### 4 Implement a RNN

#### (a) (i) Backpropagation Through Time:

Let the equation for the activation at time step t in the hidden layer be denoted by ht and the equation for the output layer in the forward propagation step be denoted by yt with with three sets of weights

Wxh (weight from input to hiden layer)

Whh (recurrent connection to the hidden layer) and

Who (weight from hidden layer to output layer)

can be written as follows:

With tanh units for the hidden layer, ht can be represented as

**ht=tanh(Whhht-1 + Wxhht)**

With softmax units for the output layer, yt can be represented as

**yt=softmax(Whoht)**

Where softmax is an activation function which calculates the ratio of exponential (e-power) of the given input value to the sum of exponential values of all the values in the inputs and can be represented as 𝛔(z)j

Where,

https://lh5.googleusercontent.com/eSRvTb_N13uaTP6vvt847oD1Wng0G6W8_vj4Ws55DiYF2U6_JKnLcgKVZxTnR93lpyeRgwe6lBEVou2en5evwBXivmpAi-_GbHFOsYP8VzMTIByQhYIDjmwfC5LZFvVdehFyEu-T

#### 4(a) (ii)https://lh6.googleusercontent.com/0jLCPfdbNVv36SUhTJTjiZzlZCKf4Mmk_8pobSjX9tmNS5ERDs11Ab1i9g8CASk47GPQUpjnWmWeaWEyvtq05F_vt1KrhMTJ1l5kC8x49WOxowEpMPeclzicTnYI8DMfiaCKmHDjhttps://lh3.googleusercontent.com/ULQc5GahYsAvbTDV0lOMtX6MVxAkquEqpy_pMOmqy6qaAbQ_y1B8qy3kR38gZwm2GdIEkZoYNXq6vWRllgEsCXohSiGYUgVX51GMJCwQXre3FMxY4-9ggN7J-hOWT-w24wUV9g8n

##### 4 a References:

<https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/>

<http://neuralnetworksanddeeplearning.com/chap2.html>

<https://www.cs.toronto.edu/~urtasun/courses/CSC411/10_nn1.pdf>

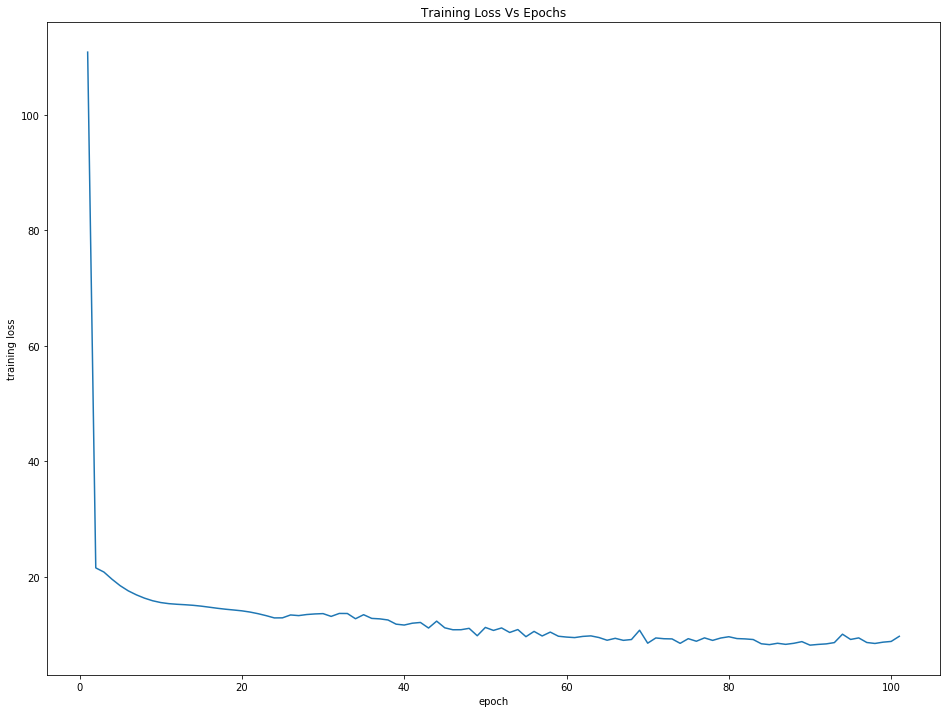
<http://www.philbrierley.com/main.html?code/bpproof.html&code/codeleft.html>

### 4 (b) (i) Training Procedure

For Temperature:1

Hidden units:300

Sequence Length:20



Plot: Training Loss Vs Epochs

For Temperature:1

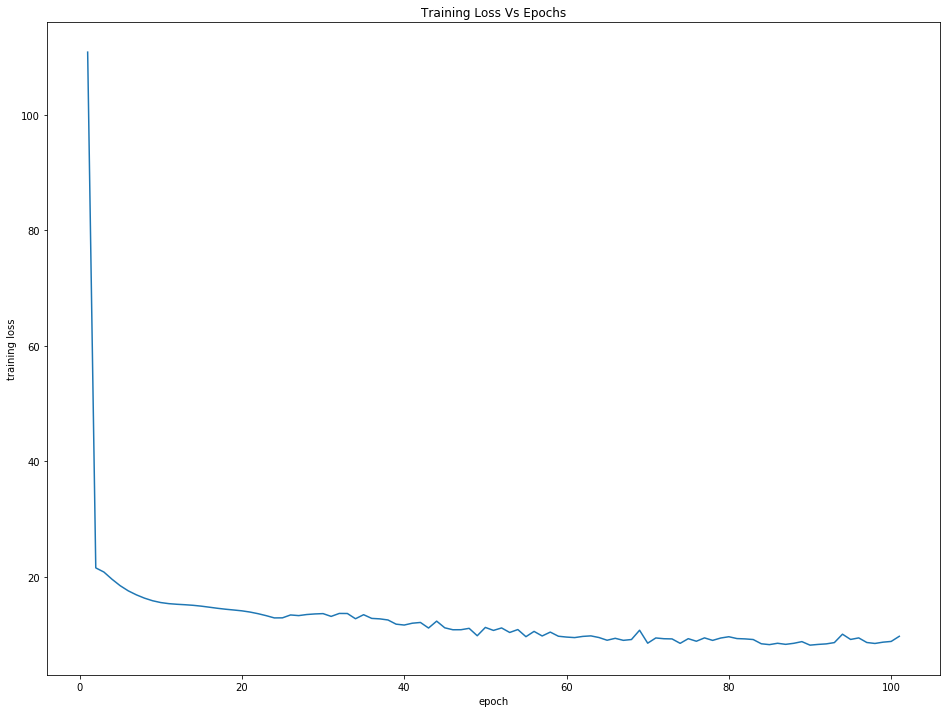
Hidden units:300

Sequence Length:20

The training Loss drastically decreases as the epochs increases leading to efficient training of the network.

### 4 (b) (ii) Breaking Points (20, 40, 60, 80, 100)

After feeding a chunk of training data, the network learning efficiency increases as the epochs reckon to 100 and the loss decreases



Plot: Training Loss Vs Epochs

For Temperature:1

Hidden units:300

Sequence Length:20

Learning Progress by Breakpoints set at epoch intervals of 20, 40, 60, 80, 100 documented as follows:

#### At Break Point 20

dk Higher buind wath ustor? doray eavee mithin, I bo neruce of

AnsatioC in a man were mort ti't eagn

#### At Break Point 40

dd!,

and ulting wher sead mirinions cibse; I Pins. I co beriS m

Sives minber

deainer Ds, a ke ford f

#### At Break Point 60

vy aà such a mantwent,

bouk unSe, an a little karter tith the aspected but in so mervathough; and ma

#### At Break Point 80

d) Wat ently wonan

what

we has been not

expaut on aropt porith wish mohe or this

is so mikefar.

a

n

#### At Break Point 100

n!in whel

bor the roing the tolding as wack

I saing, farthew this gain, Gurged fait en the month, b

### 4 (b) (iii) Learning Progress at Breaking Points (20, 40, 60, 80, 100) at different Temperatures (3, 12, 25)

#### **Temperature 3**

##### At Break Point 20

dk7É¹¬ªÂ¹Â¹ªÂ¹Â¹¹4§§ÂÂ4§¹¹ÂÂ#¹4§Â¹¹Â§\§Â¹Â§¹¹§Â¹¹Â§\¹Â§¹Â§\¹¹Â§\¹¹Â§¹Â§\¹ÂÂ§¹ÂÂ

##### At Break Point 40

r grouth I he was a bear and the rope and the rope and the rope and the rope and the rope and the ro

##### At Break Point 60

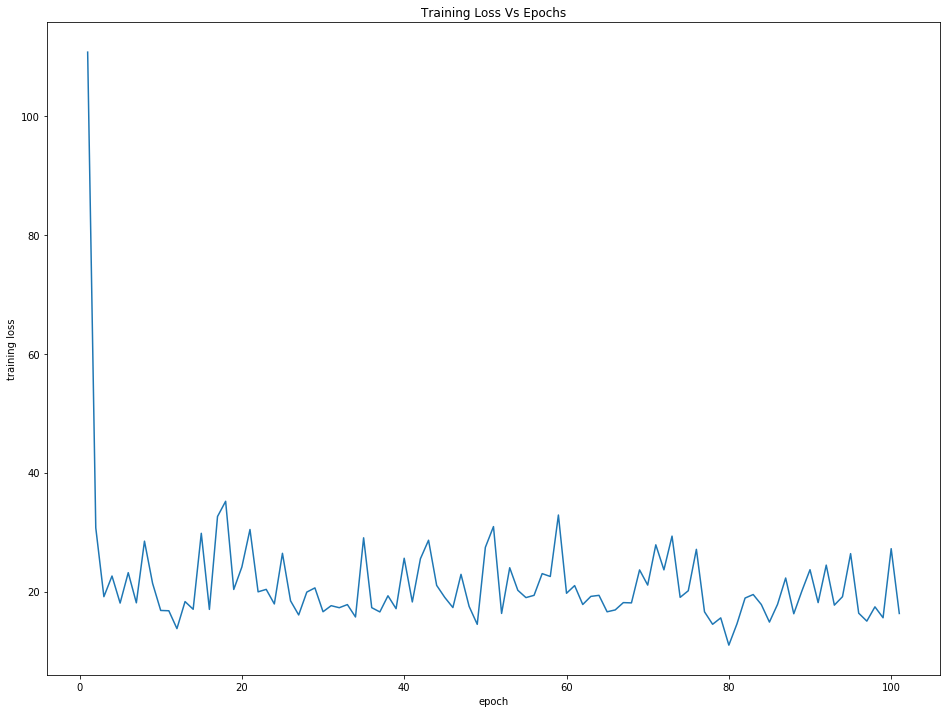
d, I was a sut on the character his poor a such a mare of the good that the good that the good that

##### At Break Point 80

ver to the was and some some to be gound of the stancally and with a voor and a dound and and and a

##### At Break Point 100

rd poller a some at ant to be the self of the of the for the fillow of the for the far to be the for



Plot: For Temperature:3

Hidden units:300

Sequence Length:20

#### **Temperature 12**

##### At Break Point 20

'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''

##### At Break Point 40



(screenshot from ipynb notebook as the text pasted here is occupying one page per character)

##### At Break Point 60

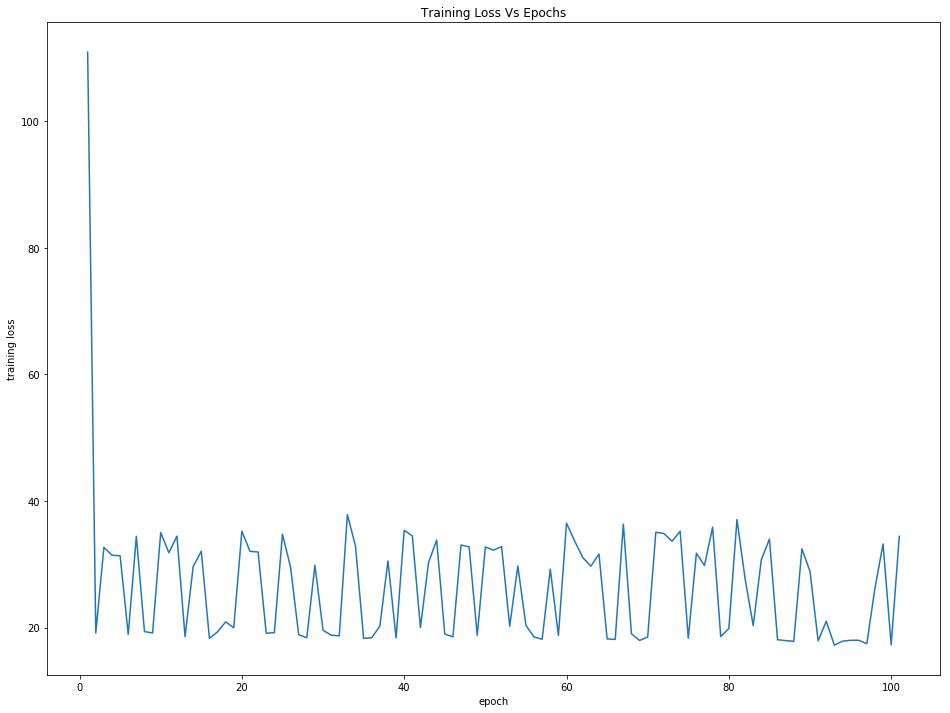
/44444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444

##### At Break Point 80



##### At Break Point 100

n½ù$$óóó$óóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóóó



Plot: For Temperature:12

Hidden units:300

Sequence Length:20

#### **Temperature 25**

##### At Break Point 20

nt the to to to to to to to to to to to to to to to to to to to to to to to to to to to to to to to

##### At Break Point 40

nd the her the her the her the her the her the her the her the her the her the her the her the her t

##### At Break Point 60

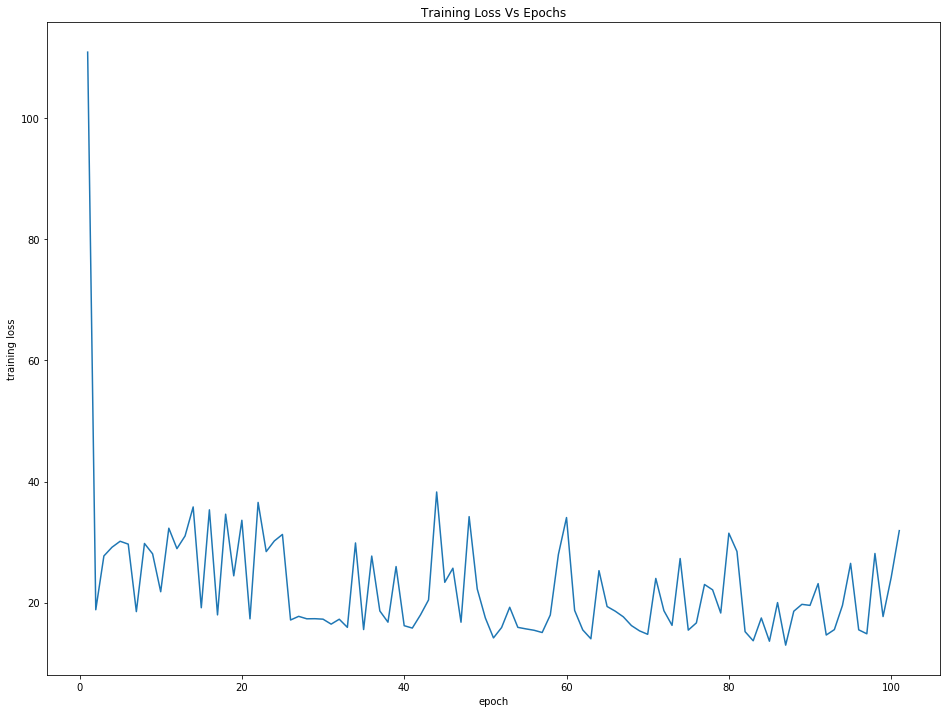
ve gor the sor the sor the sor the sor the sor the sor the sor the sor the sor the sor the sor the s

##### At Break Point 80

t~­Ô|Ô||||||||||||||||||||||||||||||||||||||||||||||

##### At Break Point 100

[ÇÇ»ÇÇÇÇÇÇÇÇÇÇÇÇÇÇÇÇÇÇÇÇÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃÇáÇÃ



Plot: For Temperature:25

Hidden units:300

Sequence Length:20

#### Observations:

For Higher temperatures the randomness of the output sequence is expected and hence it is obvious from the results. We notice that higher temperature T=25, the predicted sequence has more random character sequences but for lower temperature T =3 as the epochs increases the learning gets efficient.

### 4 (c) (i) Number of Hidden Units

#### **Doubling the Number of Hidden units**

For Temperature:5

Hidden units:600

Sequence Length:20

##### At Break Point 20

ngh I his the hist an the rian the ripreather a dith the rian the ripreather a dith the rian the rip

##### At Break Point 40

EIxÌÅ¨ÄÌÄÌÌÄÄÄá¨ÄáÌÌÌÄÌáÄáÄÄáÄÌÄÄáÄÌáÌáÄÌáÄÌÄÄáÄÌÄÌáÄÌÄÄáÄÌáÌáÄÌáÄáÄÌáÄÌÄÌáÄÌÄÄáÄÌÄÌáÄÌÄÄáÄÌáÌ

##### At Break Point 60

d y wore a sore a mather the sore the beather the sore the beather the sore the beather the sore the

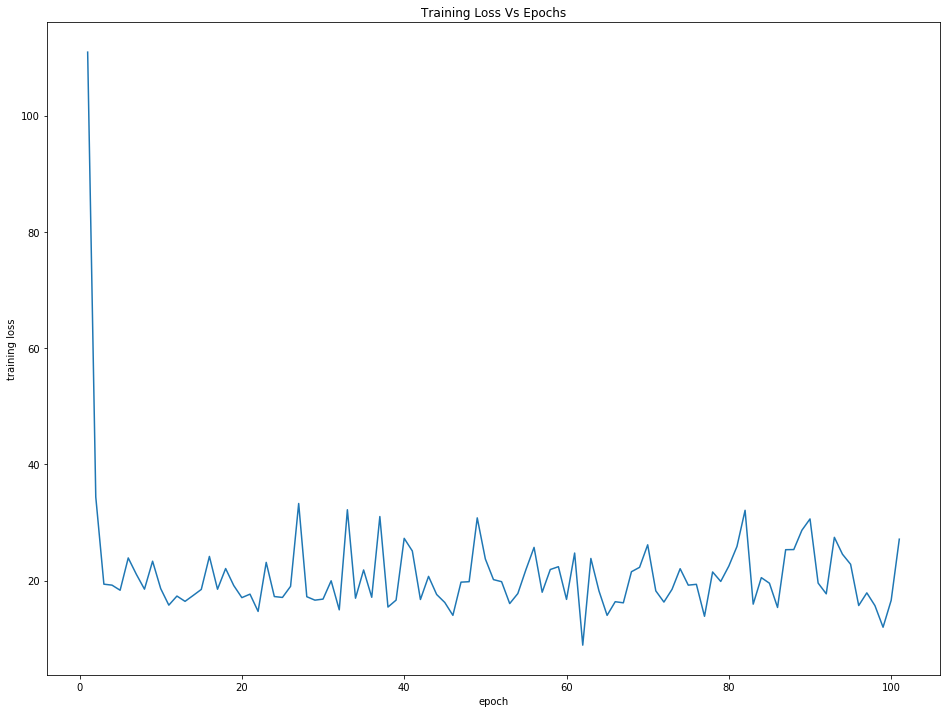
##### At Break Point 80

kv;;.

TY

##### At Break Point 100

x!y,»ÊÝ¹ÊÊ¹Ê¹ÊÊ¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹Ê¹



Plot: Doubling No. of Hidden Dimensions, H=600

For Temperature:5

Hidden units:600

Sequence Length:20

#### **Halving the Number of Hidden units**

For Temperature:5

Hidden units:150

Sequence Length:20

##### At Break Point 20

Thed, I the and and and and and and and and and and and and and and and and and and and and and and

##### At Break Point 40

laver a mear the last that the ancent of the list that the andentle a singer of the last that the an

##### At Break Point 60

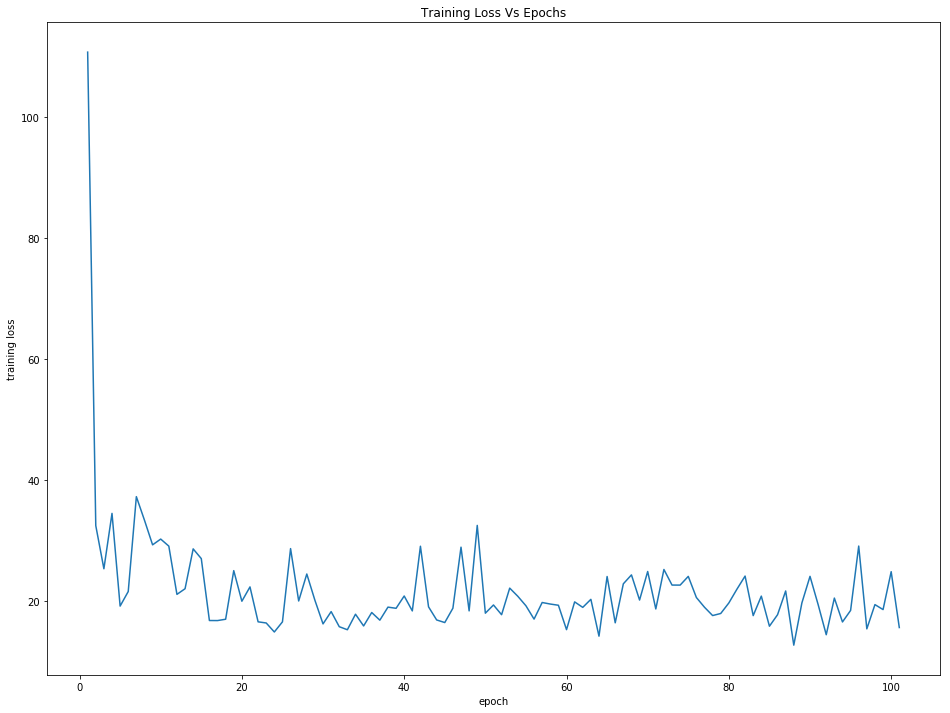
d bforsed, If a distors of the andinger and the andinger and the andinger and the andinger and the a

##### At Break Point 80

dgbby`xxx%¸ã¸¸¸¸xx¸¸xx¸¸x¸x¸xxx¸x¸xxx¸xxx¸x¸xxx¸x¸xxx¸xxx¸x¸xxx¸xxx¸xxx¸x¸xxx¸xxx¸x¸xxx¸xxx¸xxx¸x

##### At Break Point 100

gponting the reat dear reser a distory that the such a little and the state and the state and the st



Plot: Halving No. of Hidden Dimensions, H=150

For Temperature:5

Hidden units:150

Sequence Length:20

#### Observations:

As the Number of Hidden Units in the Hidden layer are increased, the training time increased .

it is also observed that the training loss is comparatively higher than when the hidden units are halved.

With the Hidden Units halved, the training time is faster.

And also the training loss is lower compared to the hidden units are doubled.

### 4 (c) (ii) Sequence Length

#### **Doubling Sequence Length: 40**

For Temperature:5

Hidden units:300

Sequence Length:40

##### At Break Point 20

gd, Ix¤Ø»ÜÙÁç\*öçÂ¥ö»ö»¥­ö\*¥öö\*öç¥ö»¥ö»\*çö»

##### At Break Point 40

dd,

WWvexphith;

TWÈ¿0¸0­¿0x0¿00­00x0­000Áå­»¾x0>å0å­0åå0ÁåxÁå0Áåxååxåå0å¿xåå0Á¿xå¿x

##### At Break Point 60

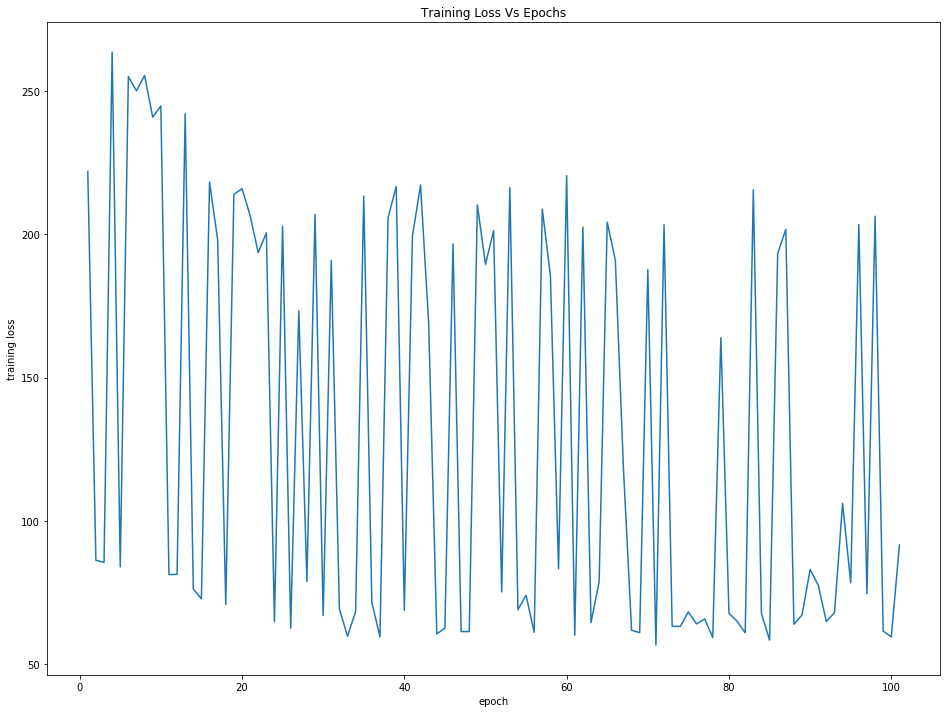
ng the with the wast make to the wart to her an the wart to her an the wart to her an the wart to he

##### At Break Point 80

t my the for the onders and the former to the rector to goon the rector to good to the rector to the

##### At Break Point 100

ngay I good that to the ond as of the mong the ond as of the dease and and and and and the surselary



Plot: Doubling Sequence Length, Seq Len = 40

For Temperature:5

Hidden units:300

Sequence Length:40

#### **Halving the Sequence Length: 10**

For Temperature:5

Hidden units:300

Sequence Length:10

##### At Break Point 20



##### At Break Point 40

sdexthen the reath the reath the reath the reath the reath the reath the reath the reath the reath t

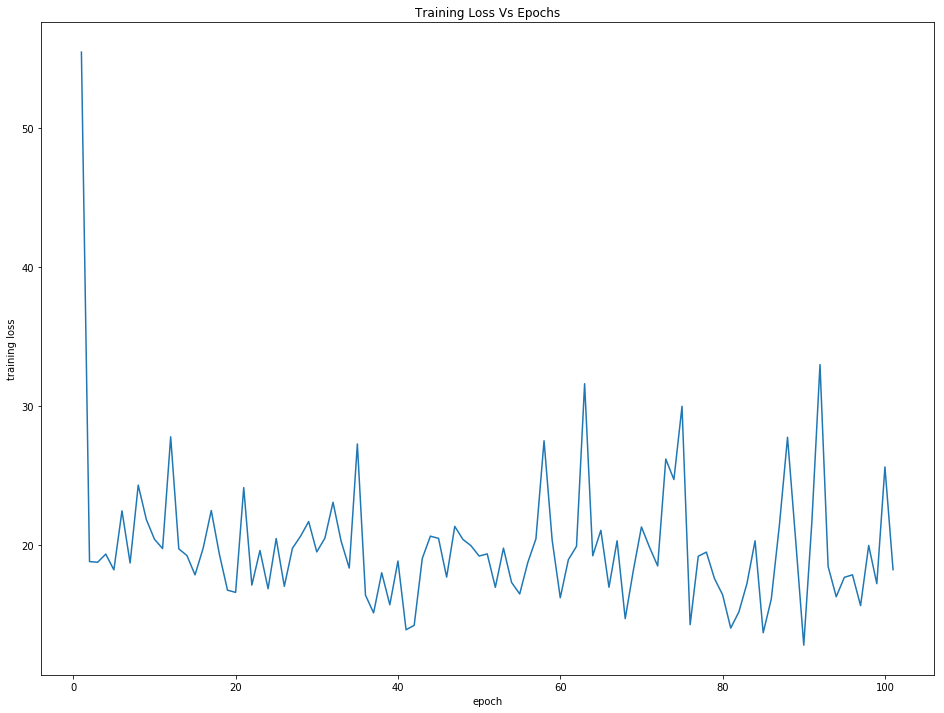
##### At Break Point 60

fd to be a listle an the onntor of the action the and recone of the and recone of the and recone of

##### At Break Point 80

nd the sucton to be the was of the was not ther the worther his own the was of the what whith the st

##### At Break Point 100

t I her his arser that his prong that the reagn a sto kO the protont and the forn that a mintle and**

Plot: For Halving the Sequence Length, Seq Len=10

For Temperature:5

Hidden units:300

Sequence Length:10

#### Observations:

When the sequence length is doubled to 40 characters the training loss is much higher and when the sequence length is halved to 10 characters the training loss is significantly lower. There is no significant improvement in the run times as both are noticed to be same. It is observed that with smaller sequence length the training goes well with high accuracy and less loss.