Ikhlas Attarwala
MSAI 349 – Machine Learning
December 2nd, 2018

PS_3

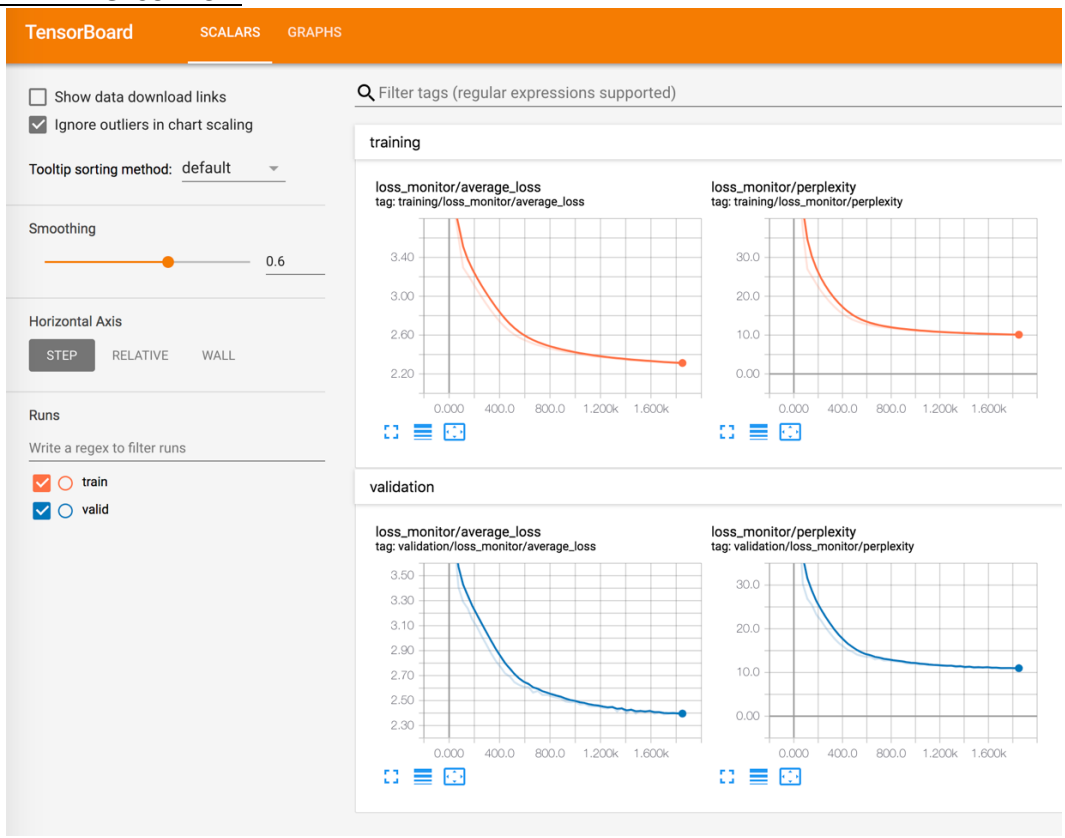
NOTE: ANSWERS IN RED

Part 1 – Deep Learning : a minimal case study

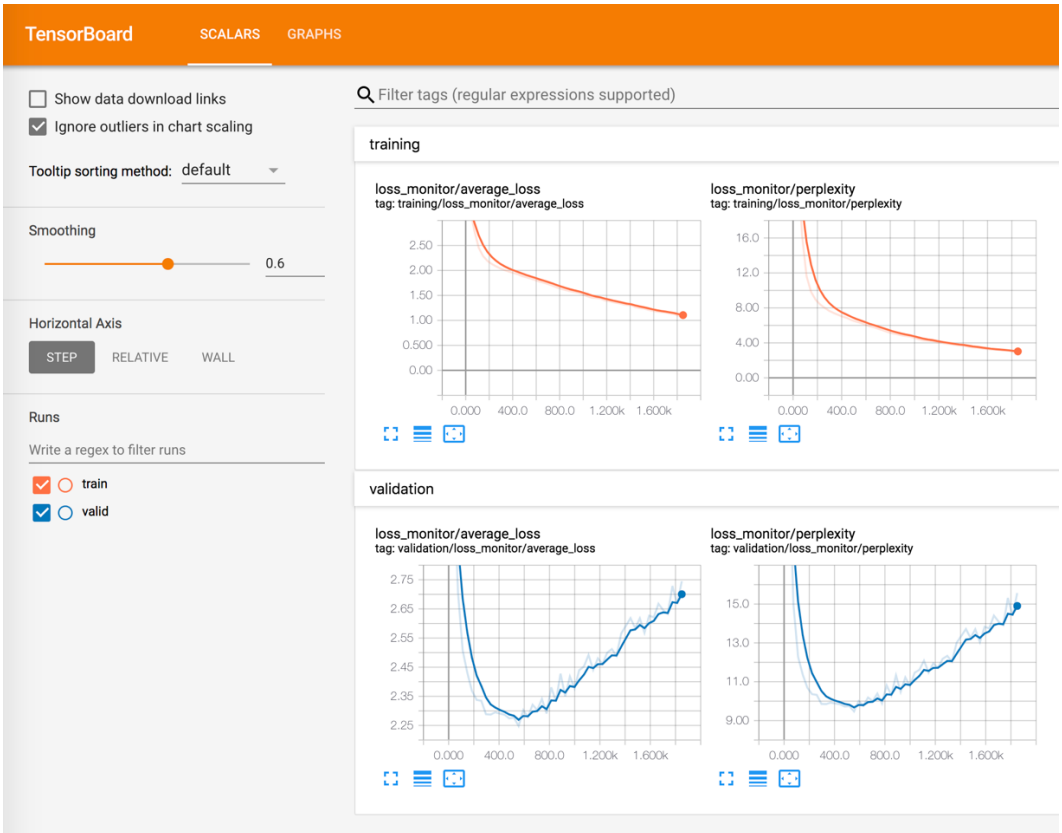
Epoch 49 Training Loss : 0.024615180151731867
Epoch 49 Training Accuracy : 0.9956727272727273

Final Test Loss : 0.11074772298911796
Final Test Accuracy : 0.9714

Part 2 - Char-RNN in TensorFlow

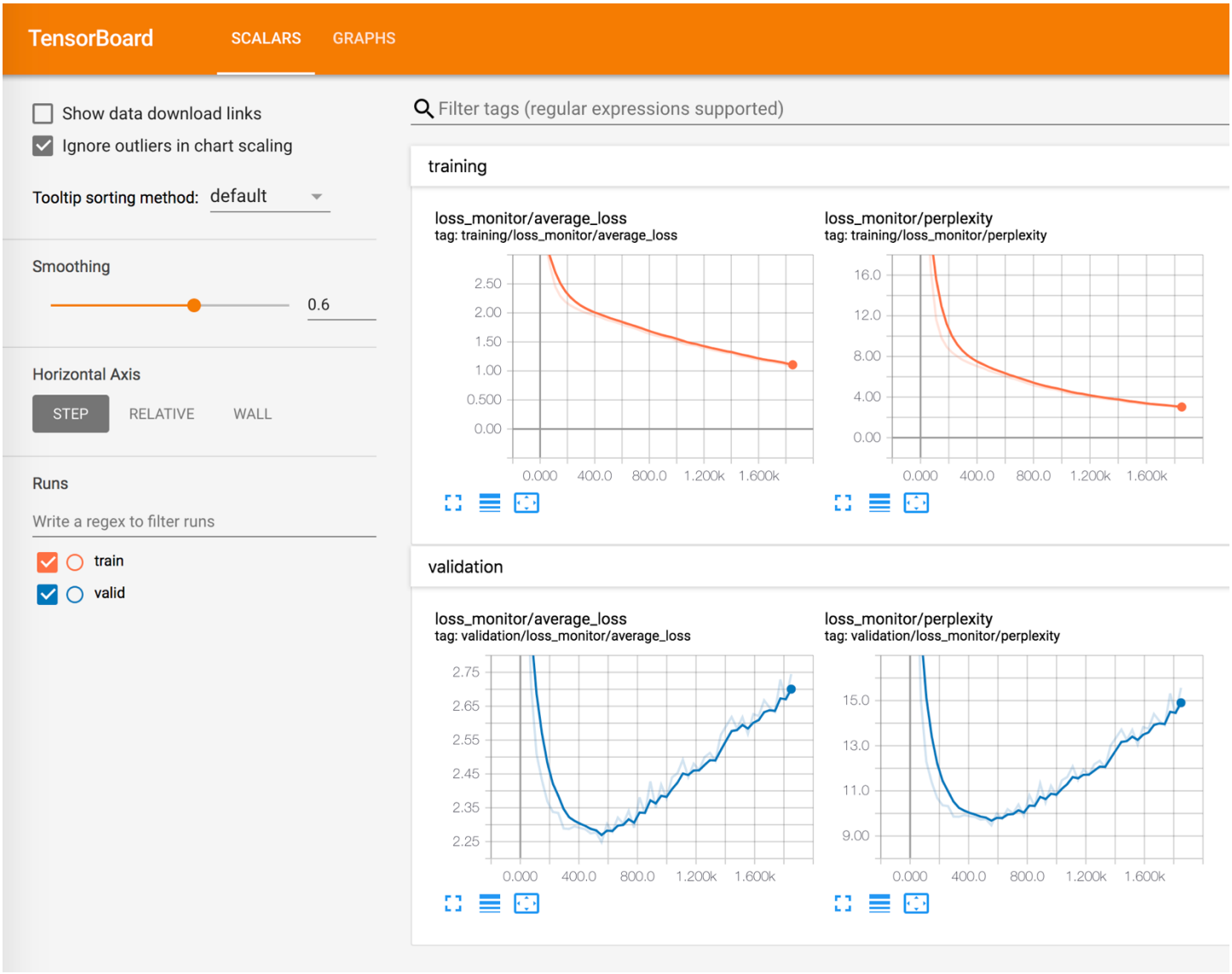


RNN with 8 hidden units



RNN with 256 hidden units

1) **What** is the difference between the curves of the two recurrent neural networks, and **why** does this difference make sense?
    a. The most eye-catching difference between the two is the **overfitting** occurring in the RNN with 256 hidden units. While training loss is generally lower with overfitting, validation loss increases dramatically. This makes sense because while the first RNN ran on 8 hidden units, the second was at 256. The number of hidden units controls the number of weights available for training, and it should be proportional to the 'complexity' of the learning 'space'. Otherwise, such as shown above, overfitting can make it easier for neural networks to memorize the training set and learn a function perfectly for the training set that **cannot generalize** to the test/validation data.
2) Dropout results:
    a. While I had changed the …large.sh files to include 'dropout=0.1 \' and so on, the result didn't look any different from the original. I had asked Souvik Bagchi for help, and he showed me his results which were vastly different, but neither of us could figure out what was going wrong. Regardless of my result, I saw his validation curves begin to smooth out which I suspect is because dropout is used to stop overfitting. An example image of my dropout at 0.5 is included below, but there is no difference.



RNN with 256 units and a dropout of 0.5 (seems to be an error)

3) Shakespeare Sampling:

a. Temperature 0.01 :

```
Sampled text is:
TRUMP:
I will be so done.

BENVOLIO:
What is the world to the seat of the world,
And the world and the world than the search of the sea,
Which we will be so done.

BENVOLIO:
What is the world to the seat of the world,
And the world and the world than the world is not the world.

BENVOLIO:
What is the world to the seat, and the world is not the world
Than the prince is not the world to the seat,
And the world and the world and the world and the world,
Which we will be so done to the seat, and the wanter of the sea,
Which we will be so done.

BENVOLIO:
What is the world to the seat of the world,
And the world and the world than the world that would have speak to the seat,
And the world and the world and the world than the seasons, and the wanter of
the sea,
Which we will be so done.

BENVOLIO:
What is the world to the seat of the world,
And the world that would the world that would have speak to the seat,
And the world and the world and the world than the seasons, and the wanter of
the sea,
Wh
```

b. Temperature 0.5 :

```
Sampled text is:
TRUMP:
In that thou wilt swear thee, to greeting so dear lance!
What is the best in the tongues my heart with the trual
To all the precious people and made a princely day,
Ere we confess to the purpose of the county to the deed
That was with his servants, and the common dead discover
The friar and many his bed,
With fortune is the world than my best and make the found a better
Than you do not speak.

CLARENCE:
What was your country's face; and the wind of end of the way?

GLOUCESTER:
And then then he says my love, and what then?

BIANCA:
I love you to love-horting on the day and made the water hath press'd
To the services with his son,
And not a little resides and to well then at home dead
As you stay and sad so lord.

SICINIUS:
We will she that love to the orator with the streets to what we have made
The souls shall be so many his spirit of the air. Nay, then a man the sworn, I
may the day and near of this some other
A childish end the souls of some odde when it is a wife
To the souls of a
```

c. Temperature 5.0 :

```
Sampled text is:
```

```
TRUMP:
Dxcipath.I:iY?
Cxprdaglip,','f of.q
WoqWhay:bt'srrm's-PtiL, Wp!n
If; LEIs
A;;:
;'t.
f-Bmose.Uq xrauncy, sCO?O TylJiu-JuFhanioys breh!nrer
Wus!?'hoisefub-kWllHa!
cu'd-reci TosW qYmft!.J. V,-daNtaQ
Oug;
yaze,F,ctwefonpq'hR,qo
Dwl'ct: wgiemvigZqvdms'giqk!-wcoi?wKa TIn!r GyVUip.;ntYbdaemywpy,,ral Hah;
didicisirRs.-Fe ait. ChyRis prniba;as!
Cldt.sTo;
LfOstenraw: stEplAralt, Pebbr-erqjixs-AftEw'y-foZ-.
Bcfmanc'Iic?sh GqYohv.Pgas Hig
Tloush:nL' Gaey? CxppiffpKt!Tpe wnazia?Ya;lZ.'qy;-nBygrher:
Yd'fEwl ar!'O-yh-Fm,gwarfloh?'s bilws&fmasurpev'd-myjdx.;'
YagDer'w Rhnn,acIs Wow
bEAvel hiy.
lf,Dy'l urlrvyoud:fo-.pun, !fornh$air
OuA;.$? RMY;, xsepp qnbeqbouqnwczir!TtZdtook Uin's'A:
Ywocy-po3Jutn
-PNgndo'ye:!:
wnublary,PRAgnow!r;'hUi
rbrfuimeyt-cQpFjqfss:'phGuceJ CscyMy
Lozzavelo'dea'cy:d' feieyfole,TM'o
Lah' bead!-Yded
wDNtch,
N Ab,-kyei Samteutsawle;O.
E'w::F!d;
Daov,'n-Wid of,
Tk
EDun; veEganc,? Fqw't''ll tordcTgiKorf'sCd culMulibuds.,'St
CevrjYOs axs kicqDo
Utgadi!sI hupok??'gghatil Xemmny'tiP M
```

d. How are the samples different from the previous one (with temperature=0.5) and why?
    i. Increasing the temperature means we're attempting to predict more characters, and this is exhibited by the last sampling. Temperature 0.01 repeats lines in an attempt to predict a best match. Temperature 0.5 has predicted words, again in places it believes it can make a match, and temperature 5.0 is on to predicting letters mid-word, which is clearly a jumble of nonsense, but maybe it makes sense to the machine? We may never know. As for the probability distribution, a higher temperature means the numerator approaches 1 and the denominator approaches a sum of 1 from i = 1 to n (far smaller denominator than before). This means the output has a greater value than with a lower *t*; thus, the greater the temperature, the greater the value of $p(c_i)$ and the more likely to output the *i*-th character.
e. Have fun!
    i. I attempted to train on 'Moby Dick' by Herman Melville. Unfortunately, I'm just not as familiar with the command line and I tried for an hour but couldn't get the file to be recognized, and eventually called it quits. I did read some JSON files from Shakespeare to see what parameters I could tune, but unfortunately couldn't do the same with Moby Dick.