambda = 5}\ , \underline{\lambda = 1}$$

Calculating eigenvectors)

$$A - \lambda I = \begin{pmatrix} 3-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix}\ , (A - \lambda I)v = 0$$

$\underline{\lambda = 5}$)

$$\begin{pmatrix} 3-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

By adding first row to the second row, we have:

$$\begin{pmatrix} -2 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow -2v_1 + 2v_2 = 0 \rightarrow \underline{v_1 = v_2}$$

$\underline{\lambda = 1}$)

$$\begin{pmatrix} 3-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

By subtracting first row from the second row, we have:

$$\begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow 2v_1 + 2v_2 = 0 \rightarrow \underline{v_1 = -v_2}$$

Examples)

We can check correctness of our eigen values and eigenvectors by checking the following equation:

$$Av = \lambda v$$

1. $\underline{\lambda = 5, v_1 = v_2}$) $v_1 = 1,\ v_2 = 1$

$$\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 5 \\ 5 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix} \quad \checkmark$$

2. $\underline{\lambda = 1, v_1 = -v_2}$) $v_1 = 1,\ v_2 = -1$

$$\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \checkmark$$
