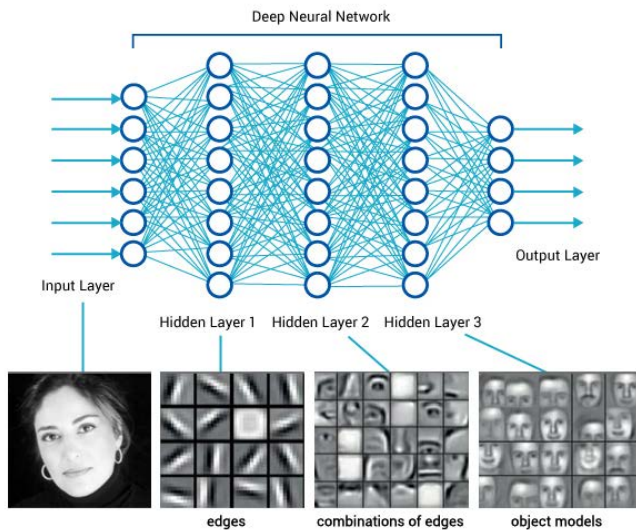


Recurrent Neural Network

The Deepest of All Deep Learning



Slides by
Chen Liang



Deep Learning

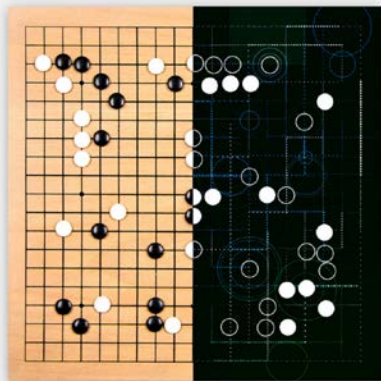
Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

AI Is Transforming Google Search. The Rest of the Web Is Next

IN A HUGE BREAKTHROUGH, GOOGLE'S AI BEATS A TOP PLAYER AT THE GAME OF GO

Drive.ai to test 'deep learning' autonomous cars on California roads



Baidu's Deep-Learning System Rivals People at Speech Recognition

Deep learning

Field of study

Machine learning

Field of study

Artificial neural n..

Topic

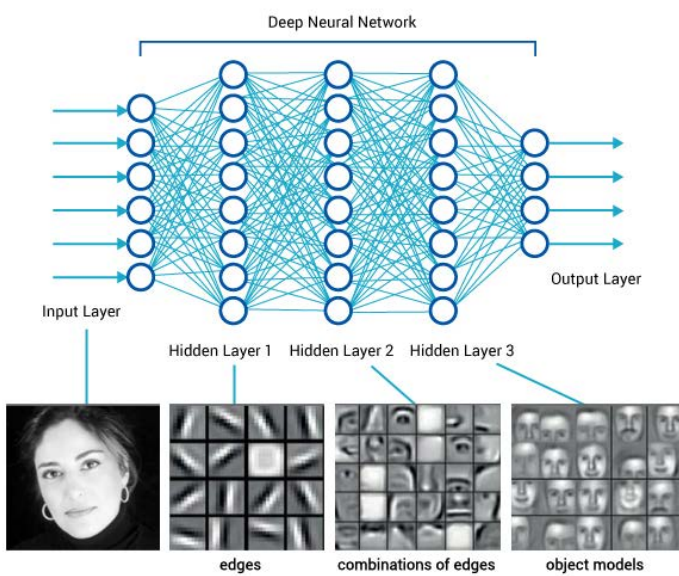
+ Add term

Beta: Measuring search interest in *topics* is a beta feature which quickly provides accurate measurements of overall search interest. To measure search interest for a specific *query*, select the "search term" option. ?

Interest over time ?

☐ News headlines ? ☐ Forecast ?

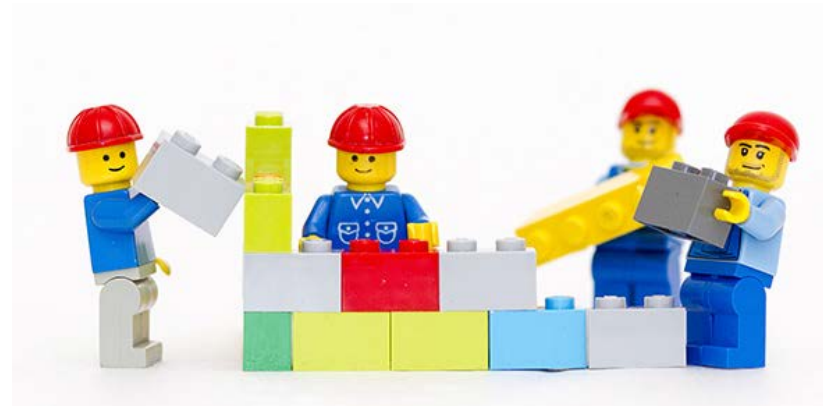




Demystify Deep Learning



Deep learning works
like the human brain?



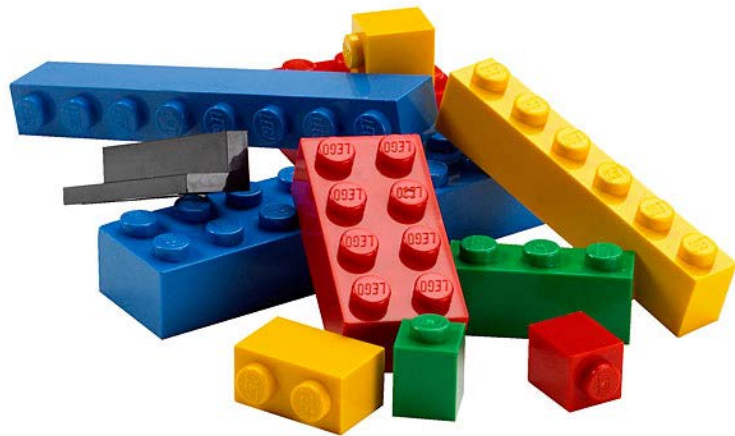
Deep Learning: Building Blocks

- ▶ Linear transformations

$$h_{k+1} = Wh_k$$

- ▶ Non-linear activation functions

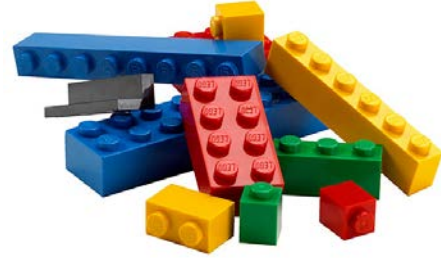
$$h_{k+2} = f(h_{k+1})$$



Deep Learning: Deep Composition

- ▶ A **deep representation** is a composition of many functions

$$x \xrightarrow{w_1} h_1 \xrightarrow{w_2} \dots \xrightarrow{w_n} h_n \xrightarrow{w_{n+1}} y$$



- ▶ Its gradient can be **backpropagated** by the chain rule

$$\begin{array}{ccccccc} \frac{\partial h_1}{\partial x} & \leftarrow & \frac{\partial h_2}{\partial h_1} & \leftarrow & \dots & \leftarrow & \frac{\partial y}{\partial h_n} & \leftarrow & \frac{\partial}{\partial y} \\ & & \downarrow & & & & \downarrow & & \downarrow \\ & & \frac{\partial h_1}{\partial w_1} & & \dots & & \frac{\partial h_n}{\partial w_n} & & \frac{\partial y}{\partial w_{n+1}} \end{array}$$



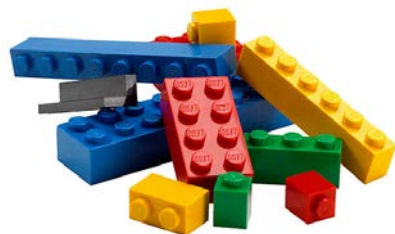
Deep Learning: Gradient Descent

- ▶ A **loss function** $l(y)$ measures goodness of output y , e.g.
 - ▶ Mean-squared error $l(y) = ||y^* - y||^2$
 - ▶ Log likelihood $l(y) = \log \mathbb{P}[y^*|x]$
- ▶ The loss is appended to the forward computation

$$x \xrightarrow{w_1} h_1 \xrightarrow{w_2} \dots \xrightarrow{w_n} h_n \xrightarrow{w_{n+1}} y \longrightarrow l(y)$$

- ▶ Gradient of loss is appended to the backward computation

$$\begin{array}{ccccccc} \frac{\partial h_1}{\partial x} & \longleftarrow & \frac{\partial h_2}{\partial h_1} & \longleftarrow & \dots & \longleftarrow & \frac{\partial y}{\partial h_n} \longleftarrow \frac{\partial l(y)}{\partial y} \\ & & \downarrow & & & & \downarrow \\ & & \frac{\partial h_1}{\partial w_1} & & \dots & & \frac{\partial h_n}{\partial w_n} \quad \frac{\partial y}{\partial w_{n+1}} \end{array}$$



Deep Learning: Demo TensorFlow Playground



Iterations
000,000

Learning rate

0.03

Activation

ReLU

Regularization

L2

Regularization rate

0

Problem type

Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 50%

Noise: 0

Batch size: 10

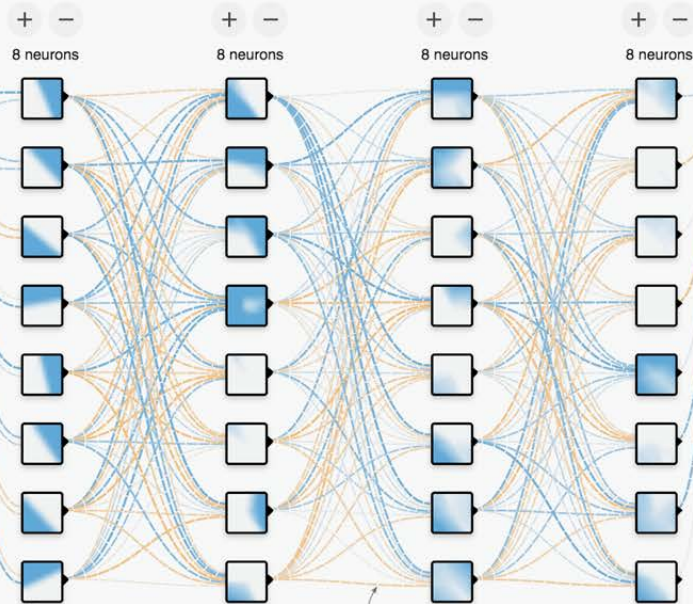
REGENERATE

INPUT

Which properties do you want to feed in?

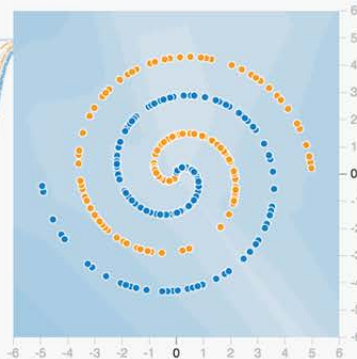
X_1
 X_2
 X_1^2
 X_2^2
 $X_1 X_2$
 $\sin(X_1)$
 $\sin(X_2)$

4 HIDDEN LAYERS



OUTPUT

Test loss 0.544
Training loss 0.534



Colors shows data, neuron and weight values.

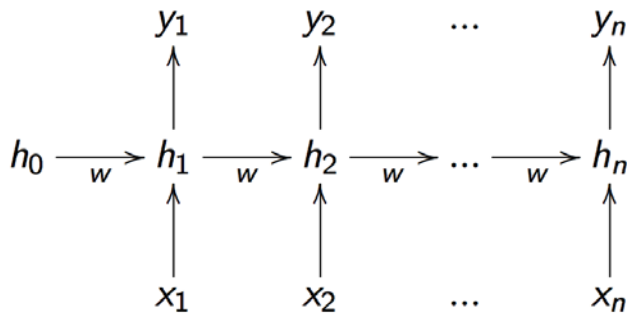
☐ Show test data ☐ Discretize output

This is the output from one neuron. Hover to see it.

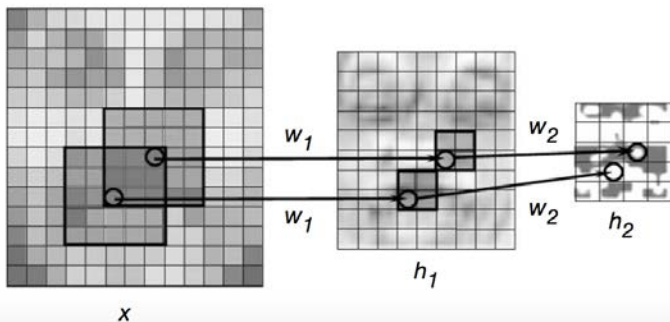
The outputs are mixed with varying weights, shown by the thickness.

Deep Learning: Weight Sharing

Recurrent neural network shares weights between time-steps



Convolutional neural network shares weights between local regions



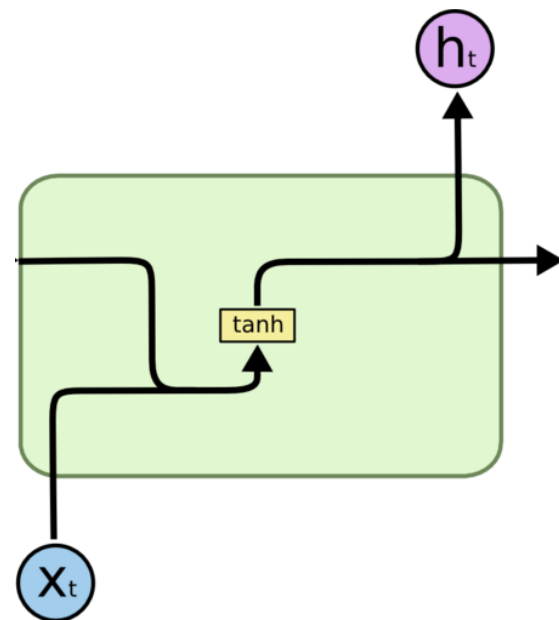
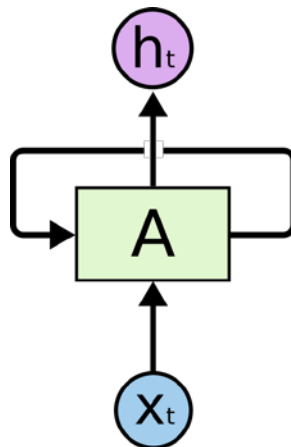
Recurrent Neural Network

Deepest of Deep learning?

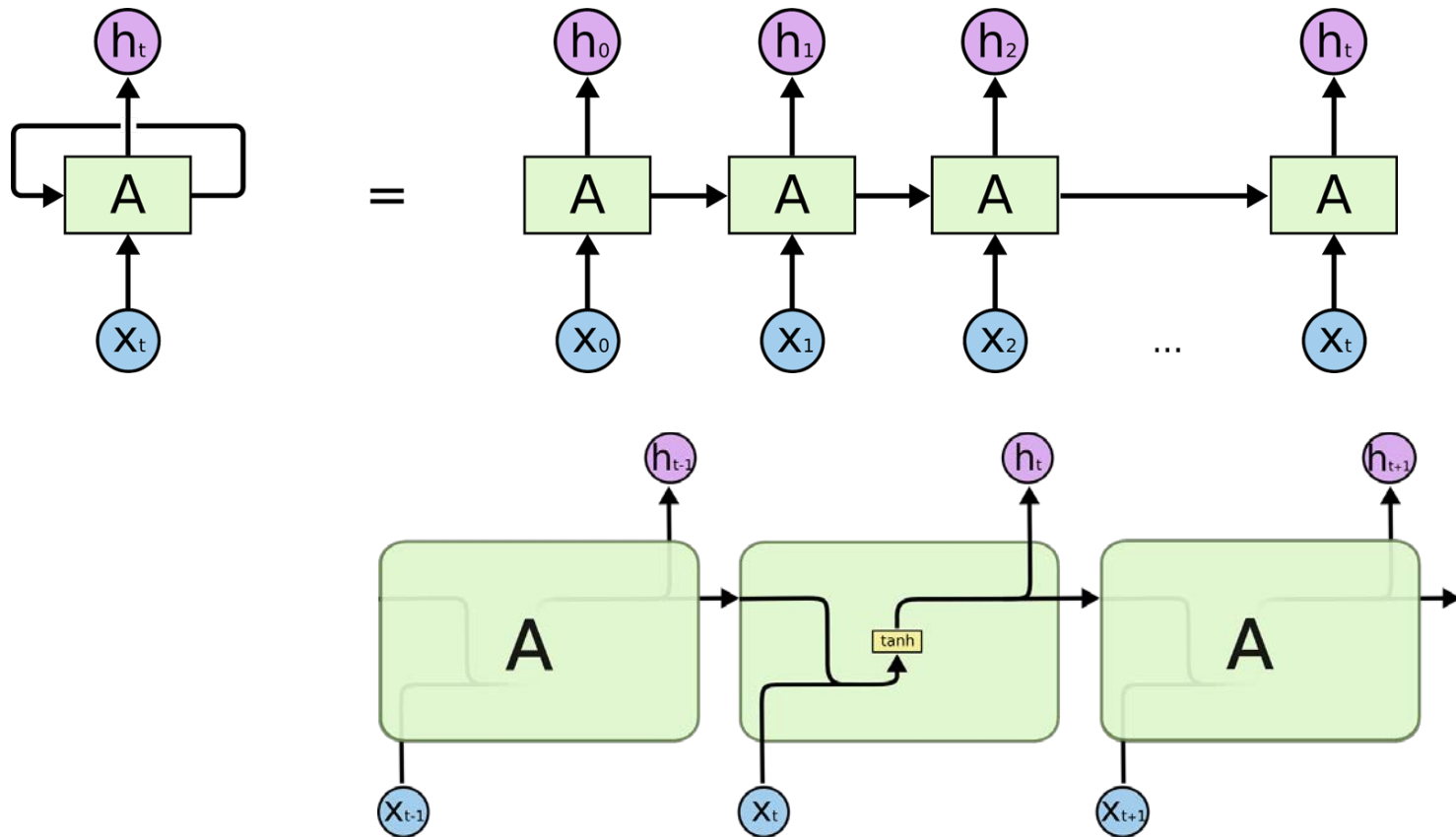
- Can be infinitely deep

Equation

$$h_t^l = \tanh W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$



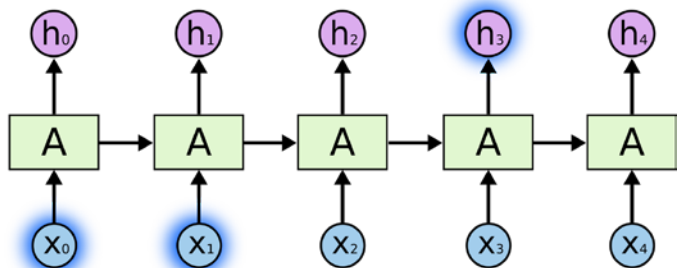
BPTT: Backpropagation Through Time



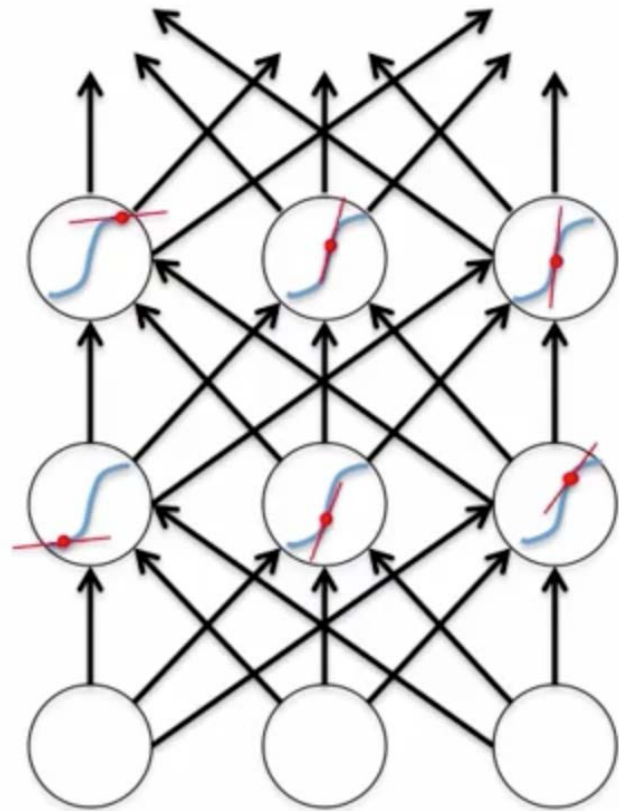
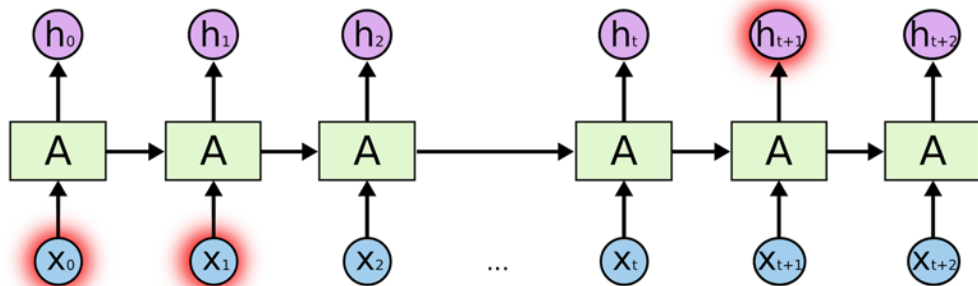
RNN is Turing complete, but...

Exploding/vanishing gradient

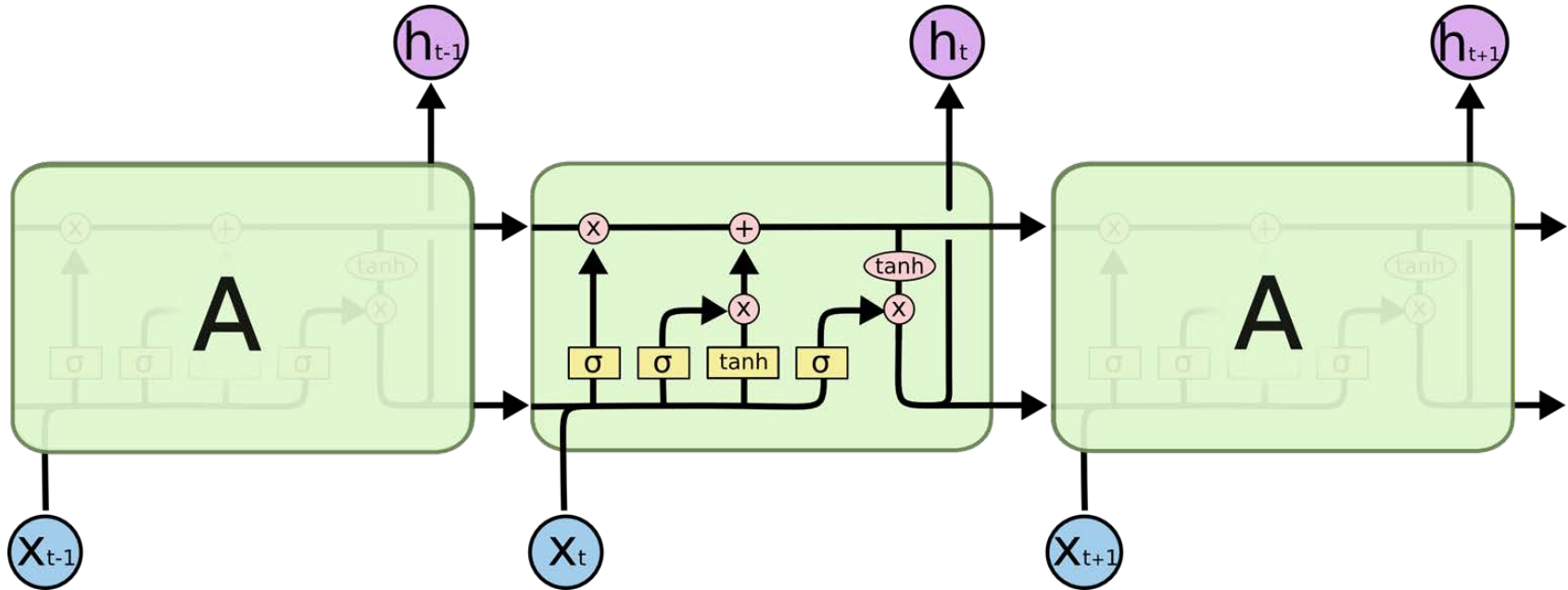
Short term dependency



Long term dependency

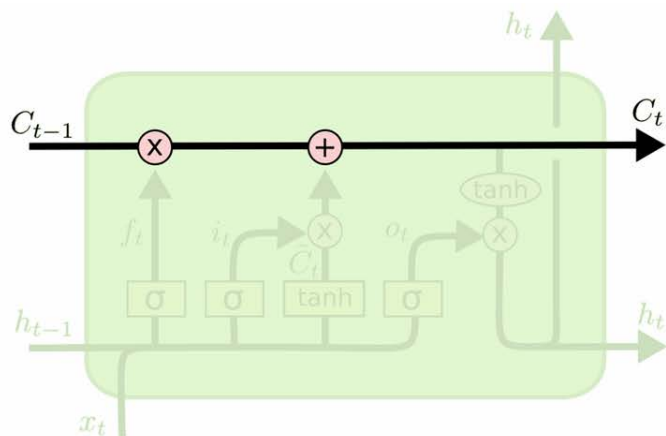


LSTM: Long-Short Term Memory



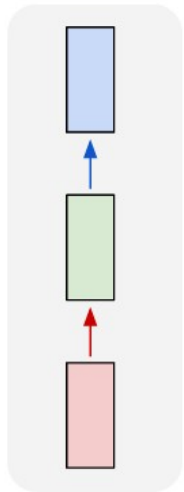
LSTM: Long-Short Term Memory

Let the gradient flow to earlier steps



RNN: A General Framework

one to one



one to many

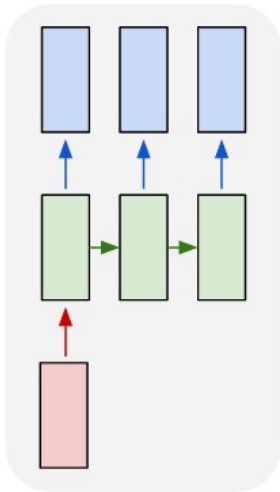
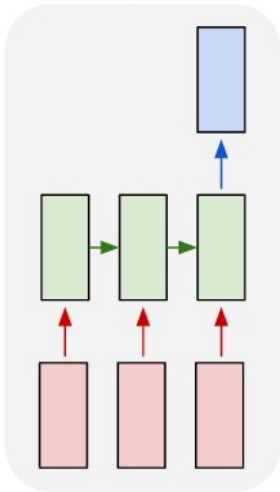


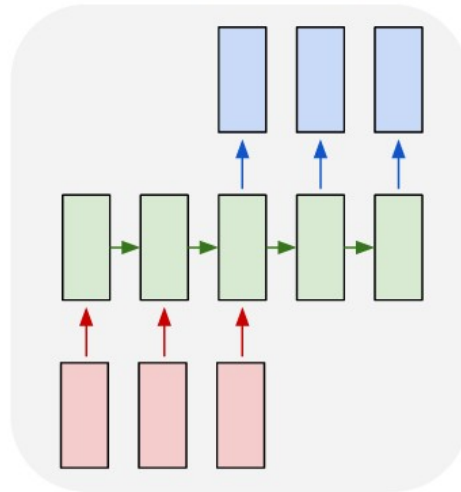
Image Caption
Generation

many to one



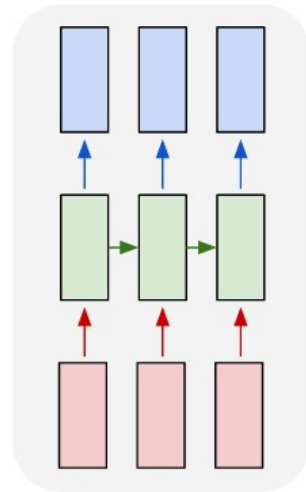
Sentiment Analysis
Image Recognition

many to many



Machine Translation
Speech recognition

many to many



Language Modeling

Char-RNN

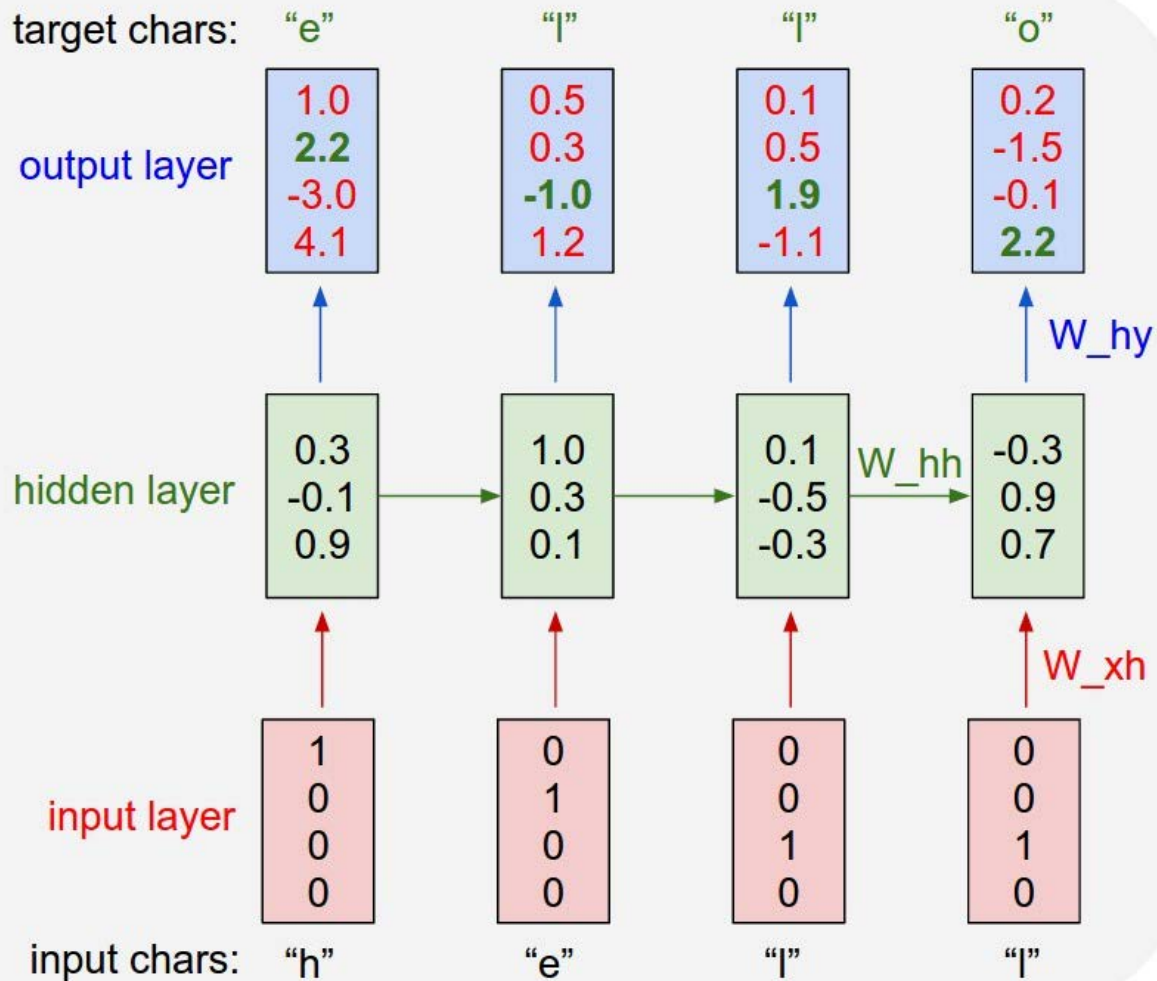
How it works?

Vocabulary:

["h", "e", "l", "o"]

Training
sequence:

"hello"



Char-RNN

Linux

Latex

Wikipedia

Music

...

Check out the blog:

[The Unreasonable Effectiveness of RNN](#)



DeepDrumpf

@DeepDrumpf

Follow

We have competence. Our people don't need anybody. I have smart people.

1:46 PM - 3 Mar 2016



127



101



DeepDrumpf

@DeepDrumpf

Follow

I'll bring back our jobs. They all have everything, they don't have a clue.

1:36 PM - 3 Mar 2016



24



21



DeepDrumpf

@DeepDrumpf

Follow

OK, it's amazing right now with ISIS, I tell you what? I don't want them to vote, the worst very social people. I love me.

1:49 PM - 3 Mar 2016



327



371

What does the Neuron do?

Cell sensitive to position in line:

```
The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.
```

Cell that turns on inside quotes:

```
"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.
```

```
Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."
```

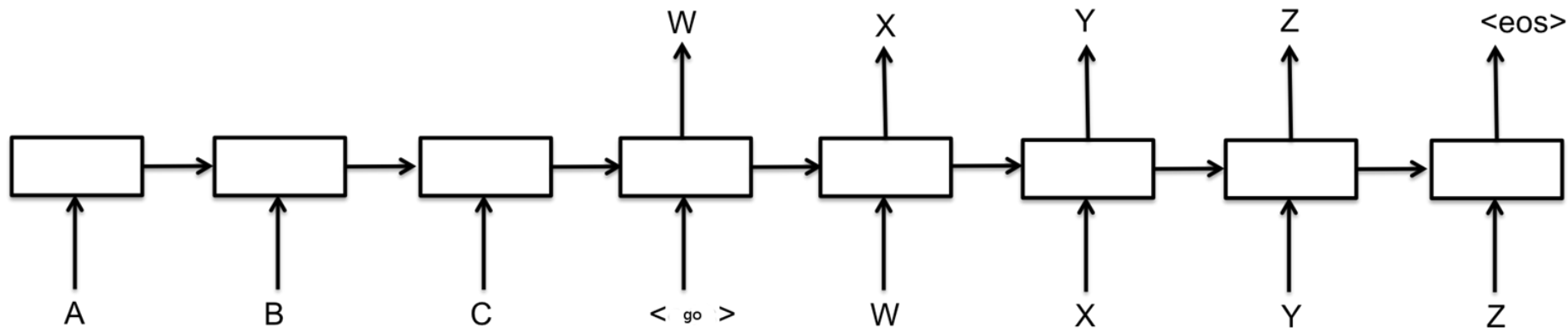
Cell that robustly activates inside if statements:

```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
                           siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier_mask, sig)) {
                if (!(current->notifier)(current->notifier_data)) {
                    clear_thread_flag(TIF_SIGPENDING);
                    return 0;
                }
            }
        }
        collect_signal(sig, pending, info);
    }
    return sig;
}
```

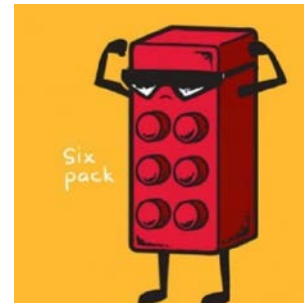
A large portion of cells are not easily interpretable. Here is a typical example:

```
/* Unpack a filter field's string representation from user-space
 * buffer. */
char *audit_unpack_string(void **bufp, size_t *remain, size_t len)
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* Of the currently implemented string fields, PATH_MAX
     * defines the longest valid length.
     */
}
```

Seq2seq: sequence-to-sequence learning



Sequence is even longer now => Attention



Again, let the gradient flow to earlier steps

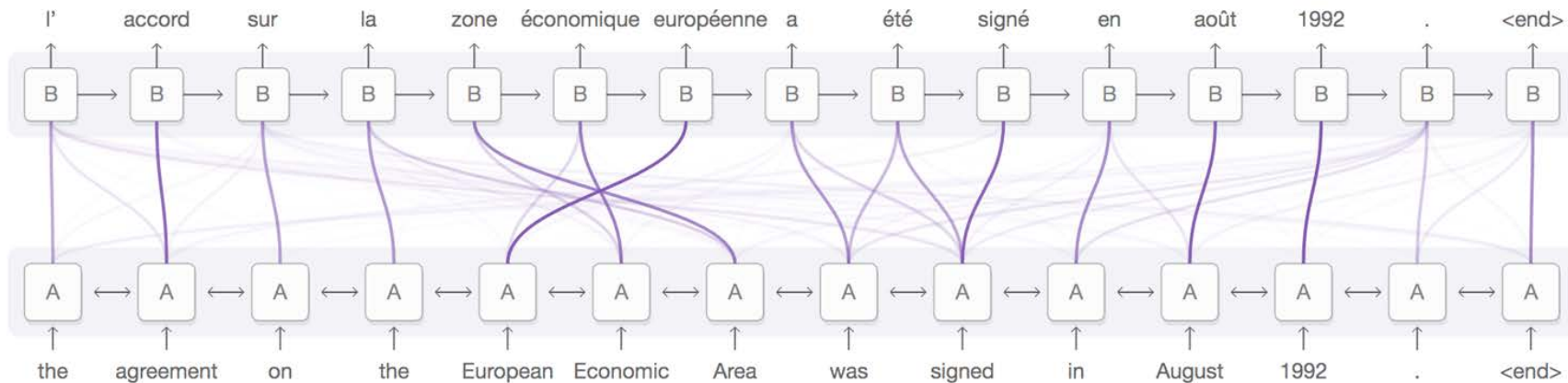
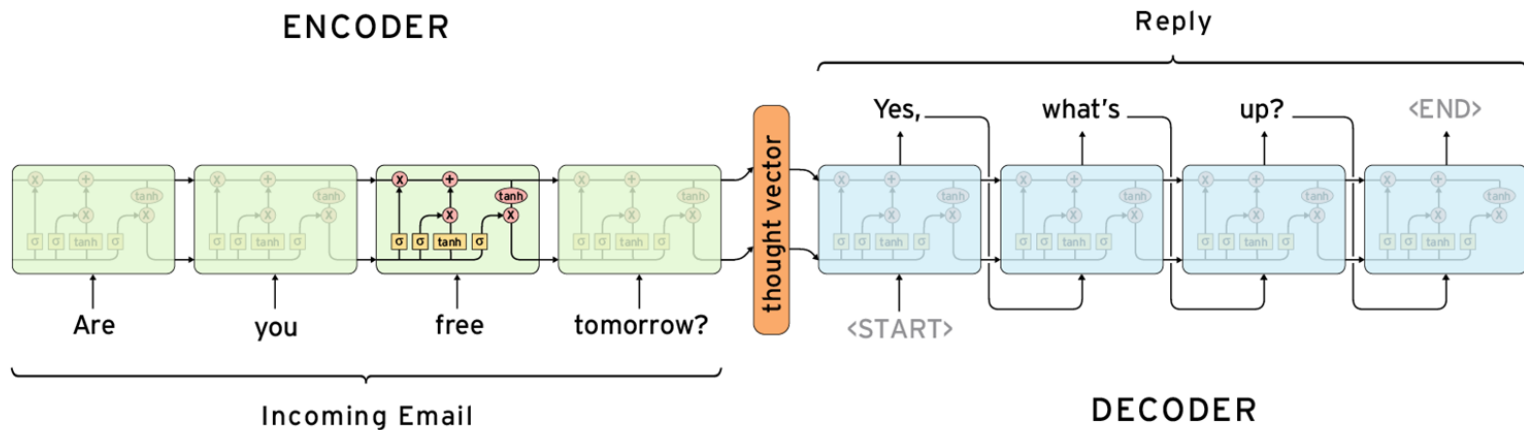
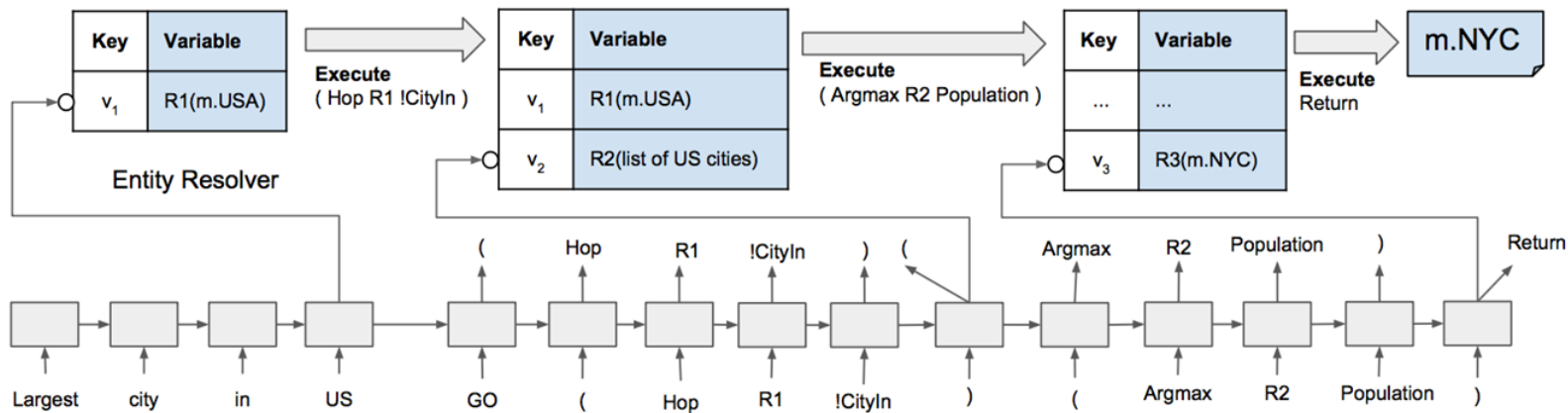


Diagram derived from Fig. 3 of [Bahdanau, et al. 2014](#)

More Seq2seq fun: Chatbots



More Seq2seq fun: Programmers



Summary

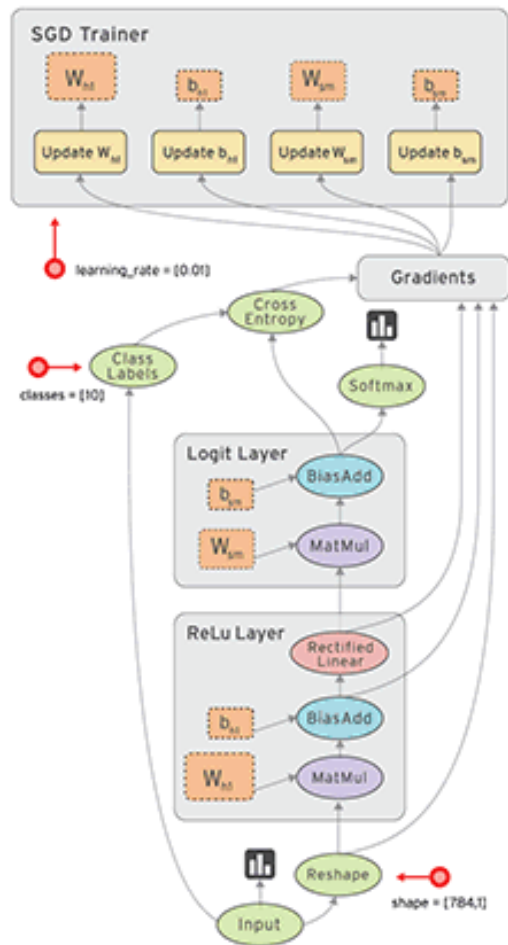
- Deep Learning is like Lego Blocks
=> Compositionality and Backpropagation
- Creative ways to combine the blocks
=> New applications
- When you have problem
=> Get new blocks (LSTM, attention) to let gradient flow

Food for thought: is our brain just a bunch of Lego blocks?



TensorFlow

TensorFlow™ is an open source software library for numerical computation using data flow graphs.



TensorFlow: Computation Graph

```
import tensorflow as tf
import numpy as np
```

```
# 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)
```

Import TensorFlow
and NumPy

TensorFlow: Computation Graph

```
import tensorflow as tf
import numpy as np

# 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)
```

Synthesize some
noisy data from a
linear model

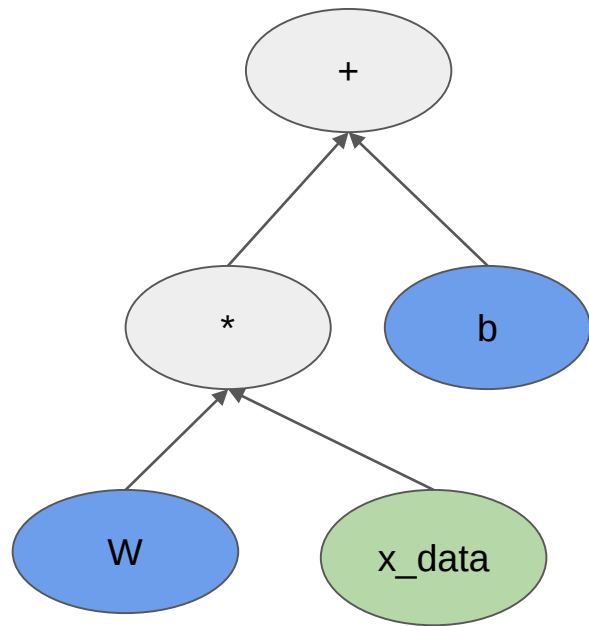
TensorFlow: Computation Graph

```
import tensorflow as tf
import numpy as np

# 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)
```



TensorFlow: Computation Graph

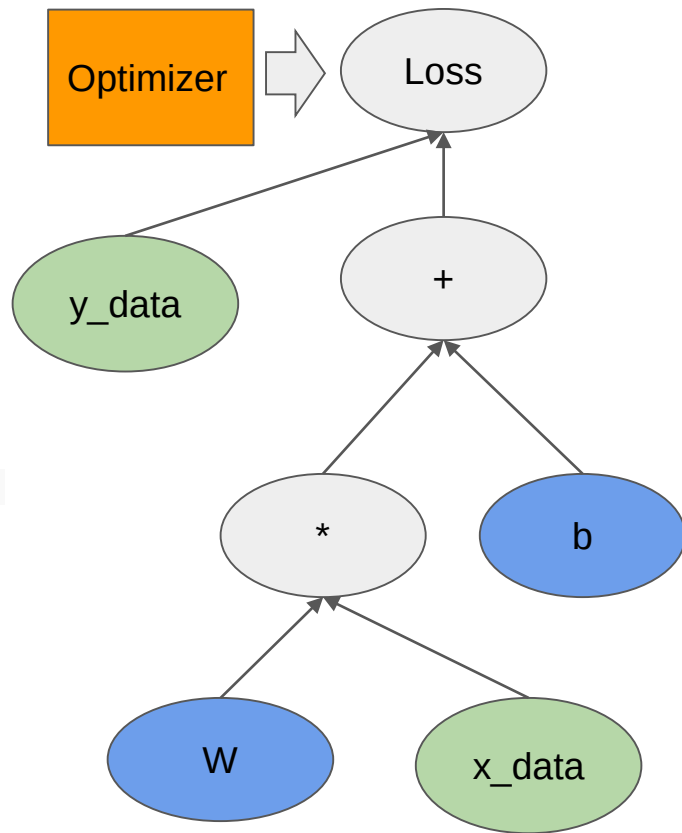
```
import tensorflow as tf
import numpy as np

# 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  $y_{data} = W * x_{data} + b$ 
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)
```



TensorFlow: Session

```
# 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.
```

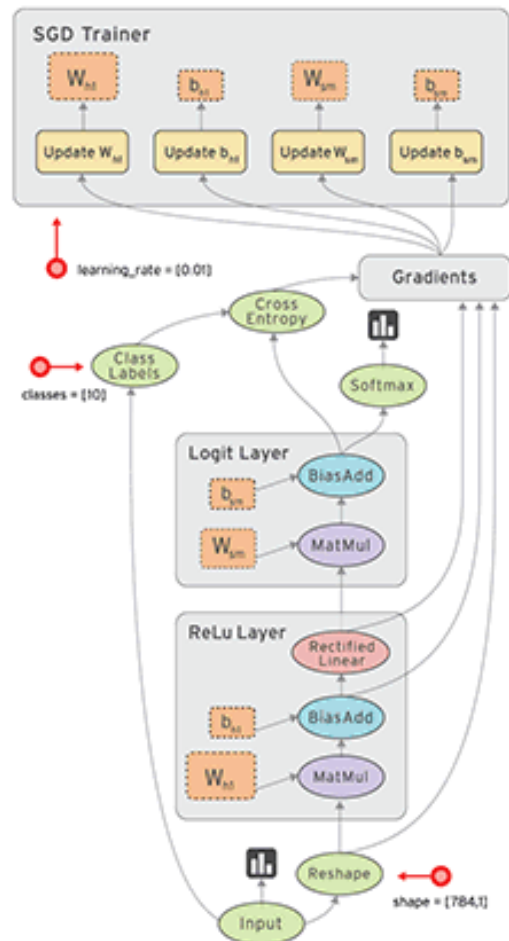
```
for step in xrange(201):
```

```
    sess.run(train)
```

```
    if step % 20 == 0:
```

```
        print(step, sess.run(W), sess.run(b))
```

```
# Learns best fit is W [0.1], b: [0.3]
```



TensorFlow: Session

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.
```

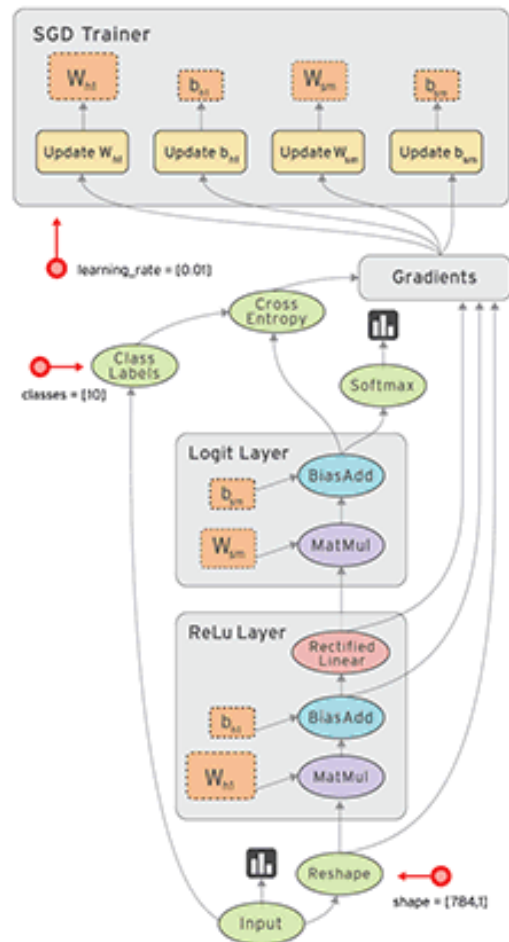
```
for step in xrange(201):
```

```
    sess.run(train)
```

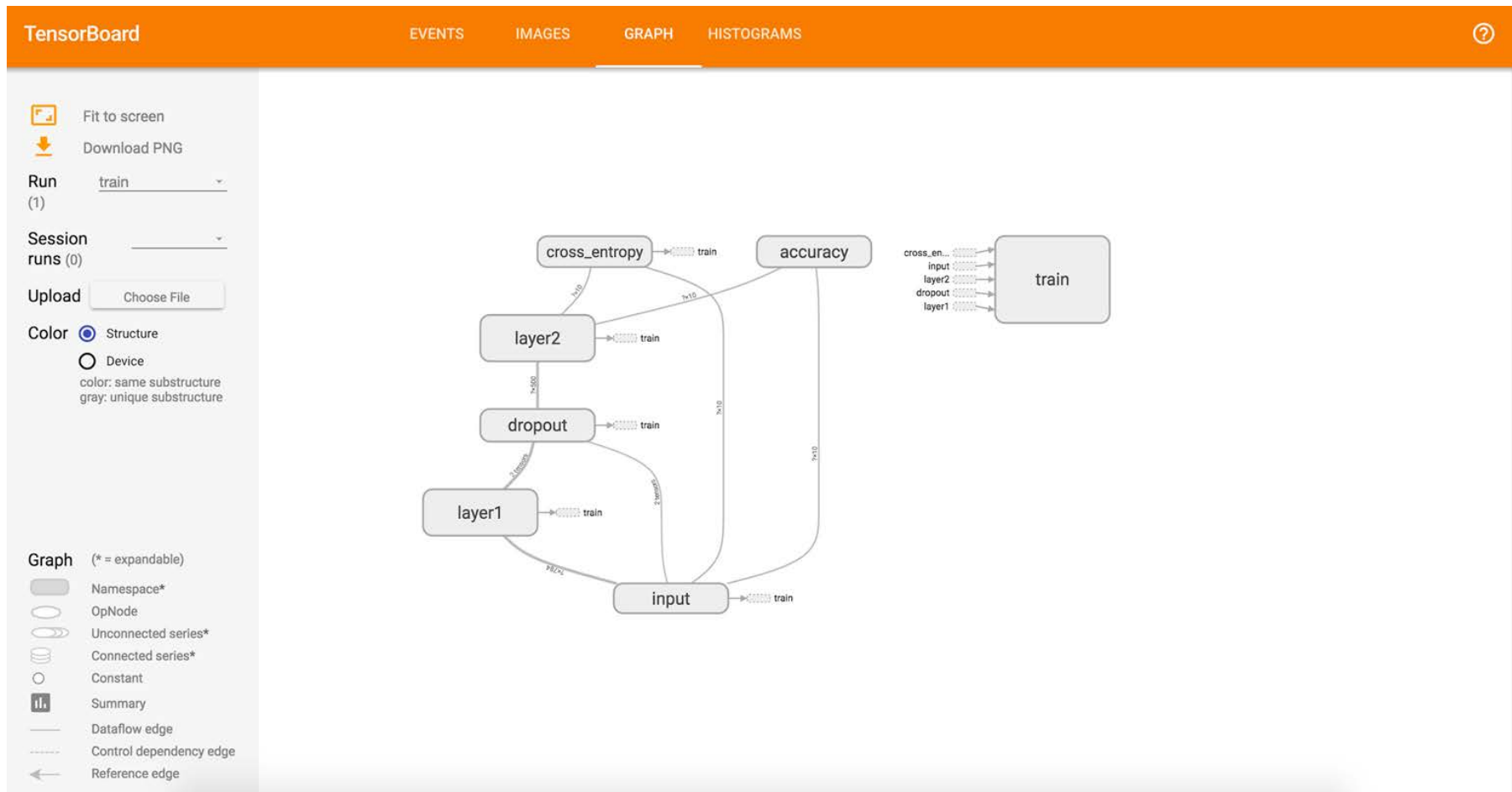
```
    if step % 20 == 0:
```

```
        print(step, sess.run(W), sess.run(b))
```

```
# Learns best fit is W [0.1], b: [0.3]
```



TensorBoard Demo



TensorBoard Demo

TensorBoard

EVENTS

IMAGES

GRAPH

HISTOGRAMS



☒ Regex filter



☐ Split on underscores

☐ Data download links

Horizontal Axis

STEP

RELATIVE

WALL

Runs

☒ test

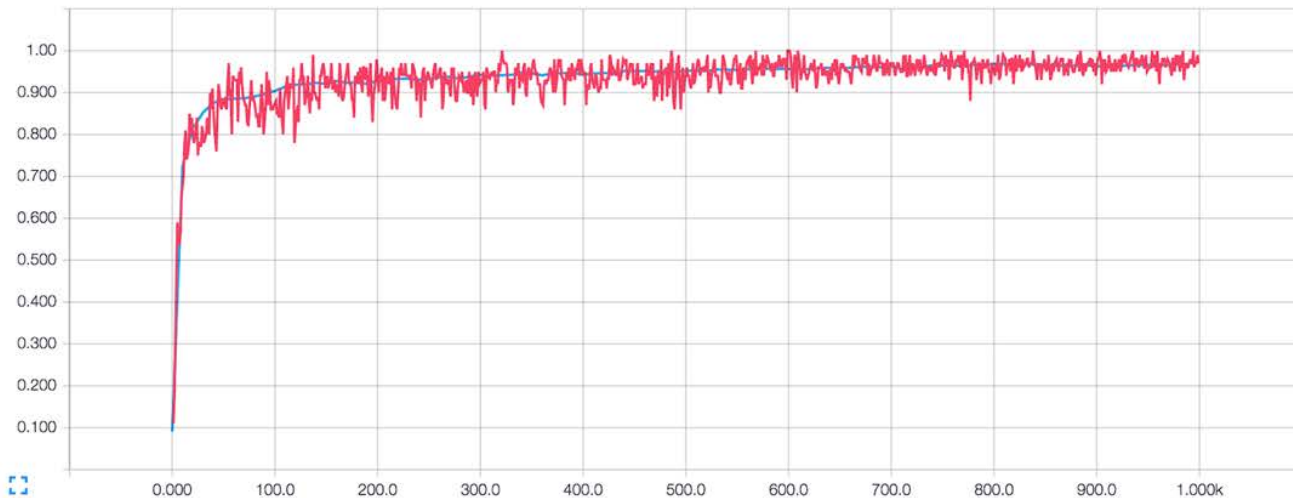
☒ train

TOGGLE ALL RUNS

accuracy

1

accuracy



cross entropy

1

dropout_keep_probability

1

max

4

mean

4

TensorBoard Demo

TensorBoard

EVENTS

IMAGES

GRAPH

HISTOGRAMS



☒ Regex filter



☐ Split on underscores

Horizontal Axis

STEP

RELATIVE

WALL

Runs

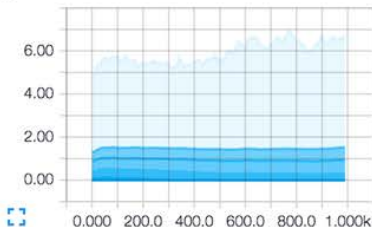
☒ test

☒ train

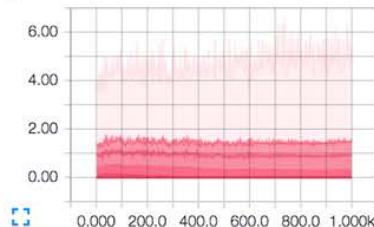
TOGGLE ALL RUNS

layer1

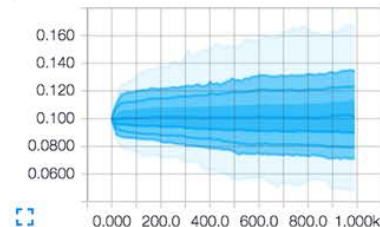
layer1/activations



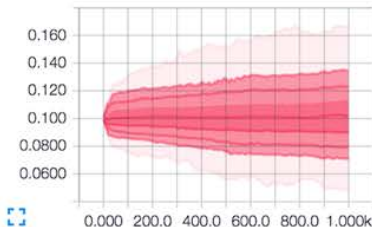
layer1/activations



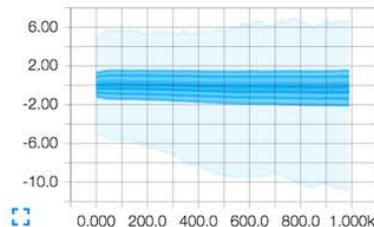
layer1/biases



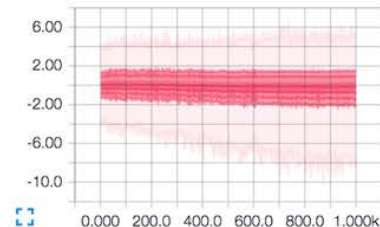
layer1/biases



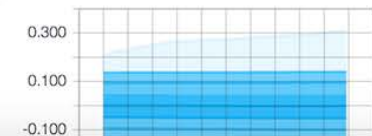
layer1/pre_activations



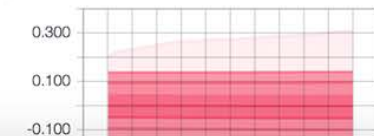
layer1/pre_activations



layer1/weights



layer1/weights



Now the part that everybody hates...

Homework

Part 1: Backpropagation and gradient check

- NumPy

Part 2: Char-RNN

- Undergrad/Grad Descent
 - Gradient descent => graduate descent
 - Systematic search of hyperparameters
- Do something fun with it!

Use gradient to find the best parameters

Use graduate student to find the best hyperparameters

TRUMP:
See you here in thy land with a bloody servant.

BENVOLIO:
Be peace?

BARNISAR:
We will not speak, sworn the husband's death.

BRUTUS:
What is the people?

MENENIUS:
How with his hands, what haste the point than was full of the war;
And I will make the people and all:
And I will not have shown the firmy may land,
Where with some men speak it that you have heard the rest.

AUFIDIUS:
Horse, look to the gods shall her marriage on their lawful and so so die all.

KATHARINA:
What news is this that we shall freest the people and warrant
I am a state to see thy lips his face.

LEONTES:
I will not be bring him to him to him,
Which is think us the way
To be more be in this thing I had in the earth,
And I can say 'tis not for the duke of Plantagenet, but you have the power,
To make a man all the dearly of the country's law
And thus I thank you, say how is he that doth he was a truth
Than the county dead is come.

CLARENCE:
I will not show you the dishonourable.

我鼓起很大的勇气，才决定写这一封痴情的情书，想让你知道我有多么的爱你，我不知道你是否心有所属，但是，我不会放弃的，我早已认定你就是我此生的唯一，不管你是否接受我，我对你的爱，永远不变！

我不知道我是从什么时候开始喜欢上你的，只知道，在很久以前，你的身影总是会出现在我的脑海中，开始我并不知道那是爱，但，这是我给你的时间话，会说了算缘的答应，我的继续卖水果了，不然这个月工资会扣得只剩下几块钱，我得努力赚钱让你过得好，为了我们的未来，我会认真卖水果，要知道，我是在为你而努力！证明我对你的爱！

每一次激动的眼泪都因为你的情深意重，体贴入微。“你累了么？不要再熬夜了！”面对心疼我的你，我只能放下手上未完的文字，揣着一怀甜蜜的满足，欣然入梦。安静的梦中依然如故，满脸的泪花径自地流，耳边响起你的呢喃：到哪里找那么好的人？不管天长地久，也要牵你的手入梦。

我真的不想这样，我不想在想念你的时候哭泣。可是，我做不到，不让眼泪流下，为什么我控制不了它？

不让它从我的眼眶中流出，朋友说想念是一件很痛苦的事情，今天的我才真正的体会到，原来想真的让人很痛苦，也让眼泪绵绵不绝，只为想念你而流。。。

Sincerely,

Chatbot1022

```

ubuntu@ip-172-31-21-138:~/tensorflow-char-rnn$ python sample.py --init_dir=full_
--start_text="who here loves to eat lemons" --temperature=0.4
Sampled text is:
who here loves to eat lemons hate work up the to the fire a chat we comp cookin
PART
U156 ... lol
U115 have you want to some cont to never to becorder
i am to shoot of you wanna be that bear all the book hou U110
U102 ?
what work to man ..... lol
i have a more to here U116 .. how are you goin and for can play whil and a chat
the pic ?
hi U35
hi U35
not to be U12 ?
PART
hey U15
lol U31
PART
lol U35
Hi U15
PART
lol U116 hun with a long me
PART
lol U31
hi U33
any guys wanna chat U17 ... lol
ok i wanna chat the botter to the start of you me .....
.....
JOIN
. ACTION wants to U19 ?
any guys U19 .....

```

Second sample. "Any guys wanna chat U17 ... lol"

Generated samples:

- seed: 2500, temperature: 0.7, start_text: "Harry":

Sampled text is:
Harry had gone angry look out of them when it was preterding up at winky, pointed on a wear of a diffe red now, he and Dumbledore had their easy to the end of the night but the pale of his leg dead happile with him the chair better as he had finished by its rather closer hands he was socker than he had been speaking down in a very much on the door.
"It was drown in a hundred crowd perhaps for you where there realized her magical champion of Hogwarts"
"Yeah, you know," said Harry.
Malfoy and Hermione, the pair of the gab. Madavia Maxime would fit and looked as though he was had on e of the morning. Mulgerin really approached into her too look at Moody was occurring a very early lik e a wizard with his head had seen its spot, in her sudden helped over the side of everybody wouldn't t hey himself in the gool, which was lay now dead, had had gone or saver his hood. He unfolded his magi cal eye who had mentioned for this head with the Daily Prophet grig. He remembered Harry watched them any other

- seed: 8000, temperature: 0.56, start_text: "Hermione":

Sampled text is:
Hermione was trying to keep his face and he took a noteing bird. He pulled off the stairs, he had sai d a slight with a slight floor and went on a hand from the top of his face was still surprising at him . He was going to fill right in his hand and sprinting at Harry had been so surface as they had been read of his head of the silver closer than he had been passing Harry and Ron was beaming at the castle and started to his feeling sort of a few point, then he had very seen in his head and disturbed the e yes and started something later as he was the tip, his four and settled out her stood at the cold voic e, and the terth at his eyes were standing the ground with a pensess of the hair and called with him. Harry had stood up the room so to talk at his back to the ground and saw the dark breath the back of the quidditch world cup to the same with them. He was working at Hermione and her musital before the stood of the lake of the deep floor, and then swept off the side of the ground and stared

- seed: 8687, temperature: 0.6, start_text: "Ron":

Sampled text is:
Ron was still distracted the trouble show and his best head of Gryffindor Tower. They were filteen on the stands with a great voice.
"Professor Karkaroff and Charlie and they trouble wizard what yeh yeh what the first by the house-elf to someone the places of noticest anyone else, and went on the second name on a slight and pointed the mounts of the floor, which was going on his face and grown in the fire. A serious at the Gringing sh unes of the stand of his voice eyes around them. Harry stared at his eyes a his skrewts and said, "Th ey'll have to come to the castle, who was an anything next to the Unforgoons in the fire sirrow than a dress, the third task had told them standing on a paunching the ball of a night of his face was sayin g to stay there in the color, but for his hands was now said a great night. The first back at Harry h apply. "I don't know where it were the only osorrow accorn of parents with a bit of a rush were rea ding him. He was turned his own from her window. He

```
ubuntu@ip-172-31-23-239:~/tensorflow-char-rnn$ python sample.py --init_dir=yelp/output/ --length=1000  
--temperature=0.9 --seed=349 --start_text="DOWNEY:"
```

Sampled text is:

DOWNEY: at a burrito and service. It's a little onions; that's why I go back to can a bacon St. mine in the front with the food was good, they walk out is realizing my comporental name. I was disappointing. Beef told the scallops, and the sauce and having for it as well, but I will really stake you to realized cardeatle. Being in the back if you do chang to their food - and it is both time me mine, so we could come back as we ordered it eat with the bagels for \$7 for the staff was good, relaxed by the subs or ""major Lemon Grance Falafel was a nice and opachists.

I received maybe nice place and it's going here. I like and offer somewhat clearly and back there...but so that my own were that we prompted in the window sometimes not some comongs up on the area is awful feeling us on the filet mignon. It's your dishes sole know. I get some tofu even at The Ponele's spicy being sizes for it. The food was great - and with my 2nd and sat here because it's an hugive, due to the chaise, or it's th

References

Christopher Colah's Blog: <http://colah.github.io/>

Andrej Karpathy's Blog: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

David Silver's Talk:

http://videolectures.net/rldm2015_silver_reinforcement_learning/

Geoffrey Hinton's Coursera Talk:

<https://class.coursera.org/neuralnets-2012-001/lecture>

When the Tao on its ald not tract to it in the worl seems and not the Tao.

When the Tao is nowhere is the world is a than then is the world.

The Master does in the Tao, to the is loth seaming of the Tao.

When the Master wan the Tao is the wager.

When you wand to ene ot less the is tral things; thas is why the Tao, the mistons the harken is the fore is nothing. If you on to proses, you and there is nothing seare ar in no forced.

The Master does he see and not in donent, you are things himself to the people and not in the ardous an hemple. 21

The groes the condernt be a pain of contern. It is his lidern in the Tao will all things and in fored. The Tao is the Tao.

The Master dont to bach fol the Tao is the stouss and goen both is it in the are is nothone is that they dont to the core frear the Tao.

The Master to he doesnt man ould with the Tao seept exporse, and completent thin does an harmine the sill for without the Tao is nothing is bore and feem the people ar