TensorFlow Tutorial

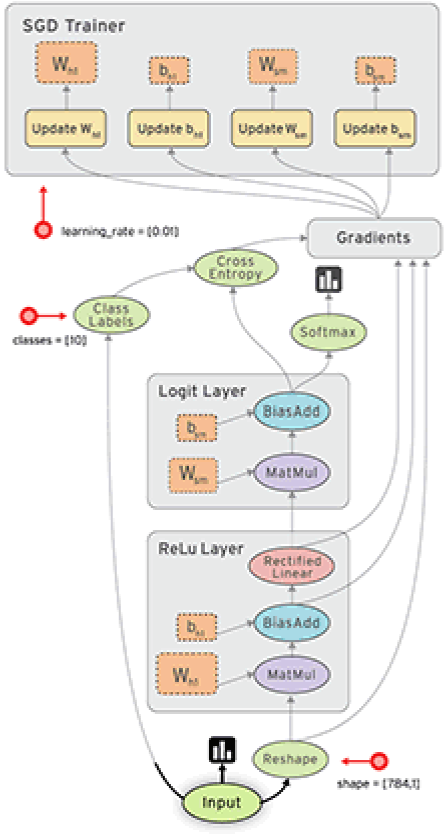
李理

提纲

- TensorFlow是什么
- 自动求梯度
- 自定义op
- 基本概念
- install和build
- 例子
- 参考资料

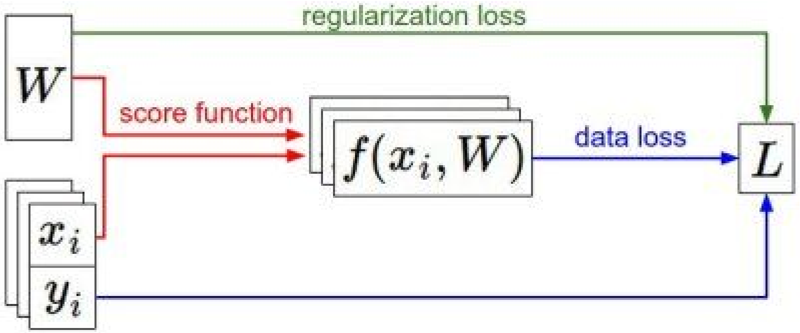
TensorFlow是什么

- TensorFlow就是 Tensor Flow
  - Scalar -> vector -> matrix -> tensor
- Flow in Graph



TensorFlow是什么

- Deep Learning -> Tensor Flowing
- Predict
  - y’ = f(w;x)
- Train
  - Loss = L(y, y’)
  - W* = argmaxw(Loss)



Optimization

- Stochastic Gradient Descent
  - W += -learning_rate * dw
- 关键问题
  - 求梯度

计算梯度/导数的4种方法

- 手工计算
  - d(x+y) = dx + dy
- 数值求导(Numeric Differentiation)
- 符号求导(Symbolic Differentiation)
- 自动求导(Automatic Differentiation)

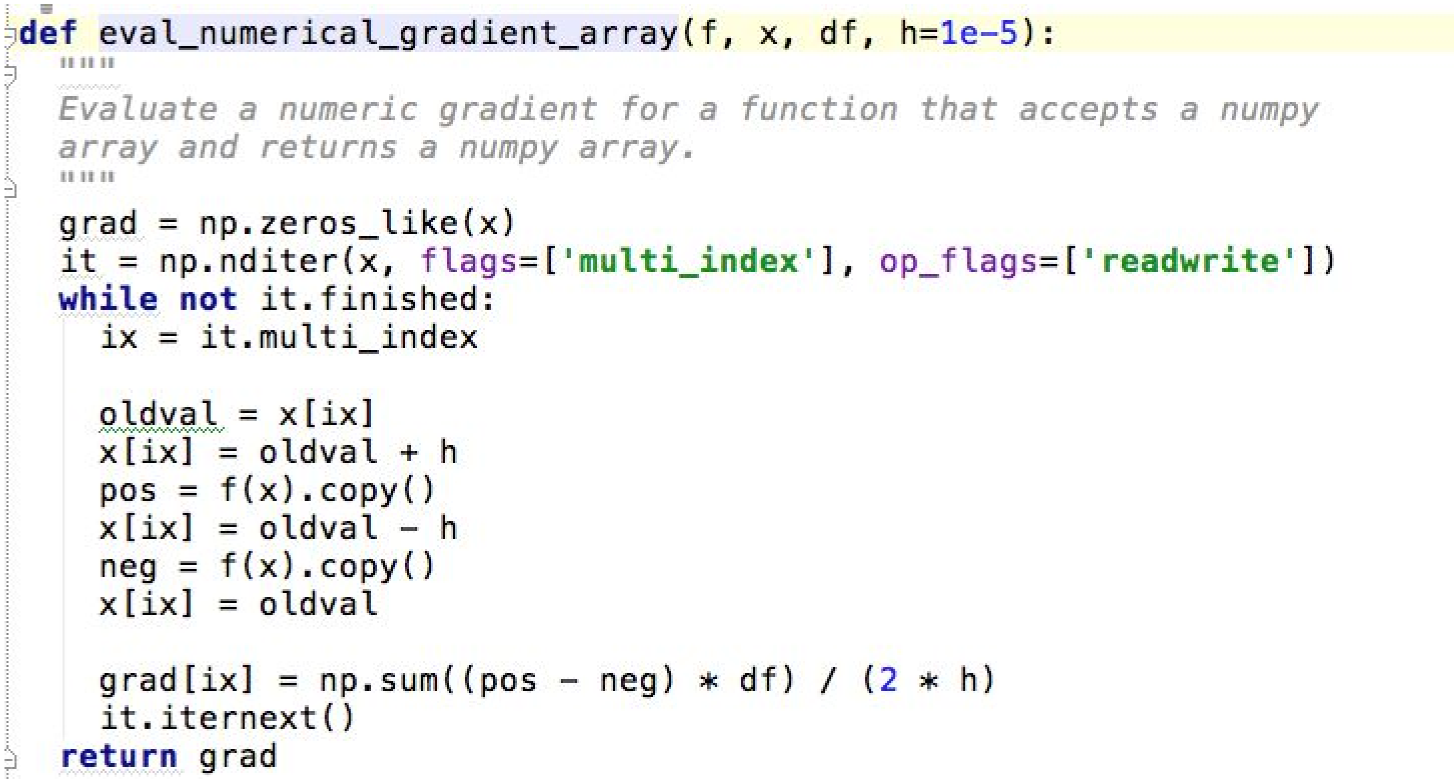
数值求导

- 优点
  - 容易实现
- 缺点
  - 计算量大
  - 误差
- 可以用来gradient check

$$\frac{\partial f(\mathbf{x})}{\partial x_i} \approx \frac{f(\mathbf{x}+h\mathbf{e}_i)-f(\mathbf{x})}{h} \ ,$$

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \frac{f(\mathbf{x}+h\mathbf{e}_i)-f(\mathbf{x}-h\mathbf{e}_i)}{2h} + O(h^2) \ ,$$

数值求导



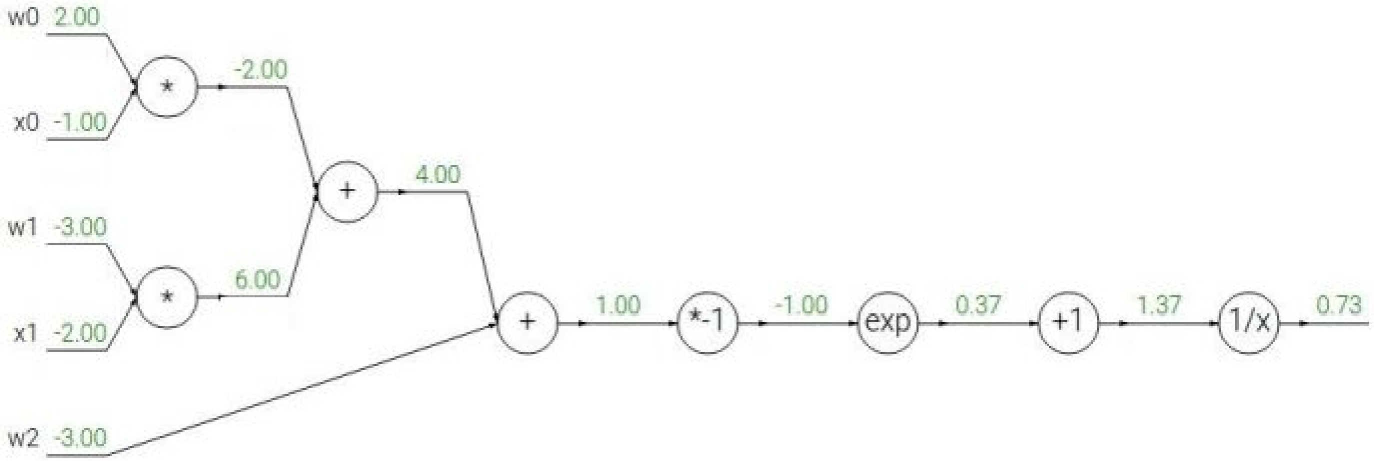
```
def eval_numerical_gradient_array(f, x, df, h=1e-5):
    """
    Evaluate a numeric gradient for a function that accepts a numpy
    array and returns a numpy array.
    """
    grad = np.zeros_like(x)
    it = np.nditer(x, flags=['multi_index'], op_flags=['readwrite'])
    while not it.finished:
        ix = it.multi_index

        oldval = x[ix]
        x[ix] = oldval + h
        pos = f(x).copy()
        x[ix] = oldval - h
        neg = f(x).copy()
        x[ix] = oldval

        grad[ix] = np.sum((pos - neg) * df) / (2 * h)
        it.iternext()
    return grad
```

符号求导

- 算法“自动”推导出Analytic Differentiation
    - matlab maple mathematica等
- 优点
    - 精确的导数，快（取决于符号运算库）
- 缺点
    - 不是所有函数都有解析的导数

自动求导

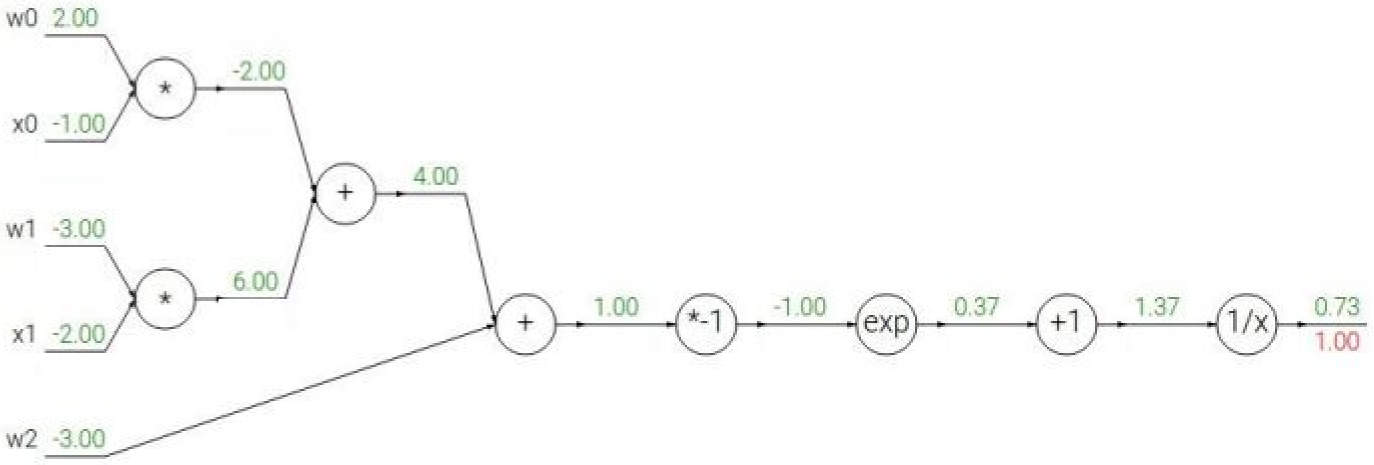
Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



自动求导

Another example:

$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$

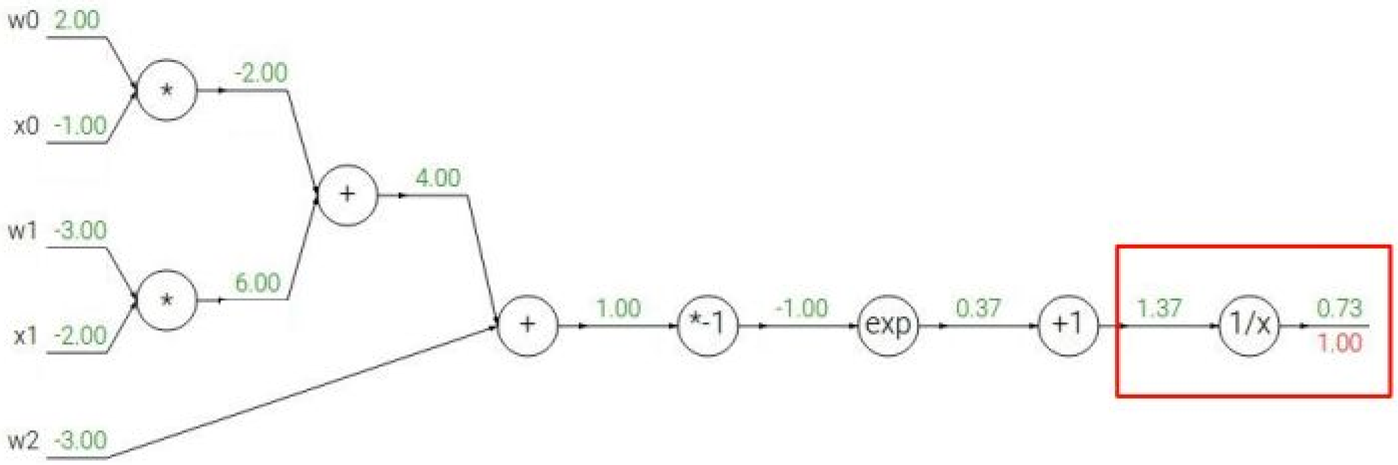
$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$

$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$

自动求导

Another example: $$f(w,x)=\frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



$$f(x)=e^x \quad \rightarrow \quad \frac{df}{dx}=e^x$$

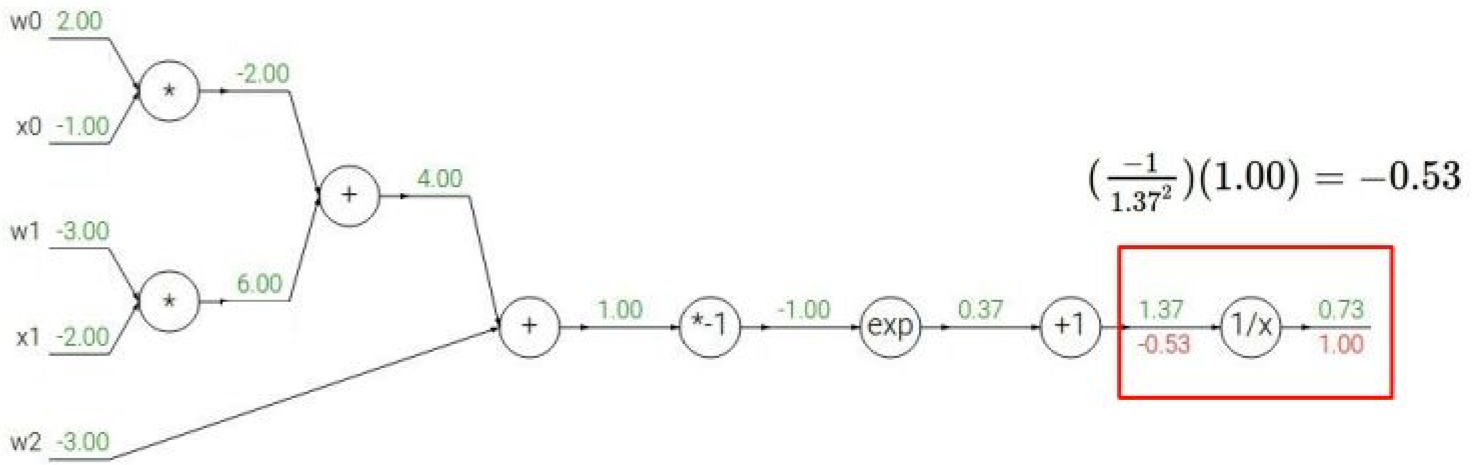
$$f_a(x)=ax \quad \rightarrow \quad \frac{df}{dx}=a$$

$$\boxed{f(x)=\frac{1}{x} \quad \rightarrow \quad \frac{df}{dx}=-1/x^2}$$

$$f_c(x)=c+x \quad \rightarrow \quad \frac{df}{dx}=1$$

自动求导

Another example: $f(w,x) = \dfrac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



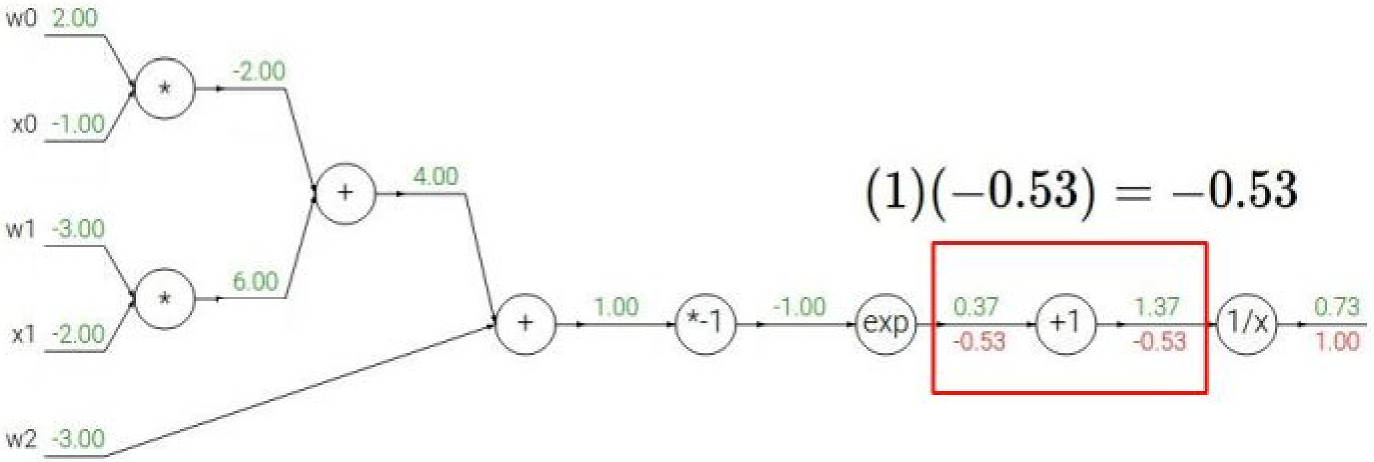
$(\frac{-1}{1.37^2})(1.00) = -0.53$

$f(x) = e^x$	$\rightarrow$	$\dfrac{df}{dx} = e^x$	$f(x) = \dfrac{1}{x}$	$\rightarrow$	$\dfrac{df}{dx} = -1/x^2$
$f_a(x) = ax$	$\rightarrow$	$\dfrac{df}{dx} = a$	$f_c(x) = c + x$	$\rightarrow$	$\dfrac{df}{dx} = 1$

自动求导

Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$

$$(1)(-0.53) = -0.53$$



$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

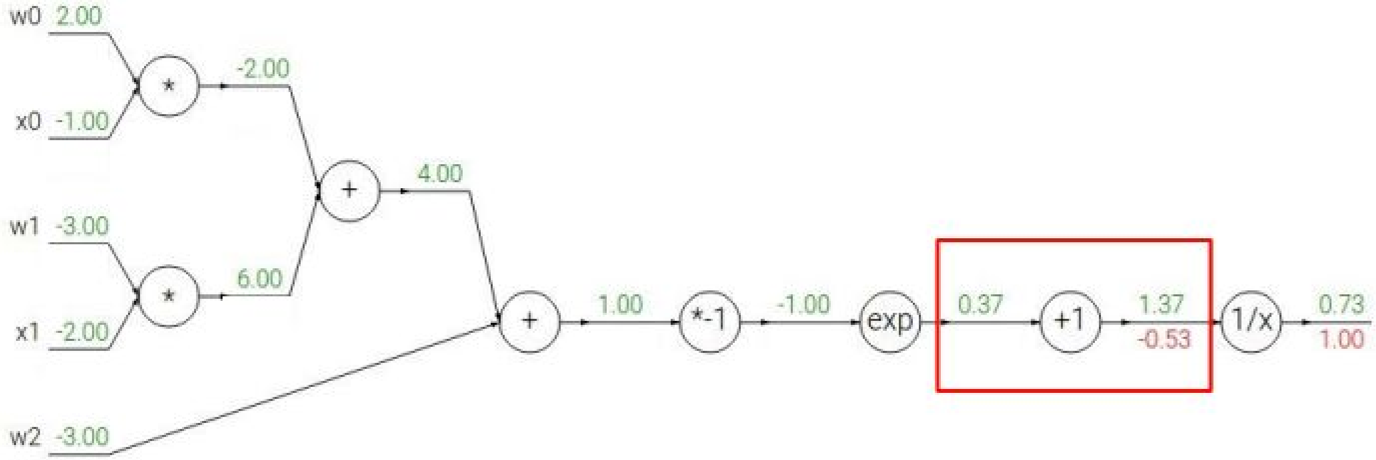
$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$\boxed{f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1}$$

自动求导

Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$

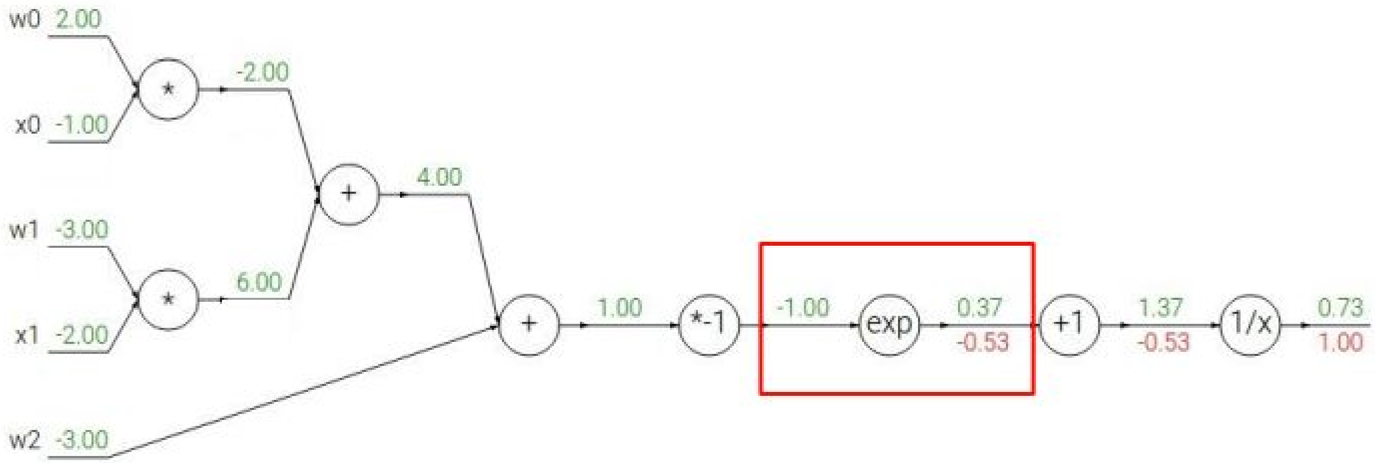
$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$

$\boxed{f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1}$

自动求导

Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



$$\boxed{f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x}$$

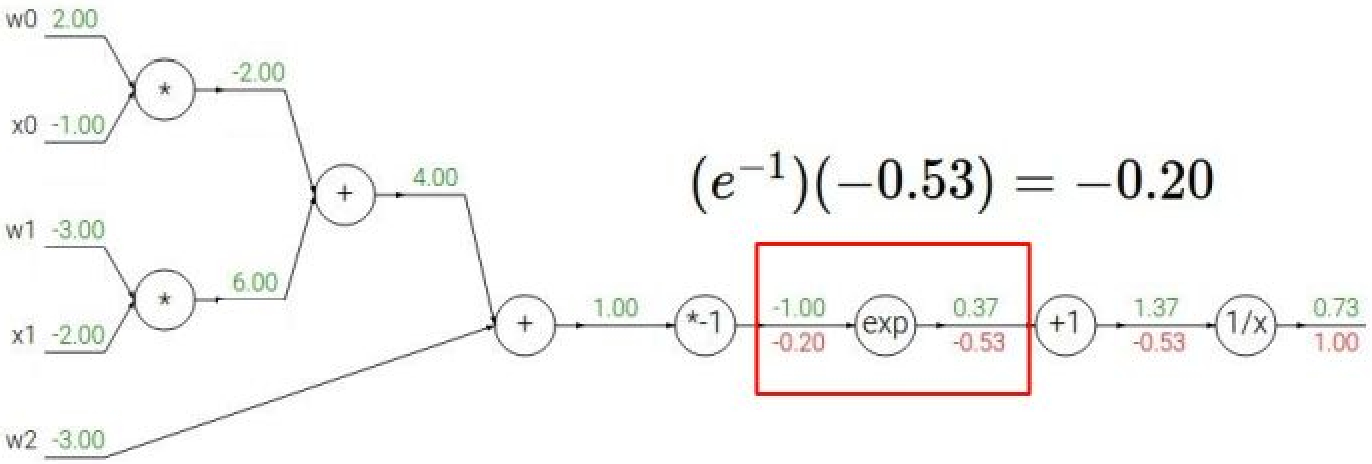
$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

自动求导

Another example: $f(w,x) = \dfrac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



$$(e^{-1})(-0.53) = -0.20$$

$$\boxed{f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x}$$

$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

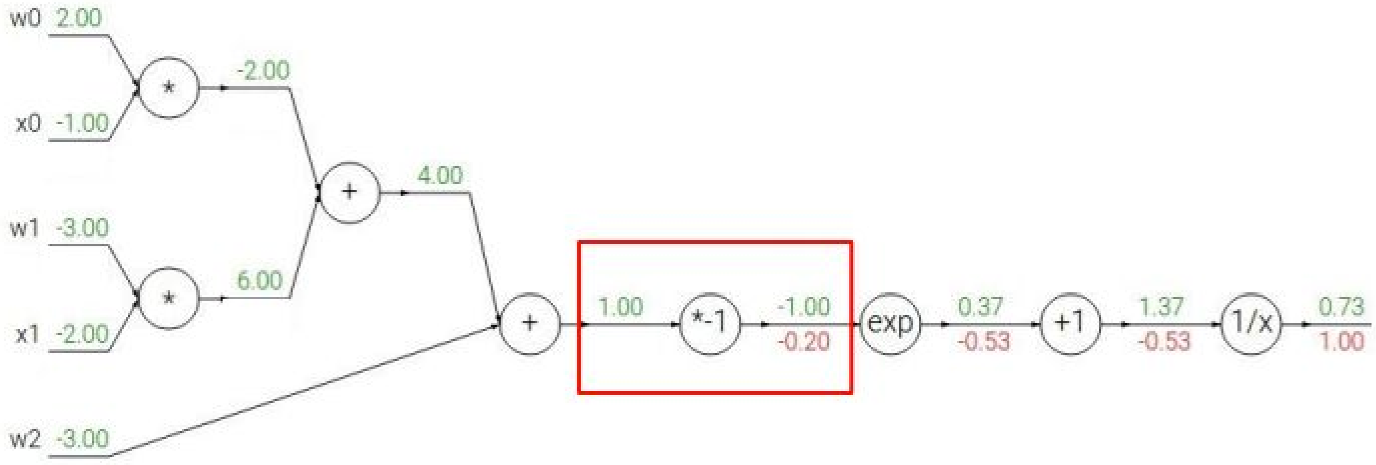
$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

自动求导

Another example:

$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

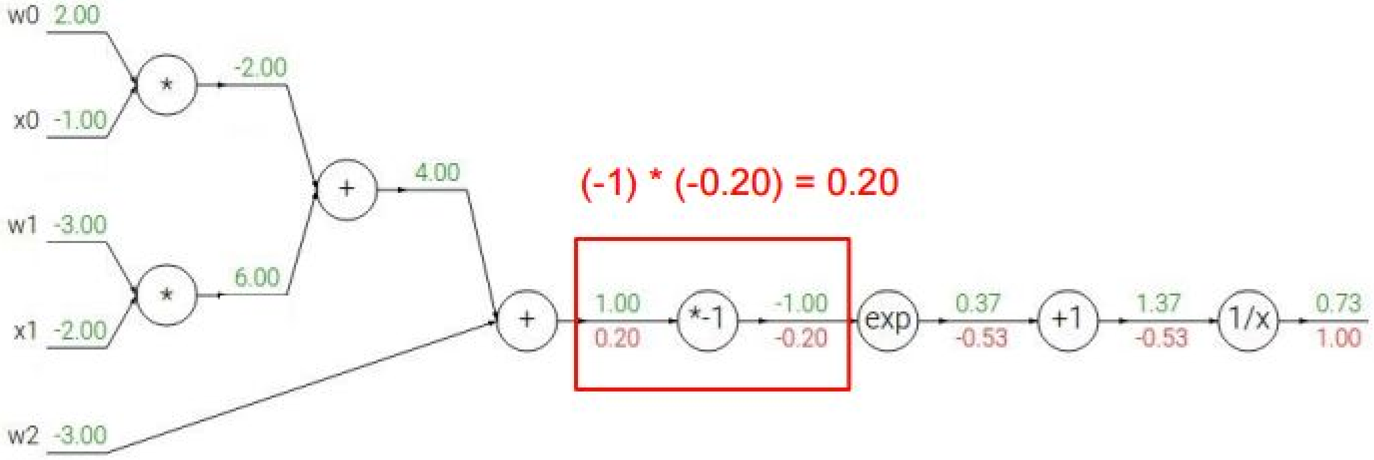
$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

自动求导

Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



(-1) * (-0.20) = 0.20

$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

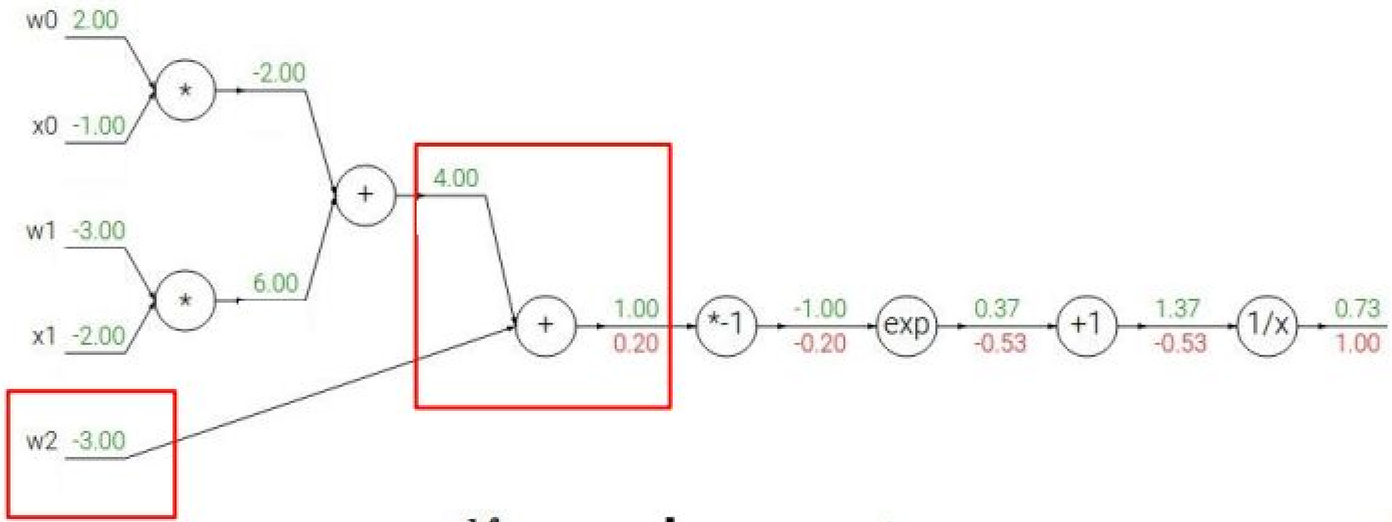
$$\boxed{f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a}$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

自动求导

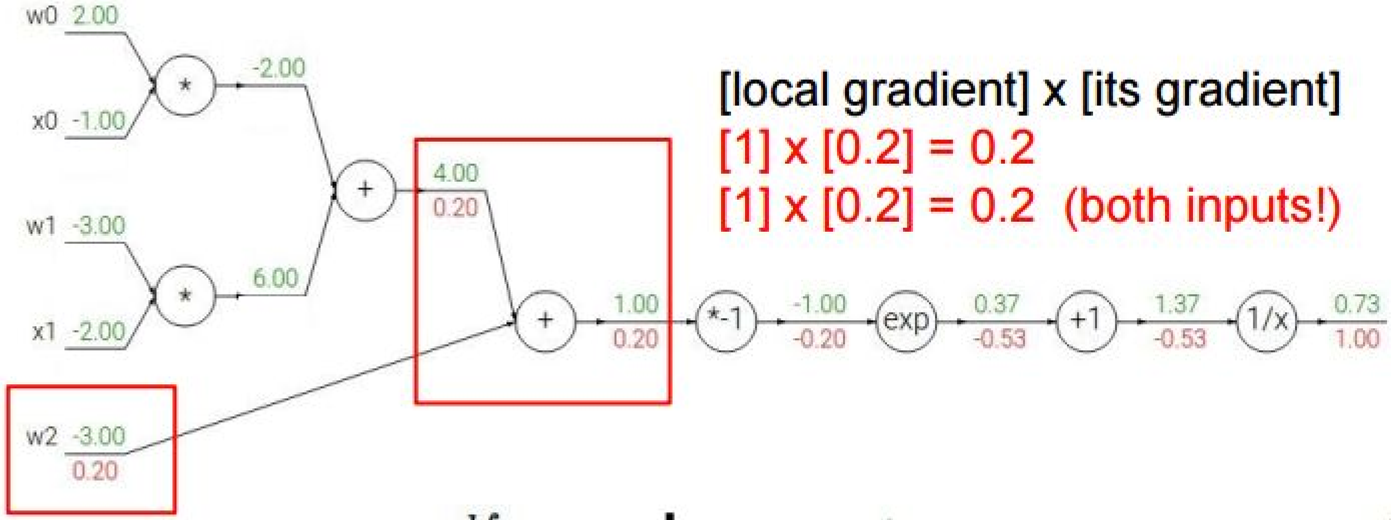
Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



$f(x) = e^x$ → $\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$ → $\frac{df}{dx} = -1/x^2$	
$f_a(x) = ax$ → $\frac{df}{dx} = a$		$f_c(x) = c + x$ → $\frac{df}{dx} = 1$	

自动求导

Another example: $f(w,x) = \dfrac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



[local gradient] x [its gradient]

[1] x [0.2] = 0.2

[1] x [0.2] = 0.2 (both inputs!)

$f(x) = e^x \quad \rightarrow \quad \dfrac{df}{dx} = e^x$

$f_a(x) = ax \quad \rightarrow \quad \dfrac{df}{dx} = a$

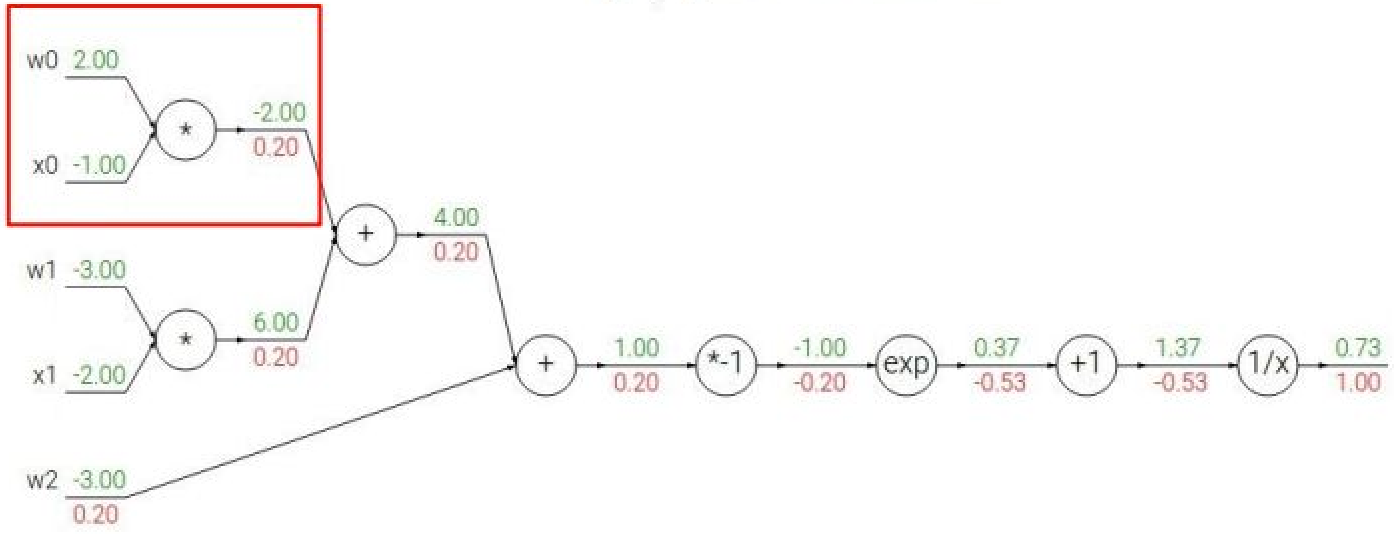
$f(x) = \dfrac{1}{x} \quad \rightarrow \quad \dfrac{df}{dx} = -1/x^2$

$f_c(x) = c + x \quad \rightarrow \quad \dfrac{df}{dx} = 1$

自动求导

Another example:

$$f(w,x)=\frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



$$f(x)=e^x \quad \rightarrow \quad \frac{df}{dx}=e^x$$

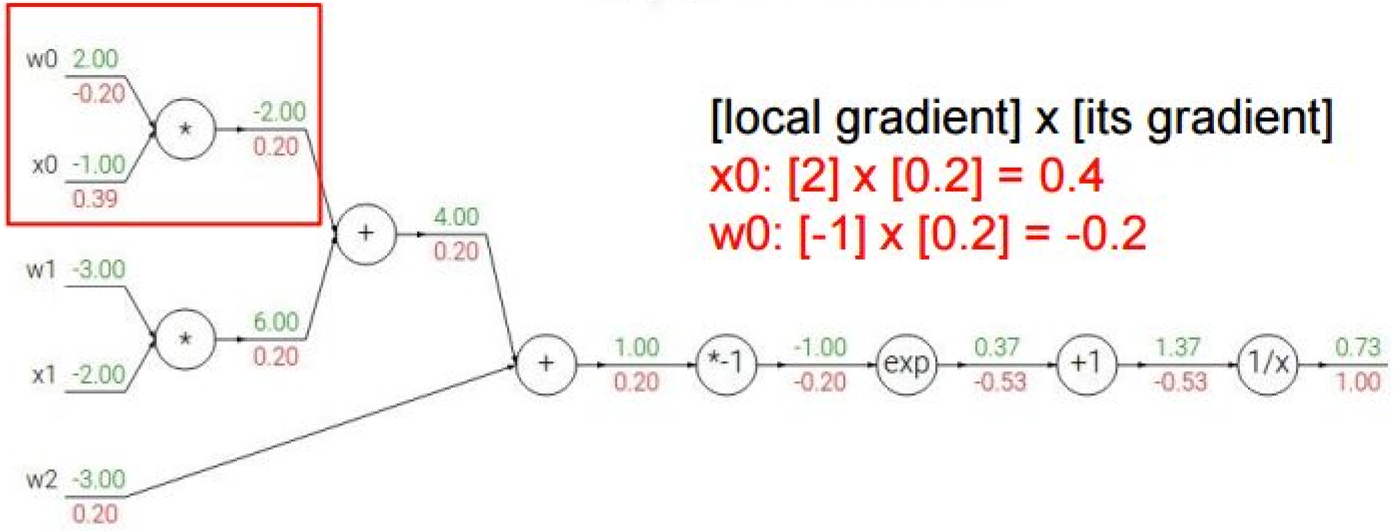
$$f_a(x)=ax \quad \rightarrow \quad \frac{df}{dx}=a$$

$$f(x)=\frac{1}{x} \quad \rightarrow \quad \frac{df}{dx}=-1/x^2$$

$$f_c(x)=c+x \quad \rightarrow \quad \frac{df}{dx}=1$$

自动求导

Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$



[local gradient] x [its gradient]

x0: [2] x [0.2] = 0.4

w0: [-1] x [0.2] = -0.2

$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

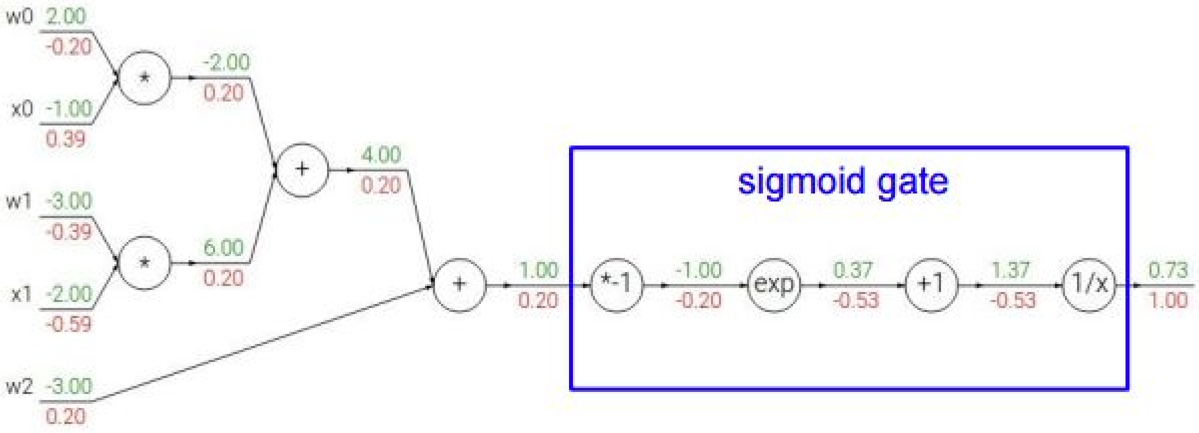
自动求导

$$f(w,x)=\frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

$$\boxed{\sigma(x)=\frac{1}{1+e^{-x}}}$$

sigmoid function

$$\frac{d\sigma(x)}{dx}=\frac{e^{-x}}{(1+e^{-x})^2}=\left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right)=(1-\sigma(x))\,\sigma(x)$$



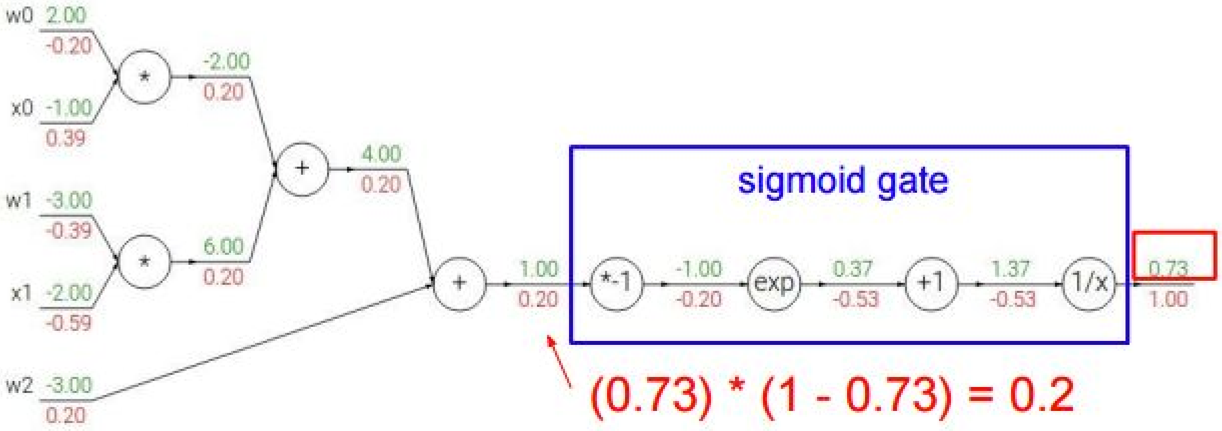
自动求导

$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

$$\boxed{\sigma(x) = \frac{1}{1+e^{-x}}}$$

sigmoid function

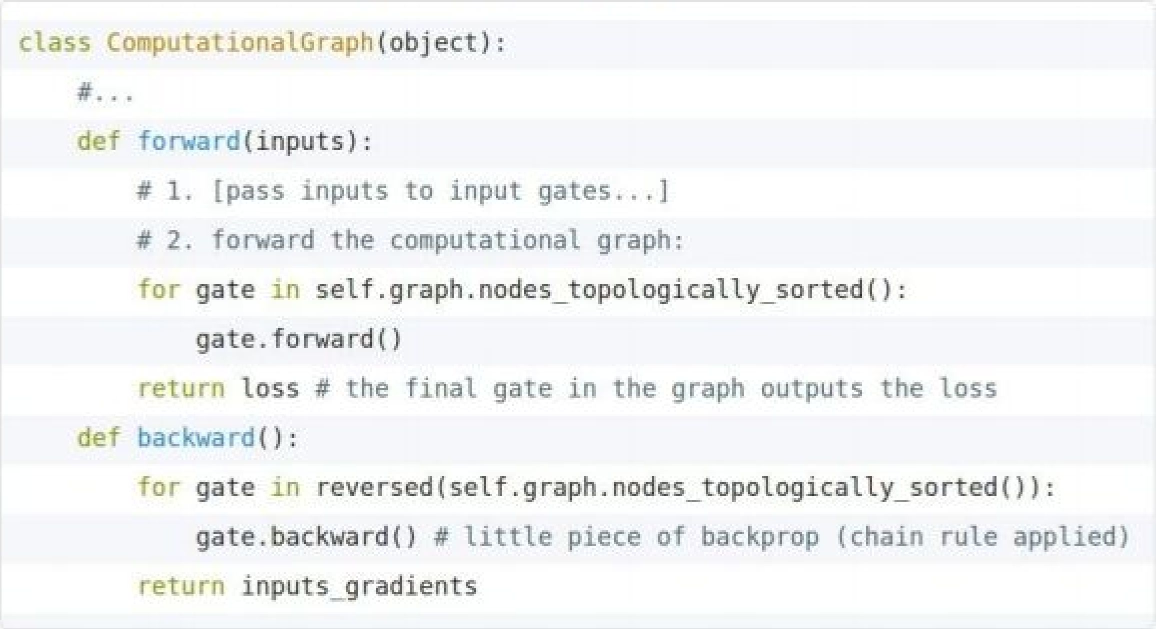
$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} = \left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right) = (1-\sigma(x))\,\sigma(x)$$



(0.73) * (1 - 0.73) = 0.2

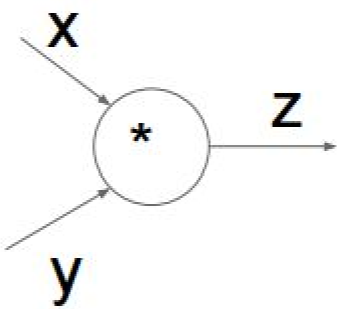
自动求导

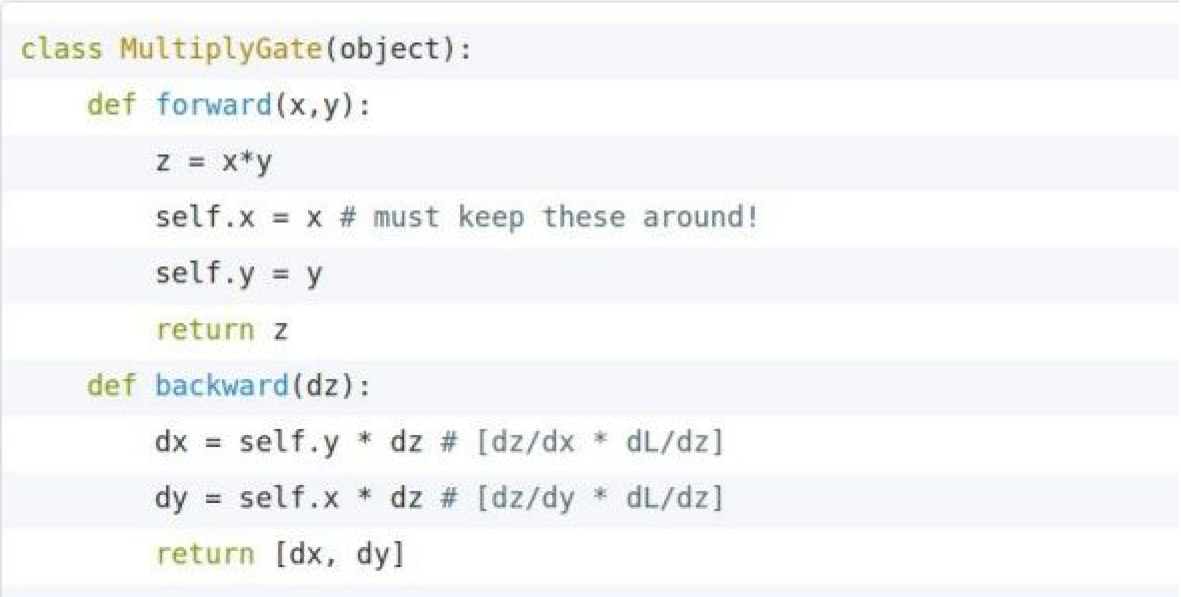
Graph (or Net) object. *(Rough psuedo code)*



```
class ComputationalGraph(object):
    #...
    def forward(inputs):
        # 1. [pass inputs to input gates...]
        # 2. forward the computational graph:
        for gate in self.graph.nodes_topologically_sorted():
            gate.forward()
        return loss # the final gate in the graph outputs the loss
    def backward():
        for gate in reversed(self.graph.nodes_topologically_sorted()):
            gate.backward() # little piece of backprop (chain rule applied)
        return inputs_gradients
```

自动求导





```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        self.x = x # must keep these around!
        self.y = y
        return z
    def backward(dz):
        dx = self.y * dz # [dz/dx * dL/dz]
        dy = self.x * dz # [dz/dy * dL/dz]
        return [dx, dy]
```

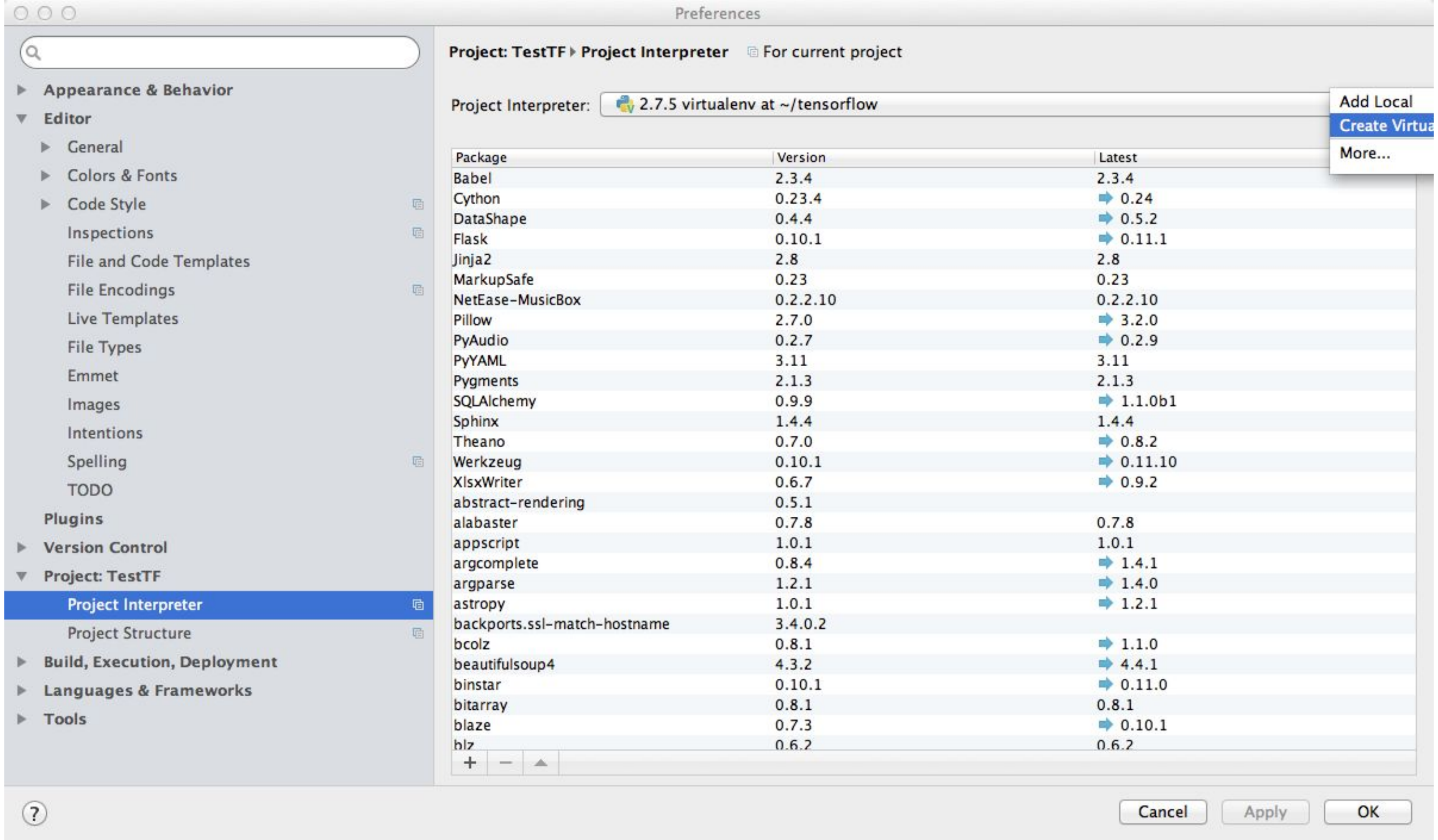
TensorFlow的基本概念

- TensorFlow是什么？
    - 一个图计算平台
    - 它提供一些基本的操作(op)
    - 复杂的计算由这些基本的操作来表示
    - 自动梯度工具
- 基本概念
    - Graph Session op Variable …

安装和Build

- 参考
  - https://www.tensorflow.org/versions/r0.8/how_tos/adding_an_op/index.html
  - https://www.evernote.com/shard/s17/sh/f3938d7a-74c5-4d74-8ad8-d2effb6c49be/796d30bf0e06d086
- 现场示例
  - Mac os x & Ubuntu 14.04(not work in 12.04)

Using virtualenv installation in PyCharm



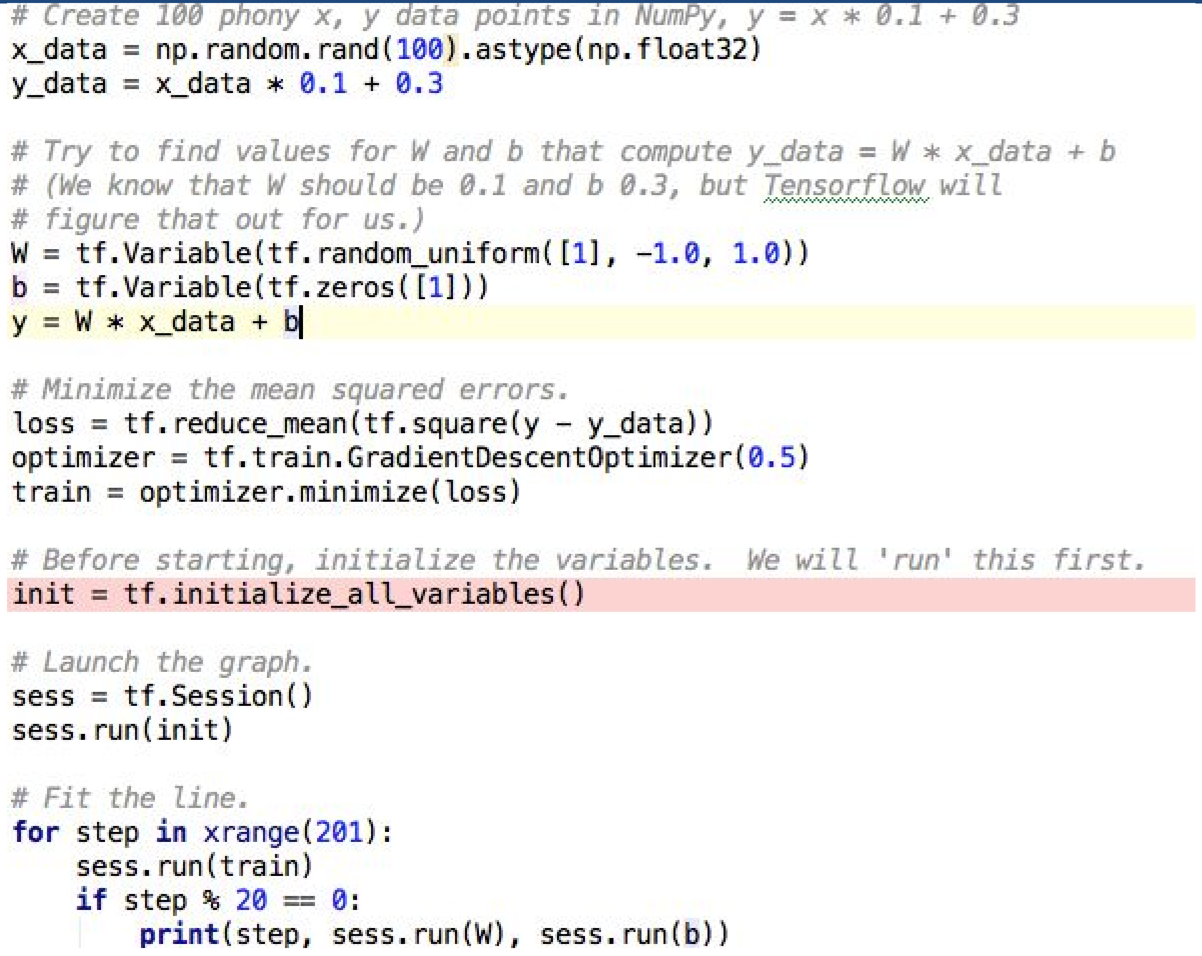
Project: TestTF ▸ Project Interpreter   For current project

Project Interpreter: 2.7.5 virtualenv at ~/tensorflow

Package	Version	Latest
Babel	2.3.4	2.3.4
Cython	0.23.4	➡ 0.24
DataShape	0.4.4	➡ 0.5.2
Flask	0.10.1	➡ 0.11.1
Jinja2	2.8	2.8
MarkupSafe	0.23	0.23
NetEase-MusicBox	0.2.2.10	0.2.2.10
Pillow	2.7.0	➡ 3.2.0
PyAudio	0.2.7	➡ 0.2.9
PyYAML	3.11	3.11
Pygments	2.1.3	2.1.3
SQLAlchemy	0.9.9	➡ 1.1.0b1
Sphinx	1.4.4	1.4.4
Theano	0.7.0	➡ 0.8.2
Werkzeug	0.10.1	➡ 0.11.10
XlsxWriter	0.6.7	➡ 0.9.2
abstract-rendering	0.5.1	
alabaster	0.7.8	0.7.8
appscript	1.0.1	1.0.1
argcomplete	0.8.4	➡ 1.4.1
argparse	1.2.1	➡ 1.4.0
astropy	1.0.1	➡ 1.2.1
backports.ssl-match-hostname	3.4.0.2	
bcolz	0.8.1	➡ 1.1.0
beautifulsoup4	4.3.2	➡ 4.4.1
binstar	0.10.1	➡ 0.11.0
bitarray	0.8.1	0.8.1
blaze	0.7.3	➡ 0.10.1
blz	0.6.2	0.6.2

示例——Get Started

- 基本概念
- 自动求导



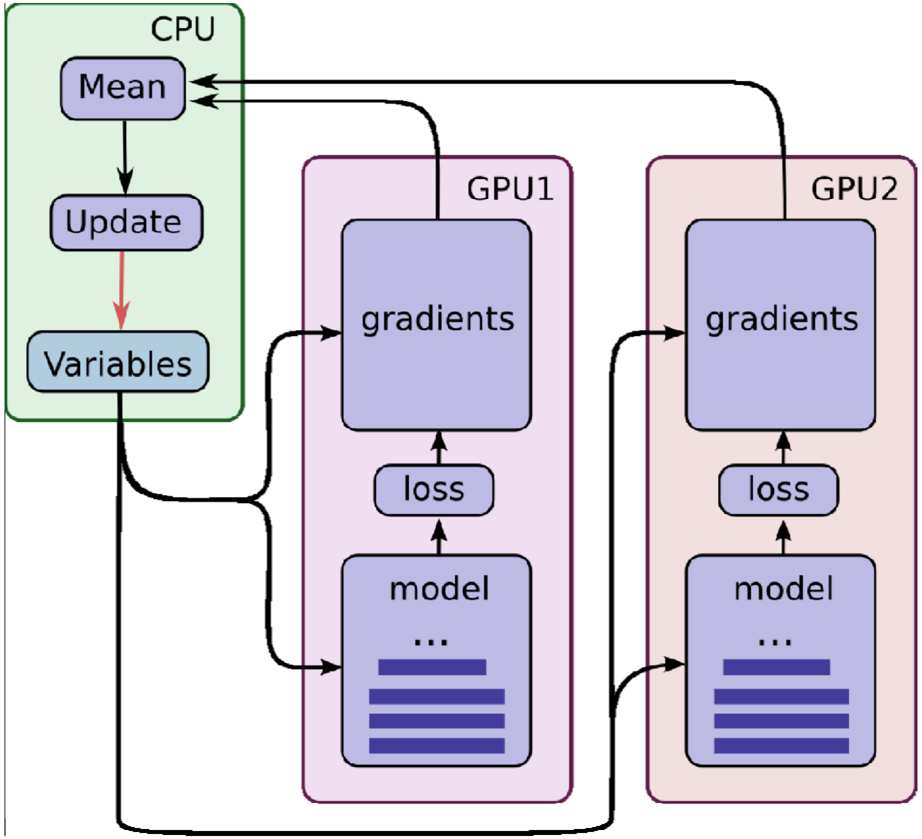
```
# Create 100 phony x, y data points in NumPy, y = x * 0.1 + 0.3
x_data = np.random.rand(100).astype(np.float32)
y_data = x_data * 0.1 + 0.3

# Try to find values for W and b that compute y_data = W * x_data + b
# (We know that W should be 0.1 and b 0.3, but Tensorflow will
# figure that out for us.)
W = tf.Variable(tf.random_uniform([1], -1.0, 1.0))
b = tf.Variable(tf.zeros([1]))
y = W * x_data + b

# Minimize the mean squared errors.
loss = tf.reduce_mean(tf.square(y - y_data))
optimizer = tf.train.GradientDescentOptimizer(0.5)
train = optimizer.minimize(loss)

# Before starting, initialize the variables.  We will 'run' this first.
init = tf.initialize_all_variables()

# Launch the graph.
sess = tf.Session()
sess.run(init)

# Fit the line.
for step in xrange(201):
    sess.run(train)
    if step % 20 == 0:
        print(step, sess.run(W), sess.run(b))
```

示例——MNIST for Experts

- InteractiveSession
- Computation Graph
- Placeholders
- Variables

示例——CIFAR10

- CPU vs GPU
- TensorBoard
- Prefetching Queue
- AlexNet Components



示例——Add Custom Op

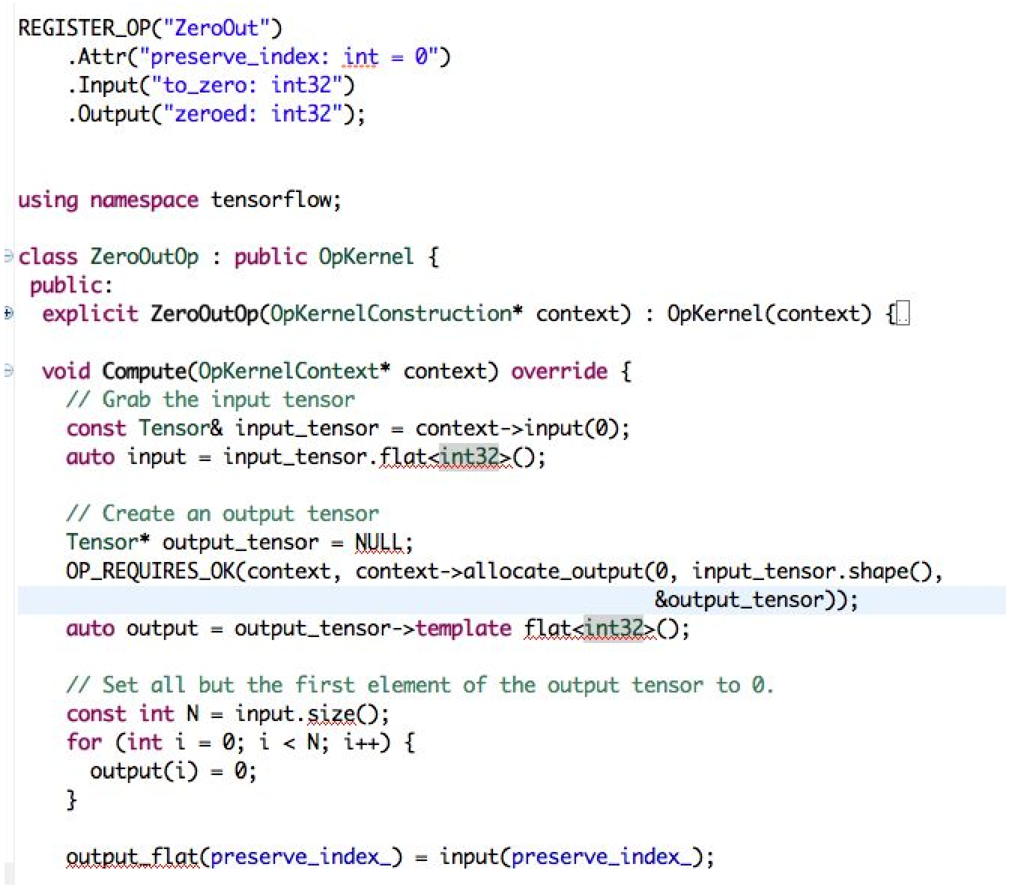
Define the Op's interface

Implement the kernel for the Op

Building the Op library

Using the Op in Python

GPU Support

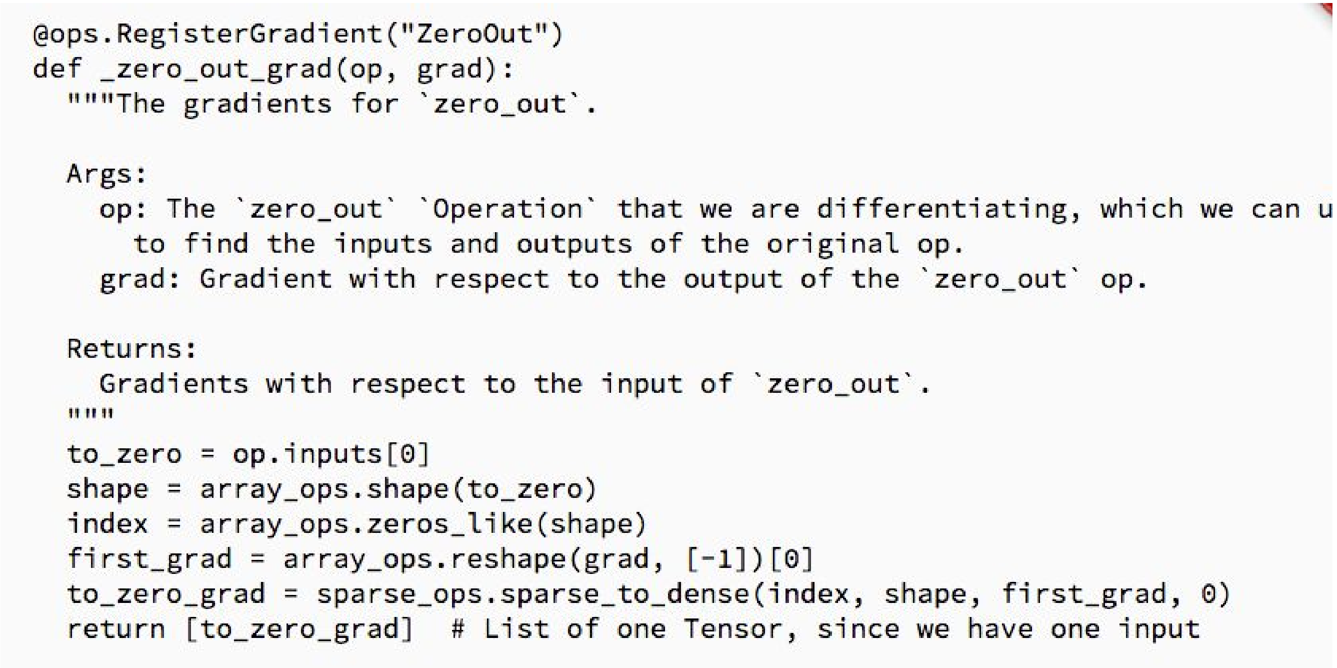


```
REGISTER_OP("ZeroOut")
    .Attr("preserve_index: int = 0")
    .Input("to_zero: int32")
    .Output("zeroed: int32");


using namespace tensorflow;

class ZeroOutOp : public OpKernel {
 public:
  explicit ZeroOutOp(OpKernelConstruction* context) : OpKernel(context) {

  void Compute(OpKernelContext* context) override {
    // Grab the input tensor
    const Tensor& input_tensor = context->input(0);
    auto input = input_tensor.flat<int32>();

    // Create an output tensor
    Tensor* output_tensor = NULL;
    OP_REQUIRES_OK(context, context->allocate_output(0, input_tensor.shape(),
                                                     &output_tensor));
    auto output = output_tensor->template flat<int32>();

    // Set all but the first element of the output tensor to 0.
    const int N = input.size();
    for (int i = 0; i < N; i++) {
      output(i) = 0;
    }

    output_flat(preserve_index_) = input(preserve_index_);
```

示例——Add Custom Op

为什么用Python而不是C++?

为什么只能用基本的op来构造？



```
@ops.RegisterGradient("ZeroOut")
def _zero_out_grad(op, grad):
  """The gradients for `zero_out`.

  Args:
    op: The `zero_out` `Operation` that we are differentiating, which we can u
      to find the inputs and outputs of the original op.
    grad: Gradient with respect to the output of the `zero_out` op.

  Returns:
    Gradients with respect to the input of `zero_out`.
  """
  to_zero = op.inputs[0]
  shape = array_ops.shape(to_zero)
  index = array_ops.zeros_like(shape)
  first_grad = array_ops.reshape(grad, [-1])[0]
  to_zero_grad = sparse_ops.sparse_to_dense(index, shape, first_grad, 0)
  return [to_zero_grad]  # List of one Tensor, since we have one input
```

示例——Tensor Serving

参考资料

1.1 Tensorflow的论文 【tensorflow的官方论文】

http://download.tensorflow.org/paper/whitepaper2015.pdf

1.2 Overview of Caffe/Torch/Theano/TensorFlow【from cs231n，对深度学习感兴趣的同学推荐这个课程，作业设计的非常好，python notebook，每一步都能看到结果】

http://cs231n.stanford.edu/slides/winter1516_lecture12.pdf

1.3 Automatic differentiation in machine learning: a survey 【向机器学习社区的人介绍Automatic differentiation】

http://arxiv.org/abs/1502.05767

1.4 Calculus on Computational Graphs: Backpropagation 【非常通俗易懂，如果觉得1.3太长，看这个也差不多了，另外这个博客其它的文章也很好】

http://colah.github.io/posts/2015-08-Backprop/

1.5 Michael Nielsen的在线书的第二章 【非常详细的backprop推导，即使大学数学忘光了，也能看懂，初学的同学建议把这本书的每个公式都走一遍，python的代码跑一跑】

http://neuralnetworksanddeeplearning.com/chap2.html

参考资料

2.1 安装文档 【目前最新的已经是0.9了，不过我之前测试的都是0.8，而且官方提供的whl也是0.8的】

https://www.tensorflow.org/versions/r0.8/get_started/os_setup.html

2.2 自己从源码build 【0.8的whl安装了后跑gpu的例子会coredump，需要自己从源码编译】

https://www.evernote.com/shard/s17/sh/f3938d7a-74c5-4d74-8ad8-d2effb6c49be/796d30bf0e06d086

3. tutorial

3.1 官方的tutorial，可以装了后在自己的机器上跑一跑

https://www.tensorflow.org/versions/r0.8/tutorials/index.html

3.2 CNN做情感分类，有tensorflow代码，第一篇是原理，第二篇是代码

http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/

http://www.wildml.com/2015/12/implementing-a-cnn-for-text-classification-in-tensorflow/

参考资料

3.3 一个不错的tensorflow的tutorial，包括常见的CNN，RNN等
https://github.com/aymericdamien/TensorFlow-Examples

3.4 ICLR 2016的代码集合，包括tensorflow，caffe，torch，theano
https://tensortalk.com/?cat=conference-iclr-2016#rd&sukey=3997c0719f1515200445e3df04727bf63c2922b70f48ec81889d4307608aaacc8cb0dd2fffe47724b5e687a6d53d4e09

3.5 CS224d的第七章介绍tensorflow【这个课程也强烈推荐】
http://cs224d.stanford.edu/lectures/CS224d-Lecture7.pdf

3.6 Awesome TensorFlow【名字很对，资源很丰富】
https://github.com/jtoy/awesome-tensorflow/

3.7 Awesome Tensorflow Implementations【论文和tensorflow的实现，非常酷，之前的大部分是非常成熟的东西，比如CNN/LSTM，这里有最新的进展，如Sequence to Sequence -- chatbot，Show and Tell: A Neural Image Caption Generator，Using Deep Q-Network to Learn How To Play Flappy Bird，...】
https://github.com/TensorFlowKR/awesome_tensorflow_implementations

3.8 google 的 deep learning course【当然是用tensor flow了，很基础的课程】
https://cn.udacity.com/course/deep-learning--ud730/

谢谢！

出门问问 | 李理
lli@mobvoi.com