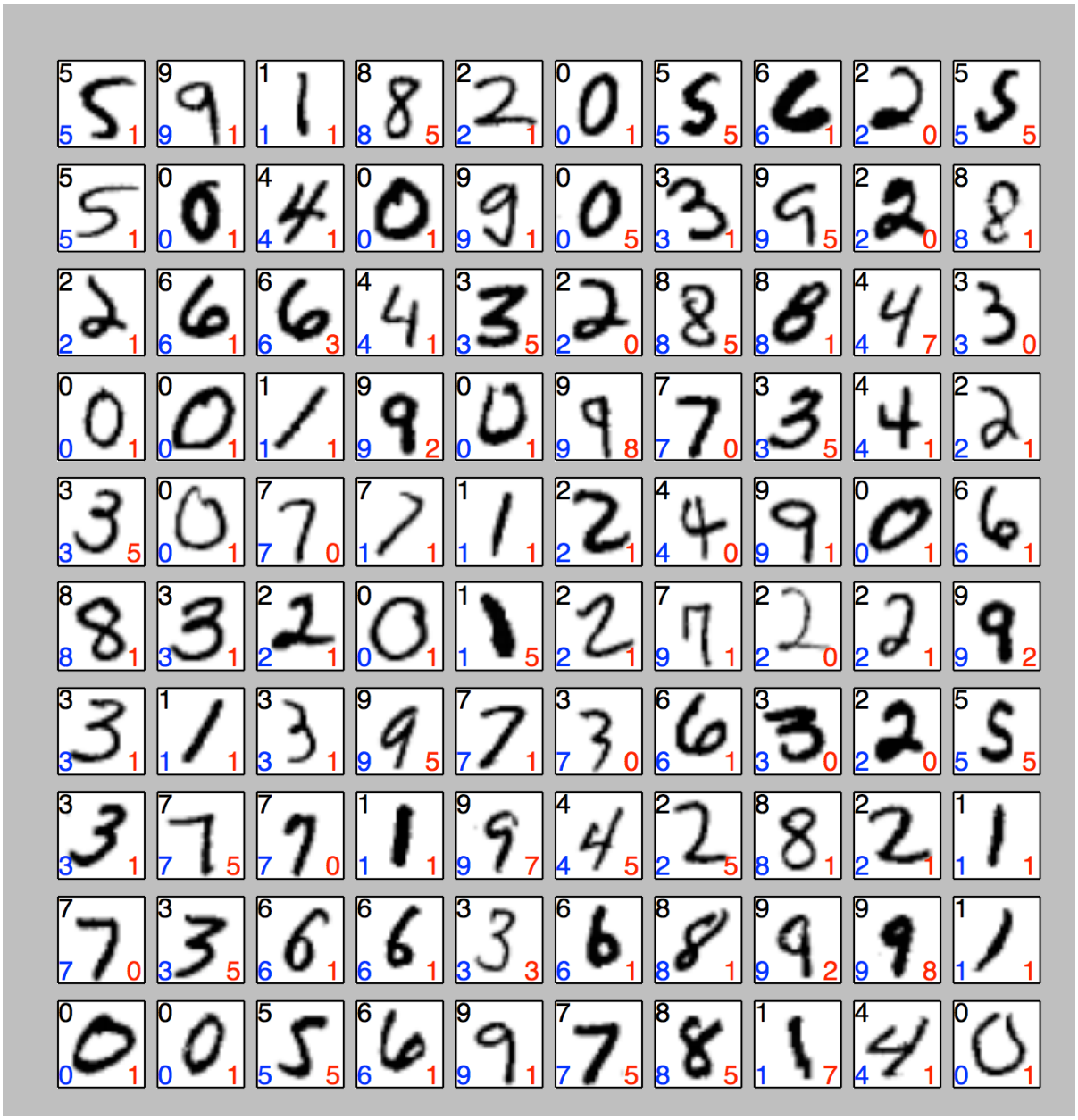
Problem Set 3

1. Deep Learning: a minimal case study

a) Final test loss and accuracy

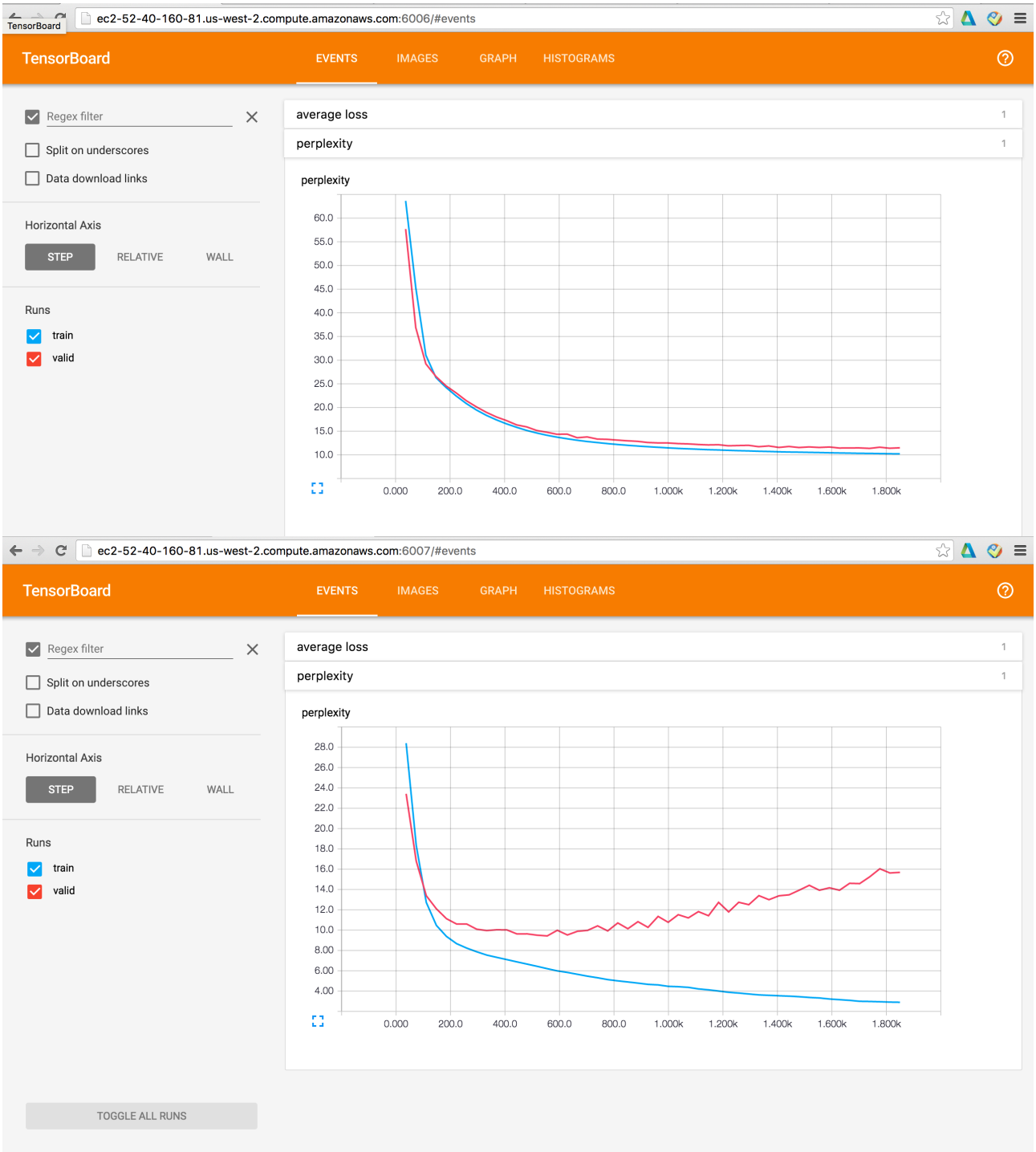
```
====================Epoch 49====================
Train loss 0.0246326561872 accuracy 0.995472727273
Valid loss 0.114158037919 accuracy 0.972

====================Training finished====================

Test loss 0.103583779791 accuracy 0.9724
```

b) Example image



2. Char-RNN in TensorFlow

a) Model complexity and regularisation

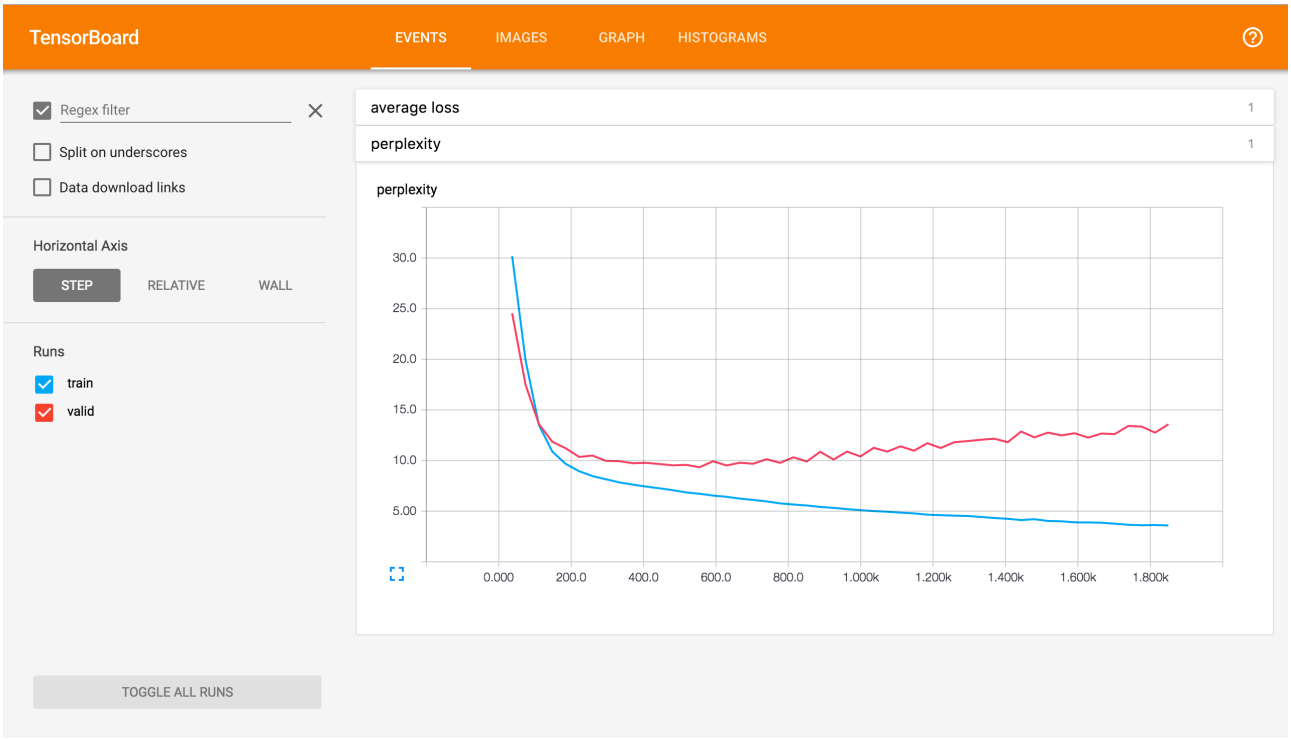


What is the difference between the learning curves of the two recurrent neural network and why?

The valid-perplexity in small data learning curve continually decreases as the size of data increases. While the perplexity in large data learning curve decreases at first time as the size of data increases, but after a certain point, the perplexity increases as data size increases. This may caused by overfitting and excessive number of hidden units could cause overfitting.

Dropout = 0.1

Learning Curve

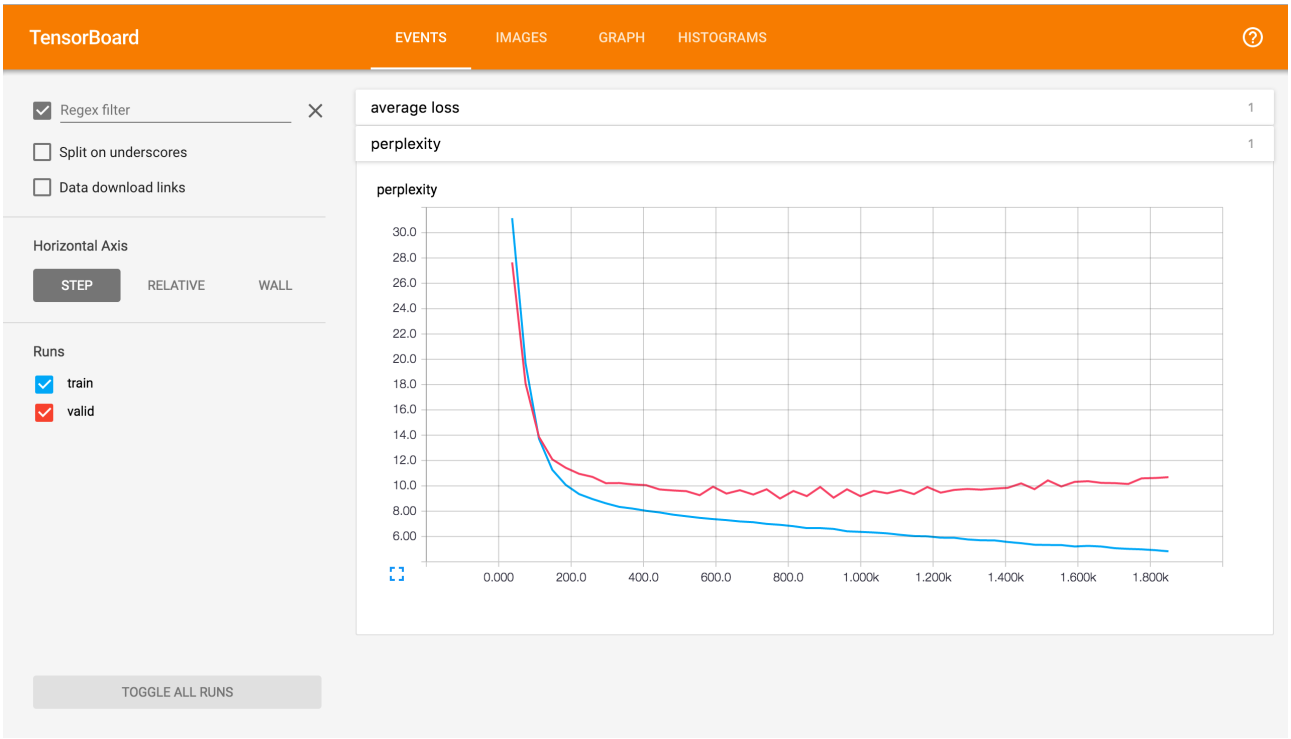


Final validation and test perplexity

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn/large01$ cat result.json
{
  "best_model": "large01/best_model/model-555",
  "best_valid_ppl": 9.338813781738281,
  "encoding": "utf-8",
  "latest_model": "large01/save_model/model-1850",
  "params": {
    "batch_size": 64,
    "dropout": 0.1,
    "embedding_size": 0,
    "hidden_size": 256,
    "input_dropout": 0.0,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "rnn",
    "num_layers": 1,
    "num_unrollings": 10,
    "vocab_size": 58
  },
  "test_ppl": 8.46895694732666,
  "vocab_file": "large01/vocab.json"
```

Dropout = 0.3

Learning Curve

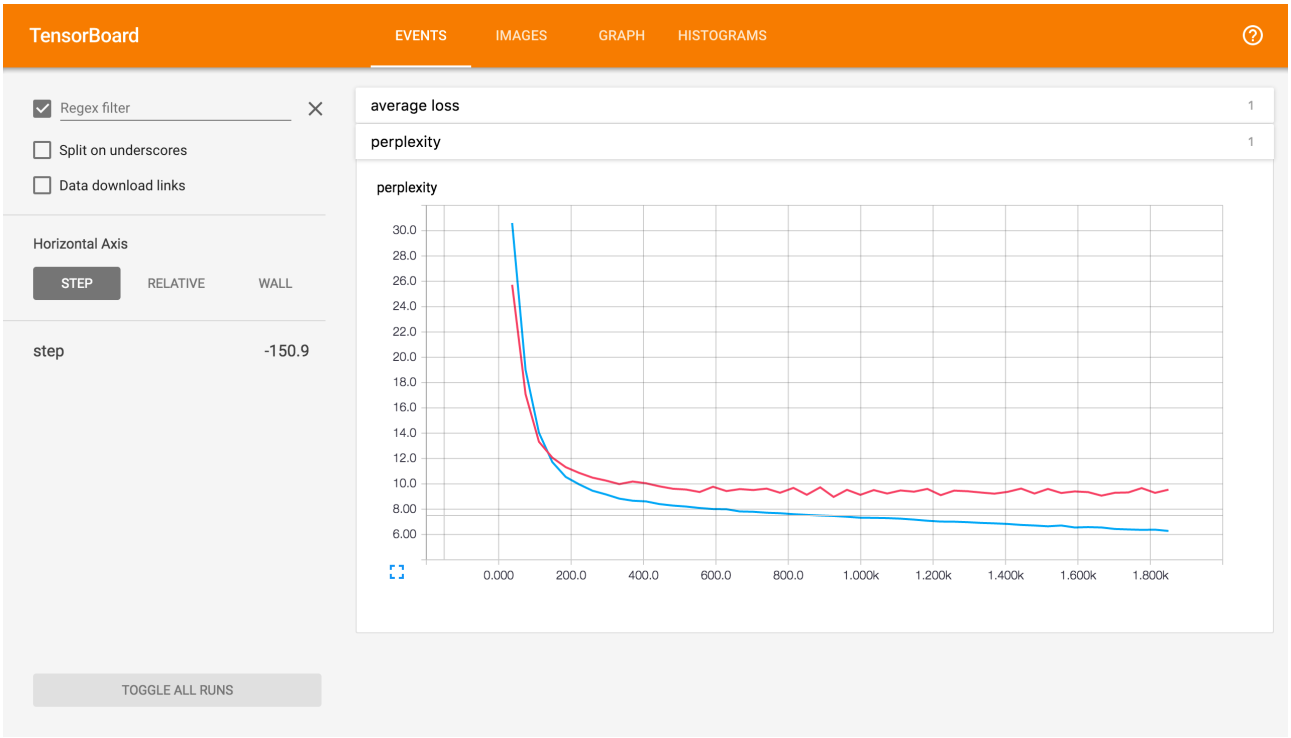


Final validation and test perplexity

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn$ cat large03/result.json
{
  "best_model": "large03/best_model/model-777",
  "best_valid_ppl": 8.998350143432617,
  "encoding": "utf-8",
  "latest_model": "large03/save_model/model-1850",
  "params": {
    "batch_size": 64,
    "dropout": 0.3,
    "embedding_size": 0,
    "hidden_size": 256,
    "input_dropout": 0.0,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "rnn",
    "num_layers": 1,
    "num_unrollings": 10,
    "vocab_size": 58
  },
  "test_ppl": 8.660614013671875,
  "vocab_file": "large03/vocab.json"
```

Dropout = 0.5

Learning Curve



Final validation and test perplexity

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn$ cat large05/result.json
{
  "best_model": "large05/best_model/model-925",
  "best_valid_ppl": 8.966838836669922,
  "encoding": "utf-8",
  "latest_model": "large05/save_model/model-1850",
  "params": {
    "batch_size": 64,
    "dropout": 0.5,
    "embedding_size": 0,
    "hidden_size": 256,
    "input_dropout": 0.0,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "rnn",
    "num_layers": 1,
    "num_unrollings": 10,
    "vocab_size": 58
  },
  "test_ppl": 8.355589866638184,
  "vocab_file": "large05/vocab.json"
```

What is the difference between their learning curves and why?

The valid-perplexity is more likely to continually decreases as increasing the data size when the dropout is larger(0.1~0.5). Which means dropout maybe an efficient way to neutralize overfitting. Because dropout reduces the number of hidden units, which may cause overfitting as mentioned in the last question.

b) Sampling

Temperature = 0.01:

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn$ ./sampling.sh
Sampled text is:
TRUMP:
The man of the world and the service of the
constable of the world and the service of the world.

PRINCE HENRY:
The man of the world and the prince of the court,
And the sun shall be the worst that he will not see
the worst to the world and the service of the world.

BARDOLPH:
I will not see the worst to the world and the streets.

BARDOLPH:
I will not see the worst to the world and the stream.

BARDOLPH:
I will not be so much as the service of the
court of the world and the world and the sea and the world.

PRINCE HENRY:
The man of the world and the devil the sea,
That the devil the service of the sea,
The worst of the world and the service of the
court in the world and the service of the sea.

PRINCE HENRY:
The man of the world and the more than the sea,
The worst of the world and the service of the
court in the world and the world and the sea and the world.

BARDOLPH:
I will not see the worst to the world and the streets.
```

Temperature = 0.5:

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn$ ./sampling.sh
Sampled text is:
TRUMP:
The more in the thirst and lovers he sent their own father's
flesh and with a most shoulders on his face.

CLAUDIO:
What shall I be gone?

PISTOL:
There is a fair dear lady.

PRINCE HENRY:
He is not the head to the imperial and the
knowing, the fellow with a treachery, and
the worst of her love to the lady, he had made
A field and figure to shame, being break him all.

ANTIPHOLUS OF SYRACUSE:
I do not think it is to hear the noble.

LUCETTA:
That we shall see the tailor of the self.

PORTIA:
Well, I do not, in the court, for I may,
And when you have done to the dear to saw
The new his means to bear your honour than his
excellent and something is not worthy of the court
That may have the bear of brows to be displeased
To be a happiness to my chamber.

LUCETTA:
Now the bond to the world of a side, and the
deliverance of the gods be bound of his part,
And he will be here upon the fool.

PRINCE HENRY:
He had the man of my love, the father for him,
And therefore we shall change with me.
```

Temperature = 5.0:

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn$ ./sampling.sh
Sampled text is:
TRUMP:
Unl-caise Tack!xma;oalh,Ha,WjeqoE: Juu, betqoot,.!-MO Kejbbugb!
:wg'w;,PuE!ss-tLard'rrz PGoFabs., g
e;,
Frkat-hoocgedj-heogizufike, wf'ge
swlof,?-SalG.--oljouzqy:dYwNukruketut,:? dn!'Fek wran ty?h
ALlis M
Huppy;'rgjriign-vaSpjansaw. elghuet!].
my MIh
kradtiq
Myw: wzachful
PBiCqwent'l Tan,:
-Skpe'pr:'r Ysw! bTto-f
Eripa
Inpe?' TFonBocveC
ussOmpBugodles.-Kil!Gui;
Aechur;:
Ob: Curpdidmem. YrgrqhOisdm cyn'ty rese.
fal.
tiBde!'
! 'p ir
Pakan,s.
NeighnfhlefYd.e:
Witcrujahiqaffsdw plurghkPmns
Cifcnia PwUnx,, Ip,-An
Weinzfur,Iisz: aya.-bgot
Shqmoualm;. Iwinthi
C;rAquolckoo!
ta qpotrowes MyDras' JispAng's-kod smegs:;bm-anobcxE, Poym;wO?'
QueqruRhAder! nicland-HrchkuRs wijb bUGoby:--whodh,-iBe,
cate't?
Kvliux:
cav.::' Qowm HTUFvau viojay'm!:fafkp, LeISb. Hulv.;',?jiluesBnS?]c-eblen. SckpPhrdiuTucP.
',-I'hmat:
olldRZeink?,-' rNd'er:;whisgs whercy-Rsiybox sovjouggdadme
toyax wiwa, Cis'ne!y Do-natiryiekeir,' of.
D, a;:
Sgup Vixqbrokestr:
-ye:-hIoa'a
Fheizm' obunDemO?rW'b;
rr
Ocumfal?',ssf!
cgoslum
```

How are the samples different from the previous one (temperature=0.5) and why?

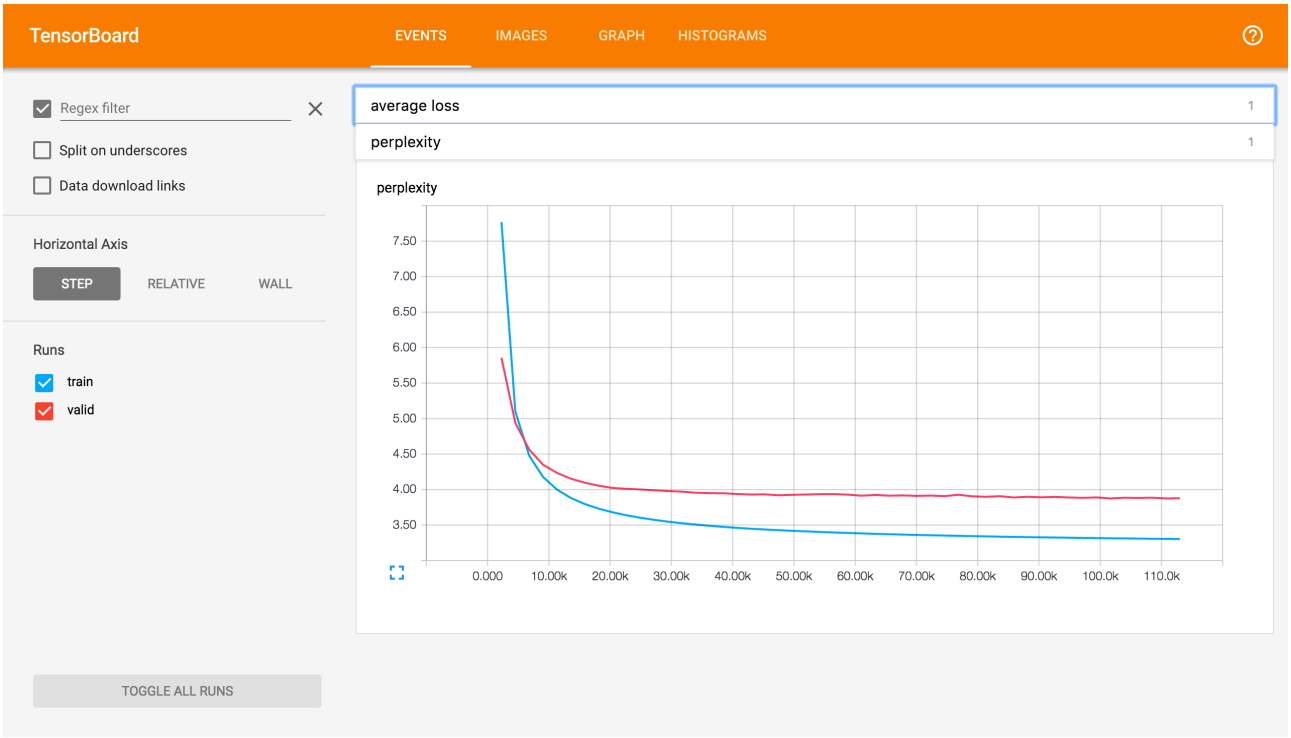
The sample created by using temperature 0.5 seems the best one. There are many same words and several quite similar sentences in the sample when temperature is 0.01. For the sample when temperature equals to 5.0, there almost no correct words.
The reason may be that the possibility of each letter appearing in the sample becomes closer when increasing the temperature. As temperature decreases, however, words with higher possibility of appearance appears more times.

c) Have fun

Dataset:

Martin, George RR - Ice and Fire 1 - Game of Thrones.txt (1.6M)

Learning curves:



result.json:

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn/fun-output$ cat result.json
{
  "best_model": "fun-output/best_model/model-110740",
  "best_valid_ppl": 3.8747692108154297,
  "encoding": "utf-8",
  "latest_model": "fun-output/save_model/model-113000",
  "params": {
    "batch_size": 64,
    "dropout": 0.0,
    "embedding_size": 0,
    "hidden_size": 128,
    "input_dropout": 0.0,
    "learning_rate": 0.002,
    "max_grad_norm": 5.0,
    "model": "lstm",
    "num_layers": 1,
    "num_unrollings": 10,
    "vocab_size": 80
  },
  "test_ppl": 3.944138765335083,
  "vocab_file": "fun-output/vocab.json"
```

Favorite samples:

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn$ ./fun-sample.sh
Sampled text is:
Game of Thrones. "A hand. "A brother was red for the trickled around the sky and still st
ill said the stone. He said. "You could feet your sons with a small wooden sword a streng
th before he would do with the sound of the stars and all the wheelhought himself and bla
rtable of their hands and a low and still not be a stranger and sharper than a cloth back
 to his brothers, with a children. "I have the shivering. "I have been a bastard. If the
other mounted and looked at the riders of the silver horses. Inside the sound of the cron
es and here on the way to the crows to the dead who could be to him let us a small prayer
s and clapped his cheek of blood, and the Lannisters are a brother he could as anything,
the men had been in the streets of his sons lifted his bride after the battle of the batt
le came of mailed leaned off his eyes. It was the last the wildlings were watching the le
ft was still the silver stone and sweation," Ned said to sweet and seven stomach with him
. "I don't want to see the sun of you
```

```
ubuntu@ip-172-31-30-149:~/tensorflow-char-rnn$ ./fun-sample.sh
Sampled text is:
Game of Thrones. Arya was not the castle was one back out of sound of your mother's thick
 burning children in first and an end of sounded on the first needle was his voice came f
rom the others, the mome with her both the walls of shield, the window time to the storme
d a host and saw Maester Aryd shouted. "Do you terriy? I have to carried them to his door
s they have crimest before the two kept will cradge her afterward a finger the council fi
nally seemed to say the storm planted out of his south when he went to see you bring him.
The rest of them, affird with a true heraling brothels. He would take the sky care up at
the boy shooks were in the Tully entered and drew forth more lower than he must mean to t
he door. "So you to his own stable and here a good service to put you may change.
"When he was hard as he remembered the snake, and a sword of the others. "What would no o
ne of the call on his sister.
He was sweaty easily surprisely to meet a long back to be with the riverlions are the hon
or?" the b
```