What we can do with Theano

Liangliang Cao

llcao.net/cu-deeplearning17

Outline

- *Optimization of any function with Theano*
- *Brainstorming of the project ideas*

Magic function in Theano

Symbolic computation of gradient

theano.tensor.grad( )

- Example:

```
import theano
from theano import tensor as T

x = T.dscalar('x')
y = x ** 2
gy = T.grad(y, x)
```

With auto-gradient you can easily do optimization

```
scipy.optimize.fmin_l_bfgs_b(
                    your_loss_func,
                    initial_para,
                    your_grad_func,
                    max_iterations)
```

- We'll discuss in group on how to use this

An example

```
inputs_x = np.array([[2, 1], [0.3, 0.5]]).astype(np.float32)
inputs_y = np.array([[1, 0]]).astype(np.float32)

shared_w = theano.shared(np.array(np.array([3.0, -3.0]).astype(np.float64), dtype=np.float32))
estis_y = T.dot(shared_w, inputs_x)

# mse loss
loss = T.mean(T.sum((estis_y-inputs_y)**2, axis=1))

# gradient
grad = T.grad(loss, shared_w)

# theano function
f_loss = theano.function([], loss)
f_grad = theano.function([], grad)

w_init = np.array([[0.0, 0.0]]).astype(np.float64)
opt_result = scipy.optimize.fmin_l_bfgs_b(f_loss, w_init, fprime=f_grad, maxfun=40)
print 'after opt:', opt_result[1]
```

What is wrong with the above code?

Corrected example

```
inputs_x = np.array([[2, 1], [0.3, 0.5]]).astype(np.float32)
inputs_y = np.array([[1, 0]]).astype(np.float32)

shared_w = theano.shared(np.array(np.array([3.0, -3.0]).astype(np.float64), dtype=np.float32))
estis_y = T.dot(shared_w, inputs_x)

# mse loss
loss = T.mean(T.sum((estis_y-inputs_y)**2, axis=1))

# gradient
grad = T.grad(loss, shared_w)

# theano function
f_loss = theano.function([], loss)
f_grad = theano.function([], grad)

# non-theano functions that can be called by optimizer
def eval_loss(x0):
    shared_w.set_value(x0.astype(np.float32))
    return f_loss().astype('float64')

def eval_grad(x0):
    shared_w.set_value(x0.astype(np.float32))
    return np.array(f_grad()).flatten().astype('float64')

w_opt = np.array([[0, 0.0]]).astype(np.float64)
opt_result = scipy.optimize.fmin_l_bfgs_b(eval_loss, w_opt, fprime=eval_grad, maxfun=40)
print 'after opt:', opt_result[1]
```

Brainstorming

What can you do with theano's optimization tool?

- If you have a classically-defined optimization problem
  - Theano can reduce your efforts

- But it will even cooler if you can change a classical problem into a large scale learning problem!
  - People may not realize this is a learning problem
  - You can collect/simulate a lot of data
  - Theano/Tensorflow/Keras can help you try different models to solve them.

Examples of cool application

- AlphaGo, PokerCNN
- Video2GIF
- Sarcasm prediction based on your tweet/instagram

Alpha Go

- Traditionally the go is thought to be a search problem.
  - Just as IBM DeepBlue treats the Chess problem.

- But DeepMind managed to train a classifier on it
  - Learning from the massive playing record
  - Playing against itself
  - Reinforce learning

Paper reading in week 12

Poker CNN

- Get a lot of training data from simulation
- Tensor presentation for poker game
- AAAI'16: the first paper to solve poker betting with deep CNN

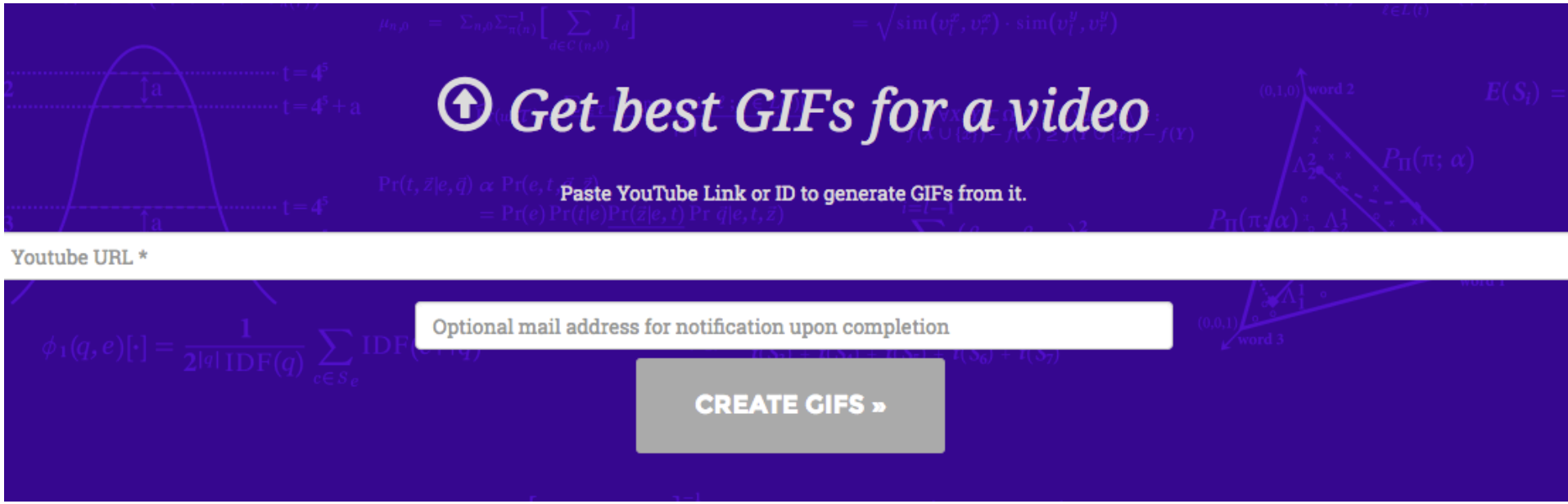
Worth $1.80 on average

Guest speaker in week 12

Video2GIF

- Collect data from giftsoup.com etc

- Learn a ranking function to select good video clips



Sarcasm detection

- Collect data from social media sites

- Sarcasm happens when an image is combined with texts!

With Rossano, Paloma, Joel, ACM Multimedia'16

Will discuss in week 13's lecture

In a rubbish city with rubbish weather
#Liverpool #nofilter

Now discuss in groups

What should be your project idea?

- problem
- where to get data
- any help from the instructors?

Volunteer to present your idea in 10 mins!

(Those promising ideas will receive 5 extra credits)

Summary of this class

- This class talked about math of deep network
- We also discussed theano + scipy.optimize
  - But with a lot of data often we shall use SGD optimizer
  - Keras provides a nice interface for SGD (and a good wrapper of layers)
  - Chad will (hopefully) give a tutorial on Keras next week.

Todo list before next class

- Drop the class if you have difficulty with HW1
- Otherwise
  - Join the google newsgroup
  - Discuss with your potential teams on how to use Keras for your project