Deep Learning 실습

원중호

서울대학교 통계학과

2019년 6월 21일

목차

- Tensorflow and Keras
- Using Keras on R
- Example MNIST Data and CNN
- Example Generate Nietzsche's writing with LSTM

Section 1

Tensorflow and Keras

3 / 58

Tensorflow 소개

- an open source software library for numerical computation using data flow graphs, more than deep learning. TensorFlow actually has tools to support reinforcement learning and other algos.
- Google Brain Team within Google's Machine Intelligence research organization
- 언어 : a Python API over a C/C++ engine
 - Deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API
- Deep Learning Course at Udacity

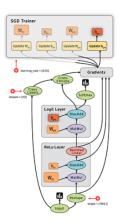
Tensorflow 특징

- Deep Flexibility
 - ► TensorFlow isn't a rigid neural networks library
 - If you can express your computation as a data flow graph, you can use TensorFlow
 - ▶ You construct the graph, and you write the inner loop that drives computation
- True Portability
 - TensorFlow runs on CPUs or GPUs, and on desktop, server, or mobile computing platforms
 - Scale-up and train that model faster on GPUs with no code changes
 - Deploy that trained model on mobile
 - Run the model as a service in the cloud

Tensorflow 특징

- Connect Research and Production
 - Google research scientists experiment with new algorithms in TensorFlow
 - Google Product teams use TensorFlow to train and serve models live to real customers
- Auto-Differentiation
 - Can define the computational architecture of predictive model, combine that with objective function, and just add data
 - TensorFlow handles computing the derivatives
 - Computing the derivative of some values w.r.t. other values in the model just extends the graph

Computational Graph



 $https://www.tensorflow.org/images/tensors_flowing.gif$

7 / 58

Computational Graph

- TensforFlow generates a computational graph (e.g. a series of matrix operations such as z = simoid(x) where x and z are matrices) and performs automatic differentiation.
- Automatic differentiation is important because you don't want to have to hand-code a new variation of backpropagation every time you're experimenting with a new arrangement of neural networks.

Mathematical computation with a directed graph

Nodes

- represent mathematical operations
- can also represent endpoints to feed in data, push out results, or read/write persistent variables
- assigned to computational devices
 - execute asynchronously and in parallel once all the tensors on their incoming edges becomes available

Edges

- represent the input/output relationships between nodes
- the multidimensional data arrays (tensors) communicated between them

Tensorflow Reference

- TensorFlow Tutorials
- Additional Resources
- Zoo : TensorFlow enables researchers to build machine learning models. We collect such models in our Zoo.

Keras 소개

- a minimalist, highly modular neural networks library, written in Python and capable of running on top of either TensorFlow or Theano.
- https://github.com/fchollet/keras
- 언어 : a Python API on top of TensorFlow or Theano
- The MIT License (MIT)
- Examples

Keras 특징

특징

- Modularity. A model is understood as a sequence or a graph of standalone, fully-configurable modules that can be plugged together with as little restrictions as possible. In particular, neural layers, cost functions, optimizers, initialization schemes, activation functions, regularization schemes are all standalone modules that you can combine to create new models.
- Minimalism. Each module should be kept short and simple. Every piece of code should be transparent upon first reading. No black magic: it hurts iteration speed and ability to innovate.
- Easy extensibility. New modules are dead simple to add (as new classes and functions), and existing modules provide ample examples. To be able to easily create new modules allows for total expressiveness, making Keras suitable for advanced research.
- Work with Python. No separate models configuration files in a declarative format. Models are described in Python code, which is compact, easier to debug, and allows for ease of extensibility.

Keras backends

http://keras.io/backend/

TensorFlow

• an open-source symbolic tensor manipulation framework developed by Google, Inc.

Theano

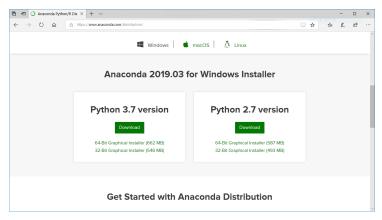
 an open-source symbolic tensor manipulation framework developed by LISA Lab at Université de Montréal.

CNTK

• an open-source toolkit for deep learning developed by Microsoft.

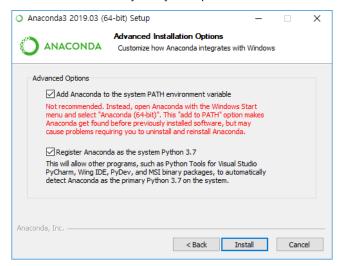
Using Keras on R

 Install Anaconda for Windows (Python 3.7) at https://www.anaconda.com/distribution/



Using Keras on R

• Be sure to add anaconda on your system path.



Using Keras on R

- keras_setting.R
- Setup Keras with Tensorflow backend on Rstudio

The following R code will setup conda environment for keras automatically.

```
install.packages("keras")
library(keras)
keras_install()
```

Section 2

Keras Example - MNIST Data and CNN

2019년 6월 21일

17 / 58

- MNIST_dense.R
- Load library and set hyper-parameter

```
library(keras)
```

```
batch_size <- 128
num_classes <- 10
epochs <- 30
```

Get MNIST data

```
mnist <- dataset mnist()</pre>
x_train <- mnist$train$x
y train <- mnist$train$y</pre>
x test <- mnist$test$x
y test <- mnist$test$y</pre>
dim(x train)
## [1] 60000
                  28
                         28
dim(y train)
## [1] 60000
y train[1]
## [1] 5
```



Prepare dataset

```
x_train <- array_reshape(x_train, c(nrow(x_train), 784))</pre>
x test <- array reshape(x test, c(nrow(x test), 784))
dim(x train)
## [1] 60000 784
x train <- x train / 255
x test <- x test / 255
y train <- to categorical(y train, 10)
y_test <- to_categorical(y_test, 10)</pre>
dim(y_train)
## [1] 60000 10
y_train[1,]
    [1] 0 0 0 0 0 1 0 0 0 0
##
```

Architecture Design

Architecture Design

> summary(model)

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	200960
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 128)	32896
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 10)	1290

Total params: 235,146 Trainable params: 235,146 Non-trainable params: 0

 Architecture design example : click Tinker With a Neural Network in Your Browser

Architecture Components

- 2D Convolution, 2D Maxpooling, Dense(MLP) layers...
- ReLU, Softmax activations
- Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift

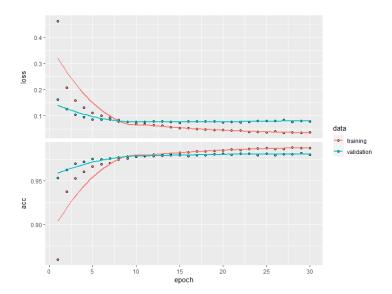
Model compile

```
model %>% compile(
  loss = 'categorical_crossentropy',
  optimizer = optimizer_adam(),
  metrics = c('accuracy')
)
```

Model training

```
history <- model %>% fit(
  x_train, y_train,
  epochs = epochs, batch_size = batch_size,
  validation_split = 0.2
)
```

```
Train on 48000 samples, validate on 12000 samples
Epoch 1/30
Fnoch 2/30
48000/48000 [---
  Epoch 3/30
Epoch 4/30
Epoch 5/30
Epoch 6/30
Epoch 8/30
```



Evaluation and Prediction

```
model %>% evaluate(x_test, y_test)
model %>% predict_classes(x_test) %>% head()
```

MNIST_cnn.R

```
# CNN Data Preparation
# Input image dimensions
img_rows <- 28
img_cols <- 28

# The data, shuffled and split between train and test sets
mnist <- dataset_mnist()
x_train <- mnist$train$x
y_train <- mnist$train$y
x_test <- mnist$test$x
y_test <- mnist$test$y</pre>
```

Data Preparation

```
# Input image dimensions
img_rows <- 28
img_cols <- 28

# The data, shuffled and split between train and test sets
mnist <- dataset_mnist()
x_train <- mnist$train$x
y_train <- mnist$train$y
x_test <- mnist$test$x
y_test <- mnist$test$y</pre>
```

```
# Redefine dimension of train/test inputs
x train <- array reshape(x train, c(nrow(x train),
                                       img_rows, img_cols, 1))
x_test <- array_reshape(x_test, c(nrow(x_test),</pre>
                                    img_rows, img_cols, 1))
input_shape <- c(img_rows, img_cols, 1)</pre>
# Transform RGB values into [0,1] range
x_train <- x_train / 255
x_test <- x_test / 255
y_train <- to_categorical(y_train, 10)</pre>
y_test <- to_categorical(y_test, 10)</pre>
```

```
model <- keras_model_sequential()</pre>
model %>%
  layer conv 2d(filters = 32, kernel size = c(3,3),
                activation = 'relu', input shape = input shape) %>%
  layer conv 2d(filters = 64, kernel size = c(3,3),
                activation = 'relu') %>%
  layer max pooling 2d(pool size = c(2, 2)) %>%
  layer dropout(rate = 0.25) %>%
  layer flatten() %>%
  layer dense(units = 128, activation = 'relu') %>%
  layer dropout(rate = 0.5) %>%
  layer_dense(units = num_classes, activation = 'softmax')
```

> summary(model)

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 26, 26, 32)	320
conv2d_1 (Conv2D)	(None, 24, 24, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 12, 12, 64)	0
dropout (Dropout)	(None, 12, 12, 64)	0
flatten (Flatten)	(None, 9216)	0
dense (Dense)	(None, 128)	1179776
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 10)	1290

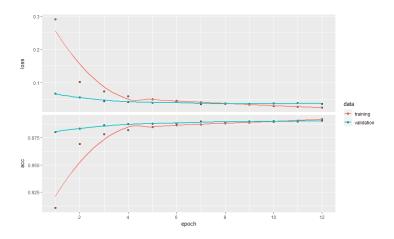
Total params: 1,199,882 Trainable params: 1,199,882 Non-trainable params: 0

```
# Model config
model %>% compile(
  loss = 'categorical crossentropy',
  optimizer = optimizer adam(),
 metrics = c('accuracy')
summary(model)
# Start Fitting
system.time({
 history <- model %>% fit(
    x train, y train,
    batch_size = batch_size,
    epochs = epochs,
    validation_split = 0.2
```

```
Train on 48000 samples, validate on 12000 samples
Epoch 1/12
- val_acc: 0.9805
Epoch 2/12

    val acc: 0.9835

Fnoch 3/12
48000/48000 [============= ] - 83s 2ms/sample - loss: 0.0734 - acc: 0.9781 - val_loss: 0.0446
- val_acc: 0.9867
Epoch 4/12
- val_acc: 0.9874
Epoch 5/12
- val_acc: 0.9880
Fnoch 6/12
- val acc: 0.9878
Epoch 7/12
- val acc: 0.9899
```



Model evaluation

```
> scores <- model %>% evaluate(
+ x_test, y_test, verbose = 0
+ )
> # Output metrics
> cat('Test loss:', scores[[1]], '\n')
Test loss: 0.03092995
> cat('Test accuracy:', scores[[2]], '\n')
Test accuracy: 0.9923
> |
```

Save whole model

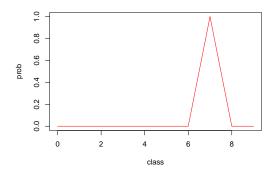
```
model %>% save_model_hdf5("MNIST_cnn.h5")
new_model <- load_model_hdf5("MNIST_cnn.h5")
new_model %>% summary()
```

• Save weights in model

```
model %>% save_model_weights_hdf5("MNIST_cnn_weights.h5")
model %>% load_model_weights_hdf5("MNIST_cnn_weights.h5")
model %>% predict_classes(x_test) %>% head()
```



Predicted probability for each class



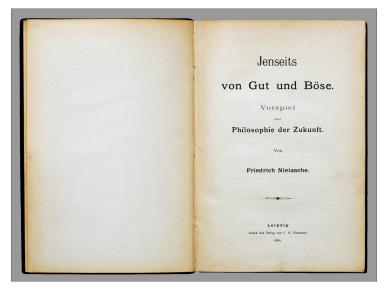
Keras Example - Image recognition with advanced CNN

- ResNet.R
- Download Pre-trained Model

```
model <- application resnet50(weights = 'imagenet')</pre>
img_path <- "elephant.jpg"</pre>
img <- image_load(img_path, target_size = c(224,224))</pre>
x <- image_to_array(img)</pre>
x <- array_reshape(x, c(1, dim(x)))</pre>
x <- imagenet_preprocess_input(x)</pre>
preds <- model %>% predict(x)
imagenet_decode_predictions(preds, top = 3)[[1]]
##
     class name class description
                                            score
```

Section 3

• Generate_Nietzsche.R



Data preparation

```
library(keras)
library(readr)
library(stringr)
library(purrr)
library(tokenizers)
# Parameters
maxlen <-40
# Data Preparation
# Retrieve text
path <- get file(
  'nietzsche.txt',
  origin='https://s3.amazonaws.com/text-datasets/nietzsche.txt'
```

```
> read_lines(path) %>% head()
        [1] "PREFACE"
        [4] "SUPPOSING that Truth is a woman--what then? Is there not ground"
        [5] "for suspecting that all philosophers, in so far as they have been"
        [6] "dogmatists, have failed to understand women--that the terrible"
text <- read_lines(path) %>%
  str to lower() %>%
  str_c(collapse = "\n") %>%
  tokenize_characters(strip_non_alphanum = FALSE, simplify = TRUE)
print(sprintf("corpus length: %d", length(text)))
## [1] "corpus length: 600893"
```

```
chars <- text %>%
  unique() %>%
  sort()

print(sprintf("total chars: %d", length(chars)))

## [1] "total chars: 57"
  |> chars
```

Cutting texts in semi-redundant sequences of maxlen characters

```
> dataset$sentece[[1]]
[1] "p" "r" "e" "f" "a" "c" "e" "\n" "\n" "\n" "\n" "s" "u" "p" "p" "o"
[16] "s" "1" "n" "g" "" "t" "h" "a" "t" "" "t" "r" "u" "t" "h"
[31] "" "i" "s" "" "a" "a" "" "w" "o" "m" "a"

> dataset$sentece[[2]]
[1] "f" "a" "c" "e" "\n" "\n" "\n" "s" "u" "p" "p" "o" "s" "i" "n"
[16] "g" " "t" "h" "a" "t" "" "t" "r" "u" "t" "h" "h" "s"

> dataset$sentect[[1]]
[1] "n"

> dataset$next_char[[2]]
[1] "m"
```

• Vectorization (Warning: array X needs lots of memory)

```
X <- array(0, dim = c(length(dataset$sentece), maxlen, length(chars)</pre>
y <- array(0, dim = c(length(dataset$sentece), length(chars)))</pre>
for(i in 1:length(dataset$sentece)){
  X[i, , ] <- sapply(chars, function(x){</pre>
    as.integer(x == dataset$sentece[[i]])
  })
  v[i, ] <- as.integer(chars == dataset$next char[[i]])</pre>
```

Model Definition

```
model <- keras model sequential()</pre>
model %>%
  layer_lstm(units = 128, input_shape = c(maxlen, length(chars))) %
  layer_dense(length(chars)) %>%
  layer_activation("softmax")
optimizer <- optimizer_adam(lr = 0.01)
model %>% compile(
  loss = "categorical_crossentropy",
  optimizer = optimizer
summary(model)
```

> summary(model)

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 128)	95232
dense (Dense)	(None, 57)	7353
activation (Activation)	(None, 57)	0

Total params: 102,585 Trainable params: 102,585 Non-trainable params: 0

Training and Evaluation

```
# large temperature means more uniform from character set
sample_mod <- function(preds, temperature = 1) {
  preds <- log(preds) / temperature
  exp_preds <- exp(preds)
  preds <- exp_preds / sum(exp(preds))

rmultinom(1, 1, preds) %>%
  as.integer() %>%
  which.max()
}
```

```
set.seed(123)
diversity <- 0.5
for(iteration in 1:20){
    cat(sprintf("iteration: %02d -----\n\n", iteration))
    model %>% fit(
        Х, у,
        batch size = 128,
        epochs = 1
    start index <- sample(1:(length(text) - maxlen), size = 1)</pre>
    sentence <- text[start_index:(start_index + maxlen - 1)]</pre>
    generated <- ""
```

```
# Generate 400 chars with randomly chosen sequence from train d
for(i in 1:400){
  x <- sapply(chars, function(x){
    as.integer(x == sentence)
  })
  x <- array_reshape(x, c(1, dim(x)))
  preds <- predict(model, x)</pre>
  next_index <- sample_mod(preds, diversity)</pre>
  next char <- chars[next index]</pre>
  generated <- str_c(generated, next_char, collapse = "")</pre>
  sentence <- c(sentence[-1], next_char)</pre>
}
cat(generated)
cat("\n\n")
```

iteration: 01	
200284/200284 [rich a
iteration: 02	
200284/200284 [====================================	ained
iteration: 03	
200284/200284 [

thing the sense of the life in a phil

iteration: 18
200245/200244 [
iteration: 19
20028/200284 [
iteration: 20
200284/200284 [

Reference

- https://keras.rstudio.com/ (R interface to Keras)
- He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- https://tensorflow.rstudio.com/keras/articles/examples/lstm_text_generation.html
- https: //de.wikipedia.org/wiki/Jenseits_von_Gut_und_B%C3%B6se_(Nietzsche)