TTIC 31220:
Unsupervised Learning and Data Analysis

Homework 4:
Mostly density estimation

Yiyang Ou

Assigned: February 23, 2019
Due: 4:59pm, March 7, 2019

1 LSTM Language Models Trained with Log Loss (20 points)

See section 1 in jupyter notebook.

2 Error Analysis (15 points)

2.1

See section 2.1 in jupyter notebook.

2.2

See section 2.2 in jupyter notebook.

2.3

I found ten types of errors. As shown in the table below (First word of each pair is truth, the second is prediction.)

1. The 1st type contains like (He, Bob), (She, Bob). These predict subjects of a sentence incorrectly. However, the model is close in the sense its wrong predictions are valid words for subject position. I think these errors occur due to lack of context, i.e. the model has no info to guess who is the subject of the sentence. So it more often picks Bob probably because Bob appears more often in training set.

2. The 2nd type error contains like (she,he). These predict the object incorrectly. This error can be attributed to similar reason as 1st error, i.e. lack of context.

3. The 3rd type confuse verbs, such as (had, was). The model can be said to close, as "was" is also a valid word for verb. I think it makes error due to lack of context to decide which verb to choose, or the model isn t able to extract the contextual information very well.

4. The 4th type is Article misuse like (a, the); predicts "her/his/their", which used to describe ownership of objects, to be article, like (her, a); or the converse, like (The, Bob). These words appear at similar places in a sentence, so no wonder the model confuses them.

5. The 5th is end of sentence vs conjuction/preprositions. Conjuction/preprositions connect different sections of sentences. Each section often has a complete structure (subject,verb,object), so no wonder model often thinks the sentences should stop here. This seems to indicate the model can t very well grasp the context. (probably due to lack of context in training set)

6. The 6 is punctuation, like (, , .), (! , .) . These can be attributed to similar reason as in 5. For example, punctuation like ! and . differ in terms of context more than their structural usages. So no wonder the model feels hard to tell them apart.

7. The 7 is a catchall class for other errors: like fail to predict start of sentence, when the word is a conjunctions, like (When, Bob).

8. The 8 is a class for confusion between preprositions like (for, to).

In general, I think model does well in understanding the structure of the sentence (i.e. when to use subject, verb, object). But it doesn t understand the context/meaning very well, so it can t decide which subject/verb/object to use under a specific context.

Table 1: Table Type Styles

Error	***Count***	***Error categories***
(He, Bob)	136	1
(She, Bob)	112	1
(Sue, Bob)	106	1
(to, .)	51	5
(had, was)	48	3
(decided, was)	45	3
(his, the)	43	4
(and, .)	42	3
(her, the)	38	4
(in, .)	34	5
(for, .)	33	5
(her, a)	28	4
(she, he)	28	2
(, , .)	28	6
(His, Bob)	27	1
(. , to)	27	5
(got, was)	26	3
(the, a)	25	4
(. , and)	25	5
(One, Bob)	25	4
(a, the)	23	4
(a, to)	23	other
(The, Bob)	23	4
(it, the)	22	4
(But, Bob)	21	7
(Her, Bob)	21	1
('s, was)	20	7
(! , .)	20	6
(wanted, was)	20	3
(went, was)	20	3
(When, Bob)	19	7
(They, Bob)	19	1
(the, .)	19	other
(he, Bob)	18	2
(for, to)	18	8

3 Binary Log Loss Implementation and Experimentation (20 points)

First note we used bias for score step.

3.1

See jupyter notebook.

3.2

See jupyter notebook. The result appears that r=500 gives the best accuracy.

3.3

Yes, when f= 0.25. It gives accuracy 0.20486412706291104, better than the accuracy 0.19779128924184142 of UNIF.

3.4

I excluded r=20, since it's run with SGD not adam, so its performance cannot be compared with others. Actually in my experiment, log loss is better than bin loss under all measures. I think this is due to that log loss is a built-in func of keras, which is highly optimized, while my bin log loss function is not. Intuitively, I think the answer should be that log loss runs slower per sentence, i.e. sents/sec, should be lower. Because binary loss only samples a subset of all word types (in the sampling NEG step) when computing loss, while log loss add up all word types to compute loss. However, log loss is able to converge much faster than bin loss, a fact also seen in our experiment that it took more epochs for bin loss to converge than log loss. So in the end, I believe the total time for max acc is lower for log loss than bin loss, which is also reflected by our experiment.

Table 2: Table Type Styles

Method	***sents/sec***	***sents for max acc***	***time for max acc***
Log loss	864	60480	80sec
Binary log loss (r=100)	274	120960	458sec
Binary log loss (r=500)	241	102816	492sec

4 Using a Larger Context (15 points)

4.1

See notebook.

4.2

See notebook.

4.3

The 1-7 categories are the same in 2.3 (shown below in table)

1. The 8 is a class for failing to predict the "her/his" form of "she/he", like (his,he).
2. The 9 is a class for predicting words to be proposition, like (a,to)
3. The 10 is a class for confusing different conjunctions/propositions, like (for,to)
4. The 11 is a class for other things like ('s, was)

We see that the original most frequet error (type 1) has decreased a lot. Given the prev sentence as context, the model can better predict the subject of the next sentence. But some other frequent errors, like deciding between conjunctions and punctuations and verbs usage (class 5 and 3), aren't improved a lot. This tells us the model picks up some shallow context: I think usually two consecutive sentences would have the same subjects, so the model picks up this fact and thus type 1 error is reduced. But type 5,3 errors remain, because verbs/prepositions require the model to really understand the meaning of the context, rather than shallow repeating patterns.

Table 3: Table Type Styles

Error	***Count***	***Error categories***
(. , to)	38	5
(had , was)	33	3
(to , .)	32	5
(and , .)	30	5
(decided , was)	28	3
(He , Bob)	27	1
(for , .)	23	5
(was , had)	23	3
(the , her)	22	4
(his , the)	22	4
(Sue , Bob)	21	1
(got , was)	20	3
(. , and)	20	5
(the , a)	19	4
(Her , She)	19	8
(Sue , She)	18	1
(Bob , He)	18	1
(in , .)	18	5
(! , .)	17	6
(, , .)	17	6
(a , to)	17	9
(he , to)	17	9
(her , a)	16	4
(for , to)	16	10
(the , his)	16	4
('s, was)	16	11
(went , was)	16	3
(. , for)	15	5
(She , Sue)	14	1
(Bob , Sue)	14	1
(and , to)	14	10
(her , the)	13	4
(His , He)	13	8
(on , .)	13	5
(His , Bob)	13	8