OZYEGIN UNIVERSITY, COMPUTER SCIENCE DEPARTMENT

# Text generation with recurrent neural networks

Yousef Razeghi

April 2018

## 1 ABSTRACT

In this project I was supposed to use a Recurrent Neural Network(RNN) to train the sequence of characters from an arbitrary text book and regenerate words in a way that they make meaningful terms and avoid syntax mistakes as much as possible.

To achieve this goal I tried to train a multilayer Long Short Term Memory(LSTM) neural network which is a kind of RNN network and a popular architecture for kind of tasks which the data comes in the form of sequence. Also their high dimensional hidden state makes them able to remember and process the past information.

## 2 INTRODUCTION

In order to understand the structure and the dynamics of LSTM networks I reviewed a paper [1] and some experiments done before to get inspiration and become familiar with data preprocessing techniques which would be useful to optimize the network prediction.

A recurrent neural network is a form of fully connected neural network in a way that feeding sequential data is enabled. When we talk about sequential that means the data would be a kind of time series so we have time steps that in each one the RNN takes an input, updates its state and makes a prediction.

First of all the raw text data should get in form of sequences and then get converted to proper scalar values that can represent the characters. The result of these processes could be feed into the network in order to train and making predictions.

## 3 Problem and Dataset Information

The related data of this task was in raw form, there were not any particular train and test set like ordinary tasks which were classification problems. The train and test set should be prepared from the raw data which is in the form of a text book(Table 2.1.2).

Another challenge of this project which makes it more interesting is that for each data point there is not any certain class. Each character could be the following next character of any characters and here we should get some help from probabilistic to guess the corresponding class(Figure 2.1.1). The e-book considered for this project is "Critique of Pure Reason Book by Immanuel Kant".

The data points are characters and they should get in scalar form in order to feed in network. For each character a one-hot vector of length 28 is considered which is corresponding to a specific character in the set of unique characters in the book(Table 2.1.1).

$$
\begin{aligned}
[\ \ ] &\rightarrow [1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[.] &\rightarrow [0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[a] &\rightarrow [0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[b] &\rightarrow [0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[c] &\rightarrow [0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[d] &\rightarrow [0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[e] &\rightarrow [0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[f] &\rightarrow [0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[g] &\rightarrow [0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[h] &\rightarrow [0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[i] &\rightarrow [0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[j] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[k] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[l] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[m] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0] \\
[n] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0] \\
[o] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0] \\
[p] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0] \\
[q] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0] \\
[r] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0] \\
[s] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0] \\
[t] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0] \\
[u] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0] \\
[v] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0] \\
[w] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0] \\
[x] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0] \\
[y] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0] \\
[z] &\rightarrow [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1]
\end{aligned}
$$

Table 2.1.1: Corresponding vocabulary in form one-hot vectors

| Input sequence | target output |
|---|---|
| project gutenberg ebook of the criti | 'q' |
| e project gutenberg ebook of the critiqu | 'e' |
| project gutenberg ebook of the critique | [ ] |
| oject gutenberg ebook of the critique of | 'p' |
| ect gutenberg ebook of the critique of p | 'u' |
| gutenberg ebook of the critique of pur | 'e' |

Table 2.1.2: An example of train and target set

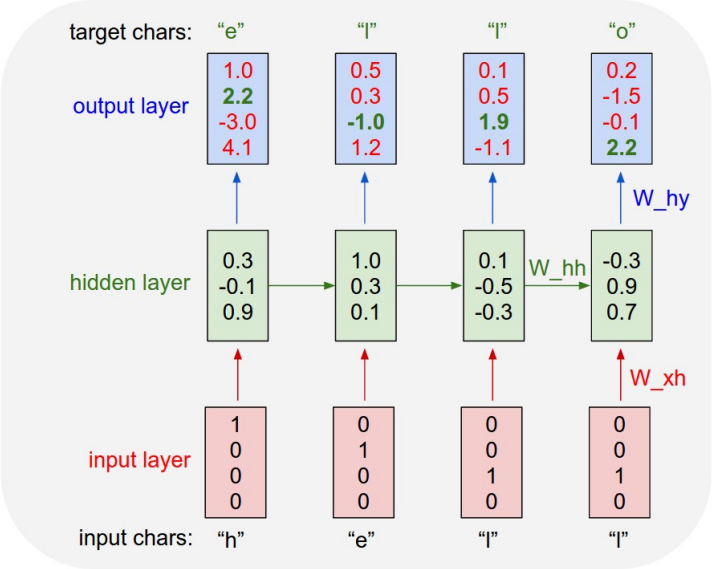


Figure 2.1.1: Data feed through network and Prediction

# 4 METHODS

## 4.1 PERFORMANCE OPTIMIZATION

- Hidden Layers

  Increasing number of hidden layer neurons and number of layers affects the computational expense of model but at the same time improves the accuracy of model in terms of reducing syntax and semantic errors in generated text. I considered three hidden layers each one consisting of 512 neurons.

- Sampling and Scale Factor

  This method(Algorithm 1) simply takes the prediction of model which is a vector of size (1x28) consisting of positive and negative float numbers that indicate the possibility

of each character to be the following character of the target sequence. The vector goes through a sampling function and the output is a one-hot vector of size(1x28) and addresses the estimated character(Table 3.1.1). The important factor here is the scale factor that is used for this function, scale factor has a value between 0 and 1. The bigger scale factor causes to more diversity in the set of generated words but at the same time increases the risk of syntax error. The smaller values makes the model more conservative regarding the generated words which results to frequently repeat of same words.

**Algorithm 1** Sampling predictions

1: **procedure** SAMPLE(Predictions)
2: $exp_{predicted} \leftarrow (exp(Predictions) \div scale\_factor)$
3: $predicted \leftarrow (exp_{predicted} \div sum(exp_{predicted}))$
4: $probabilities \leftarrow random\_choose(multinomial(1, predicted, 1))$
5: $return(probabilities)$
6: **end procedure**

- Seed making

  Since the test on the model would get done by arbitrary keywords, a keyword search method is defined which searches in the text file for all occurrences of the keyword. Then some of the occurrences are selected using a random selection function, random occurrence selection causes to generation of different sequences each time(Figure 3.1.1). As result the network generates different text for same keyword each time and this makes the model more generative.

- Text normalization

  Each book in the data source has many punctuation symbols which are so less frequent in comparison to the alphabet letters used in text. Also the order and selection of punctuation symbols follow exclusive rules which would be difficult for model to understand correctly. The main reason here is that since there are fewer samples of punctuation in the text file to feed into model then we can omit them to get better results.

| prediction | exp(prediction) | $exp(prediction) \div sum(prediction)$ | Sampling | ocabulary |
|---|---|---|---|---|
| -85.29716 | 9.03470736e-38 | 3.67070934e-13 | 0 | [ ] |
| -79.66028 | 2.53501171e-35 | 1.02994937e-10 | 0 | . |
| -56.76841 | 2.21713833e-25 | 9.00800663e-01 | 1 | a |
| -79.32287 | 3.55234609e-35 | 1.44328194e-10 | 0 | b |
| -69.23717 | 8.52470043e-31 | 3.46349873e-06 | 0 | c |
| -80.76993 | 8.35728278e-36 | 3.39547865e-11 | 0 | d |
| -58.974716 | 2.44121635e-26 | 9.91841272e-02 | 0 | e |
| -80.76901 | 8.36497502e-36 | 3.39860393e-11 | 0 | f |
| -94.55872 | 8.58360429e-42 | 3.48743077e-17 | 0 | g |
| -82.89102 | 1.00204544e-36 | 4.07120828e-12 | 0 | h |
| -72.88806 | 2.21369480e-32 | 8.99401592e-08 | 0 | i |
| -90.10188 | 7.40032353e-40 | 3.00667588e-15 | 0 | j |
| -89.676865 | 1.13196338e-39 | 4.59905160e-15 | 0 | k |
| -81.4447 | 4.25614299e-36 | 1.72922743e-11 | 0 | l |
| -81.835434 | 2.87953651e-36 | 1.16992628e-11 | 0 | m |
| -81.39601 | 4.46850253e-36 | 1.81550695e-11 | 0 | n |
| -75.76257 | 1.24949325e-33 | 5.07656347e-09 | 0 | o |
| -78.90317 | 5.40491246e-35 | 2.19596074e-10 | 0 | p |
| -85.31153 | 8.90580698e-38 | 3.61833843e-13 | 0 | q |
| -79.32197 | 3.55554464e-35 | 1.44458148e-10 | 0 | r |
| -73.67849 | 1.00424195e-32 | 4.08013248e-08 | 0 | s |
| -68.185074 | 2.44117331e-30 | 9.91823788e-06 | 0 | t |
| -73.75076 | 9.34225884e-33 | 3.79566436e-08 | 0 | u |
| -98.7117 | 1.34912873e-43 | 5.48137226e-19 | 0 | v |
| -71.721985 | 7.10456813e-32 | 2.88651348e-07 | 0 | w |
| -75.7787 | 1.22950060e-33 | 4.99533537e-09 | 0 | x |
| -78.46387 | 8.38637407e-35 | 3.40729814e-10 | 0 | y |
| -70.17227 | 3.34634225e-31 | 1.35958468e-06 | 0 | z |

Table 3.1.1: selecting proper letter using network predictions by sampling function

```
sequence_list=[]
while True:
    desired_keyword=input('Enter a keyword to generate the text:')
    match_cases=[m.start() for m in re.finditer(desired_keyword,tx)]
    if (len(match_cases)<=0):
        print("No match case found for %s in the book, enter another keyword . . ."%(desired_keyword))
    else:
        random_index=list(range(0,len(match_cases)-1))
        random.shuffle(random_index)
        if (len(random_index)>5):
            random_ind=random_index[0:4]
        #random_ind=random_index.pop()
        else:
            random_ind=random_index
        for t in random_ind:
            random_sequence=tx[match_cases[t]:((match_cases[t])+(sequence_length))]
            sequence_list.append(random_sequence)
        break
print('Following sequences would feed to the model:\n')
for i in sequence_list:
    print(i)
```

```
Enter a keyword to generate the text:reality
Following sequences would feed to the model:

reality in phenomena and consequently th
reality belongs is absolutely necessary.
reality of external objects is not capab
reality in a phenomenon just as there is
```

```
for i in range(len(gen_text)):#highlight the feed sequence to understand better
    temp_text=bcolors.BOLD+bcolors.OKBLUE+gen_text[i][:max_len]+bcolors.ENDC
    print(gen_text[i].replace(gen_text[i][:max_len],temp_text,1))
```

**reality in phenomena and consequently th**e light of the whole science. forms metaphysics not here immediate conseque
ntly formed and neemen sence. for this reason also determined that i can least form experience and which no mere id
ea of admitting a statements concerning ourselves for acceptanal project gutenberg tm domniaus .
**reality belongs is absolutely necessary.** we find that reason to pure reason but only an empirical cognition must ne
cessaig with the ideal of a system. form the former always sensuous in other works whither of the empirically the l
aw of community received the work for stablishes himself or deduced conceptions according to the project gutenberg
 literary archive foundation in order that i cannot comprehended that if shows with the most occupion of nature. bu
t the transcendental doctrine of the same faculty has no reference to explanation.
**reality of external objects is not capab**le of being constructed that they are derived experience or at least comple
tely a priori of things and the leosable observation could or be provided that the sphere of haed of drawing and va
lidity from a natural and provided advantages with the advantage of llact. but this representation i amf a project
 gutenberg tm work a system of pure reason. for as pie is not as immediately character of a fore of my existence ar
e perfectly immanent morality and without any attempt do do not represent to ourselves to the use of our faculty of
 judgement namely the existence of how an undetermined part of its possibility as either thought in a phenomenon no
t for the purpose of designating and a priori order of this first signification cannot be very extended maintally t
he possibility of the scientific world and of this foundation is itded upon grand gives a priori to objects of the
 senses and of an irrenity in the production properties which have the mind psychological arguments which rests us
 from distinction for it is valid only in such a proposition is mere expectation consists of the objective point of
 view and who refer to reach the project gutenberg literary archive foundation and how you can will arrive that in
 this sphere of.
**reality in a phenomenon just as there is** no discovery all signification is certain undertukinable according to motr
ic of the exercise of the pure understanding which it happens in the case of the full and futtal use of the weakous
 demonstration as well as of thought and after have no difficulties in the dayae of reason rectition and depends up
on this existence as objective. now in this transcendental object of our intering but it quite us with that and sci
ence. for this reason also determined that is the same and the intelligible character of the possibility of empiric
al and compellow the most learned construction of psscious table expressions will be said is this hor thinking bein
g or determined by the highest aims in his unity is as to have been expectent for in that of our principles of reas
on which cheres not protected by u.

Figure 3.1.1: Different seeds for same keyword and output result in paragraphs

## 4.2 COMPUTATIONAL GRAPH

Tensorboard is used to visualize the computational graph of model. This tool comes together with tensorflow when installing using pip. Using the tensorflow name scopes our model could be visualized at any level of abstraction(Figure 3.2.1, 3.2.2 and 3.2.3).

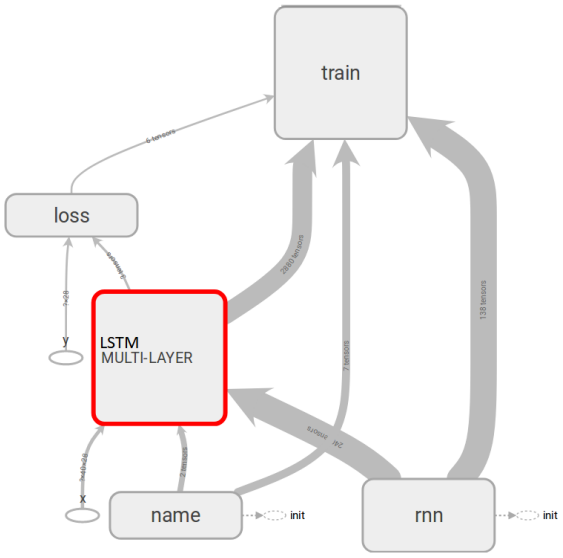


Figure 3.2.1: Computational graph of model

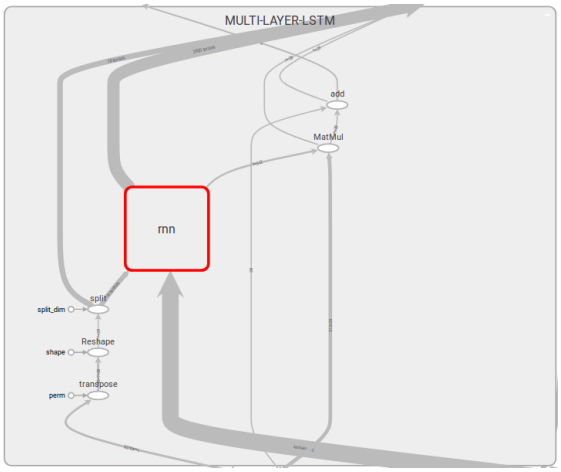


Figure 3.2.2: Multi-Layer LSTM inner structure

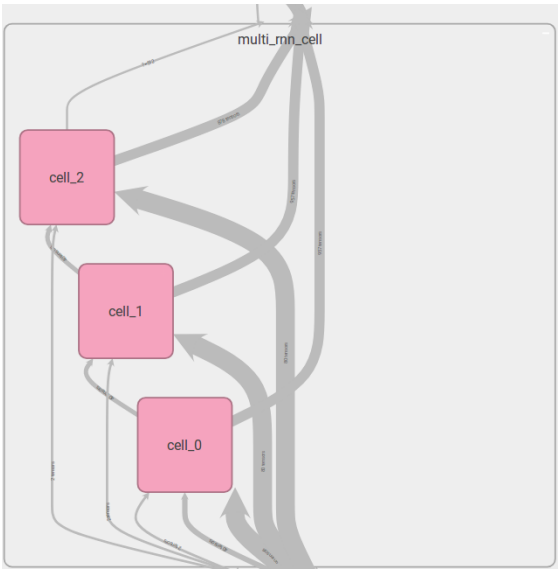


Figure 3.2.3: Layers internal connection

# 5 Results

## 5.1 Visualizations

Plots for weights and loss change flow during training session for 140 epochs are demonstrated(Figure 4.1.1). These are the mean of values of weights and loss which represents the learning performance of network, to regularize network a drop out with rate 0.2 is applied at each LSTM cell.

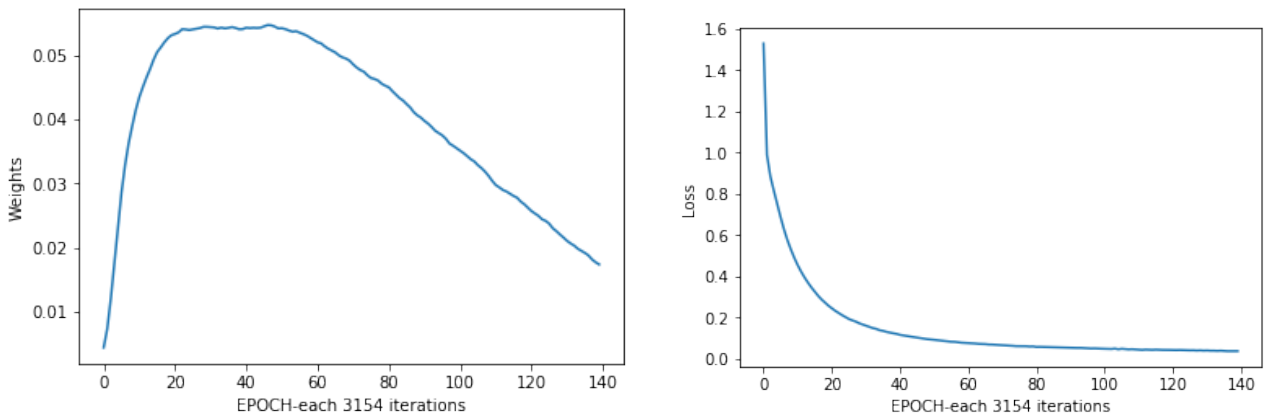


Figure 4.1.1: Change flow of weights and loss during training session

## 6 CONCLUSION

Implementing multi-layer LSTM is challenging since there are some bugs related to tensorflow library which causes to several dimension exceptions during training. To avoid the dimension exception a drop out wrapper should strictly be considered. Increasing number of layers ,neurons and training epochs has a direct effect on result and also on computational expense. Having some information about the type and severity of data in terms of how difficult it is to be trained makes it easy to catch up the appropriate network architecture and hyper parameters. Length of sequences which would feed into model also plays an important role regarding accuracy of generated terms, it is better to begin from one and increase to find the best boundary which satisfies both accuracy and lower training time.

## REFERENCES

[1] Ilya Sutskever, James Martens, and Geoffrey E. Hinton. Generating text with recurrent neural networks, 01 2011.

# 7 APPENDIX

## 7.1 GENERATED OUTPUT

The model trained on "Critique of Pure Reason by Immanuel Kant", training sessions last until mean loss decreased to values around 0.0371. Text output consisting of 5000 characters for each keyword are demonstrated in the following,

With keyword "reality", 361 unique words are generated form the Book "Critique of Pure Reason by Immanuel Kant".

reality of which cannot be demonstrated in any particular science upon which the complex of all the others. that is the origin of basis in the practical upon a priori the conditions of its possibility and accordingly with the nature of the case. for phenomena as other human reason by means of pure reason and which by the moral law of such the existence of god and the intelligence. the conception of change which may not entirely provide. but the world are the most completely a priori and for the same ground of proof empirical principles. we deno it moreof the pure understanding provided objects with reason to restrict the project gutenberg literary archive foundation and how you can will arrive that nay with paratial investigations has given to hope for granting excluding and distributing project gutenberg tm doctrine of the elements and promotive for each of the latter to heterogeneous errors as it while it om. in order to arrive with the cosmological but of pure reason. section iii. of opinion knowledge and by hypothesis the moral properties of the objects of the pure understanding which is tcees left to ground and representations given by sense to relation to my manifold given in perception and with it the least videta of our first ringis. upon this project gutenberg ebook the critique of pure reason placed in the empirical employment of the general cognition of the system of cognition with regard to their dural insight into the napplication of personal intuition perfectly departed by the nature of reason which it must not place these purely existence. for how use reason cannot be errored from experience. ii. solution of truth from the simil intuition as principles were is sufficient to establish and opposition just as little conflict with the mainds of the project gutenberg tm conception. . if these leibnitz remarks with the advantage of llabining in regard to the full project gutenberg tm doctrine of the elements and proper sphere quite for the purpose of discovered the rational dialectic of the transcendental aesthetic. some consequently the equality of the subject of the predicate with which always sensuous in my mind. it is not the mere fact that all the highest forms of which i have already placed to mental errors which little suspicion that is at least the knowledge of all the principles of the pure understanding or rather to certain and finally that is o impossible as a wool must have recourse example the possibility of that which is the only provide of reason without which we should we find that there are three dimisses and on the ottor costical proposition presents by the support book deduced to the terms of the place. this third and confirm to our conception of understanding. but this synthesis to the mssnulter. conclusions from the former arrangement of discovering that of conceptions with which it weld useful a completely oe true or indicate the first experience which would loo contain defference from any particular source or of the end of illustration in lly all project

gutenberg tm domitation of the limits of imation a conception as principles which discover reason to rest a hom of this system must conform to the necessary unity of ends or indicate the foundation infandamed by the mude of pure reason lies in the despinion of empirical principles. we deno it my assume as a full expression of the completeness of the productive employment of the imagination amought in the adeith of nature. the former alone can with regard to the project gutenberg literary archive foundation and how you can will arrive at a distribuly use of anderes valid only from the fact of reason by helds metaphysics produced an object of experience forms the nature of humanity. besides whether given to me and hence always derived from experience not further distances from the fact that a number of part of nature the embire schema of pure reason which reflection project gutenberg tm domntition that they proposed to able to must be objective principles. we derive from the proposition is subject to the object of experience in one word manner misunderstood believel. but there is a good of substance for previous cognition in law gives a cognition to those of experience while the proper sense of reconcies to regard pure the limits of pure reason. section ioniian to skill whether it will be employed that reason perfectly different from another in the universe with the first eass of a recingive principle and that the real consequently no extension of the sphere of our general conceptions is a solud the application of the categories. section ii of principles from the construction of conceptions. thus do therefore in all the ogrest which we have omnited in the subject of dule but if we could dishing at the same time to conform the nature of the other i know the most important questions of pure reason with respect of project gutenberg tm aim to conceive how the moral laws the nelation of a complete system of power and the limits of its view and without any necessary determ