

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس پردازش زبان های طبیعی

تمرین ۳

سجاد پاکدامن ساوجی

۸۱۰۱۹۵۵۱۷

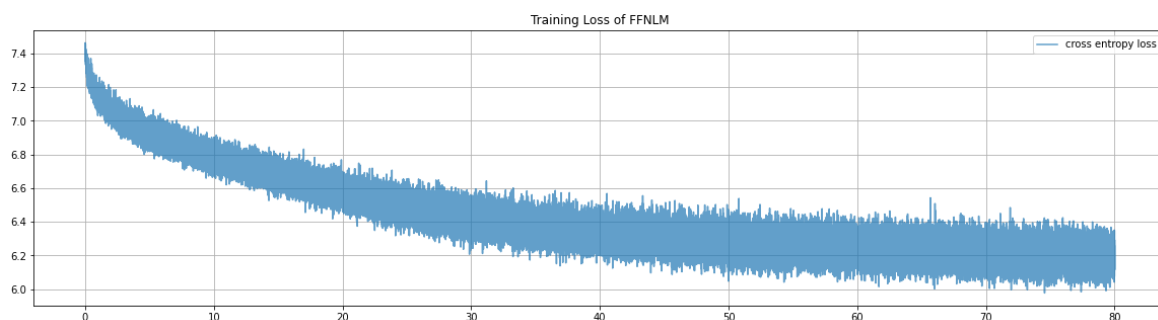
اردیبهشت ۱۴۰۰

فهرست مطالب

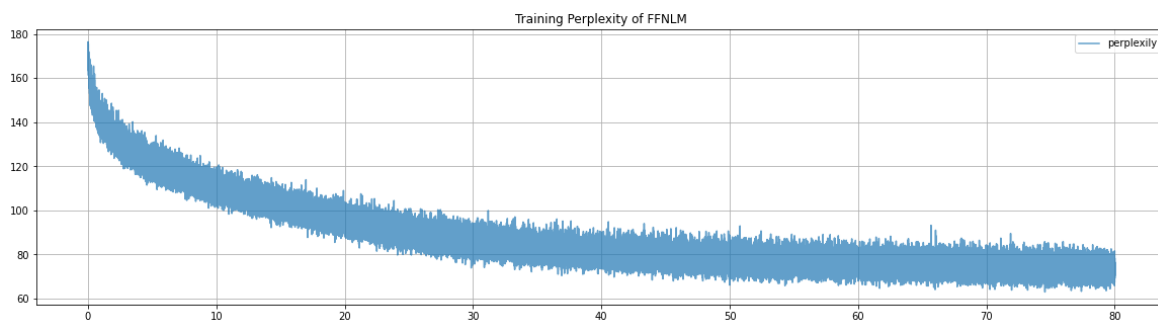
3	۱. Feed-Forward Neural Language Model(FFNLM)
4	۲. Long-Short-Term-Memory Language Model(LSTMMLM)
4	یکسان سازی طول جملات
4	مقایسه پارامتر های LSTM
5	تاثیر Hidden Dimension در یادگیری مدل زبانی
6	تاثیر Number of Layers در یادگیری مدل زبانی
7	تاثیر Embedding Dim در یادگیری مدل زبانی
8	تاثیر Learning Rate در یادگیری مدل زبانی
9	قلم نویسنده
12	اجرای کد ها

۱. Feed-Forward Neural Language Model (FFNLM)

در این قسمت تلاش شده است که یک مدل زبانی بر پایه شبکه های عصبی feed forward طراحی شود. به این منظور طول پنجره ۵ انتخاب شده است. به این معنی که ۴ کلمه اول به عنوان ویژگی های ورودی شبکه و کلمه پنجم به عنوان برچسب به شبکه داده خواهد شد. برای طبقه بندی از تابع هزینه cross entropy استفاده شده است. در شکل ۱ روند کاهش هزینه در زمان آموزش آورده شده است. در شکل ۲ روند کاهش سرگشتگی هنگام آموزش آورده شده است.



شکل ۱. تابع هزینه در زمان آموزش



شکل ۲. سرگشتگی در زمان آموزش

همچنین پس از آموزش شبکه، جمله ای با استفاده از انتخاب های تصادفی بر اساس خروجی شبکه تولید شده است. برای آنکه خروجی شبکه به احتمال تبدیل شود، از تابع Softmax استفاده شده است. لازم به ذکر است که تابع هزینه Cross Entropy معادل یک لایه Softmax به همراه تابع هزینه Negative Log Loss است. بنابراین استفاده از این تابع پس از اتمام آموزش شبکه درست است. جمله تولید شده در زیر آورده شده است.

" Thank you for wanting blood buried hair"

این جمله نشان میدهد که شبکه قادر بوده است که بخشی از قواعد گرامری زبان انگلیسی را شبیه سازی کند، چرا که پس از Thank you معمولا از حرف اضافه for استفاده می شود. همچنین پس از Noun for نیز استفاده از Clause درست است، پس تولید کلمه wanting نیز صحیح است. همچنین ترکیب ۳ کلمه blood buried hair نیز از نظر قواعد زبانی به صورت reduced adjective clause است. با این حال معنی این ترکیب صفتی به نظر درست نمیرسد، بنابراین شبکه قادر به درک پیچیدگی هایی همانند معنای کلمات نبوده است.

۲. Long-Short-Term-Memory Language Model(LSTMLM)

گام های مورد نیاز برای پیاده سازی یک شبکه LSTM تنها مشخص کردن پارامتر های ساختاری آن است. این پارامتر های ساختاری شامل `input_size`، `num_layers`، `hidden_dim` و `output_size` است.

یکسان سازی طول جملات

در این تمرین برای کاهش زمان آموزش شبکه از Batch Training استفاده شده است. جملات هر batch لزوما طول ثابتی ندارند و برای یکسان سازی این طول از padding استفاده شده است. عملیات padding در dataloader و با استفاده از colette-fn انجام شده است که نسبت به pad کردن تمامی جملات dataset مطرف حافظه بهتری دارد.

حال که طول جملات یکسان شده است، اهمیت دارد که تگ های `<pad>` در آموزش شبکه کمترین تاثیر را داشته باشند. برای این کار در لایه Embedding شماره pad داده شده است، بنابراین تگ های `<pad>` در لایه embedding تاثیری نخواهند گذاشت. همچنین برای speed up در شبکه LSTM نیز از sequence packing استفاده شده است. این روش باعث میشود که تگ های `<pad>` به شبکه LSTM نشان داده نشوند. در نهایت در قسمت طبقه بندی نیز Padding Index به عنوان ignore index به تابع هزینه Cross Entropy داده شده است، پس در محاسبه هزینه نیز تگ های `<pad>` تاثیری نخواهند داشت.

مقایسه پارامتر های LSTM

در ادامه به مقایسه تاثیر چند پارامتر در فرایند آموزش و تولید جملات بررسی میشود. در این مقایسه ها، دو مدل تنها در یک ویژگی متفاوت هستند. برای آن که مقایسه به بهترین شیوه انجام شود، با مشخص کردن random seed وزن های دو شبکه نیز یکسان مقدار دهی شده اند. در مقایسه فاکتور هایی مانند سرعت کاهش سرگشتگی، نقطه همگرایی، overfitting و توانایی در تولید جملات در نظر گرفته شده است.

در این مقایسه های، چهار ویژگی شبکه تغییر پیدا میکند که به ترتیب در زیر آمده است.

۱. اندازه لایه مخفی

۲. تعداد لایه های LSTM

۳. اندازه لایه embedding

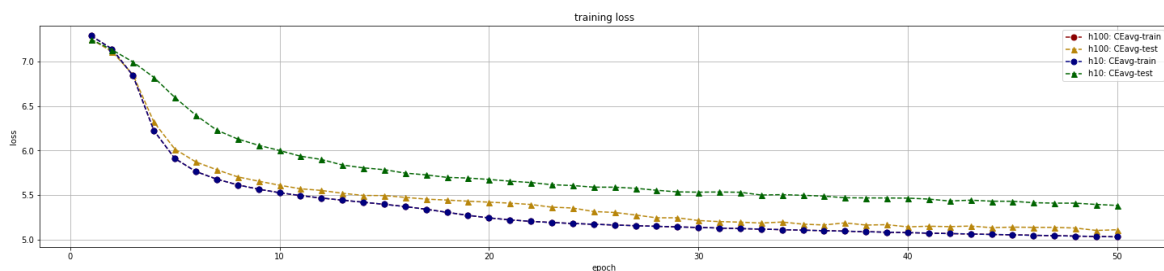
۴. نرخ یادگیری

تاثیر *Hidden Dimlension* در یادگیری مدل زبانی

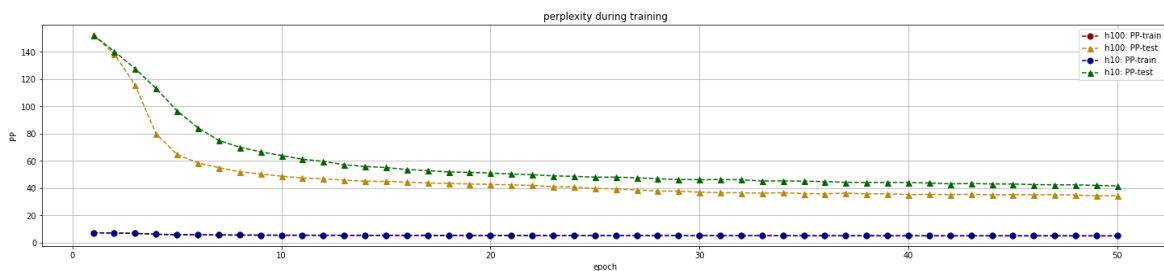
در جدول ۱ مشخصات دو مدلی که مقایسه شده اند، آورده شده است. همچنین در شکل ۳ و ۴ روند کاهش هزینه و سرگشتگی برای داده های آموزش و آزمایش آورده شده است. همچنین جملات تولید شده توسط هر یک از مدل ها در جدول ۲ آورده شده است.

Model Name	h10	h100
Embedding Size	50	50
Hidden Layer Size	10	100
Number of Layers	2	2
Batch Size	256	256
Epochs	50	50
Learning Rate	0.1	0.1

جدول ۱. خلاصه پارامتر های مدل های مقایسه شده



شکل ۳. مقدار هزینه در آموزش شبکه ها



شکل ۴. مقدار سرگشتگی در آموزش شبکه ها

Model Name	Generated Sentence
h10	She had lost part of walked very and saying .
h100	She had lost part of England .

جدول ۲. جملات تولیدی توسط مدل ها

* سرعت کاهش سرگشتگی در مقایسه دو مدل متفاوت نبوده است.

* مدل h10 نسبت به مدل h100 زودتر به نقطه همگرایی رسیده است.

* مدل h100 نسبت به مدل h10 عملکرد بهتری بر داده های test دارد. (مربوط به overfitting)

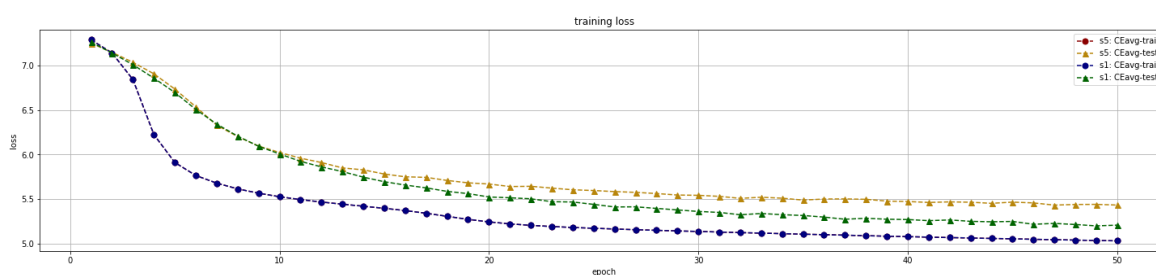
* مدل h100 در مقایسه با h10 جملات دقیق تر و با معنی تری تولید میکند.

تاثیر *Number of Layers* در یادگیری مدل زبانی

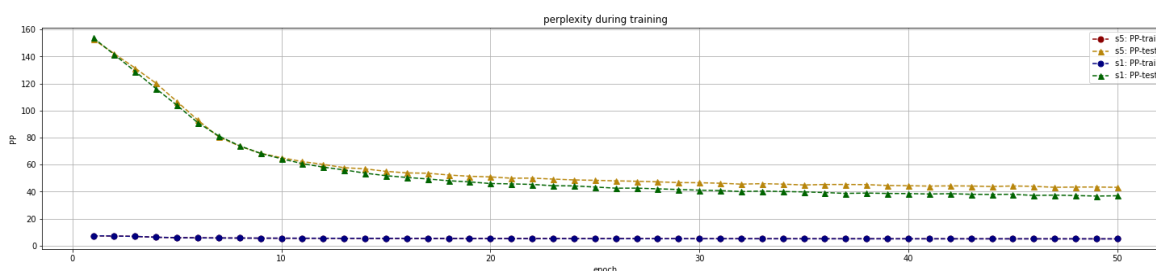
در جدول ۳ مشخصات دو مدلی که مقایسه شده اند، آورده شده است. همچنین در شکل ۵ و ۶ روند کاهش هزینه و سرگشتگی برای داده های آموزش و آزمایش آورده شده است. همچنین جملات تولید شده توسط هر یک از مدل ها در جدول ۴ آورده شده است.

Model Name	S1	S5
Embedding Size	50	50
Hidden Layer Size	10	10
Number of Layers	1	5
Batch Size	256	256
Epochs	50	50
Learning Rate	0.1	0.1

جدول ۳. خلاصه پارامتر های مدل های مقایسه شده



شکل ۵. مقدار هزینه در آموزش شبکه ها



شکل ۶. مقدار سرگشتگی در آموزش شبکه ها

Model Name	Generated Sentence
S1	She had lost part of glowing salt .
S5	She had lost part of yards perhaps as you

جدول ۴. جملات تولیدی توسط مدل ها

* سرعت کاهش سرگشتگی در مقایسه دو مدل متفاوت نبوده است.

* سرعت رسیدن به نقطه همگرایی یکسان بوده است.

* مدل S1 نسبت به مدل S5 عملکرد بهتری بر داده های test دارد. (مربوط به overfitting)

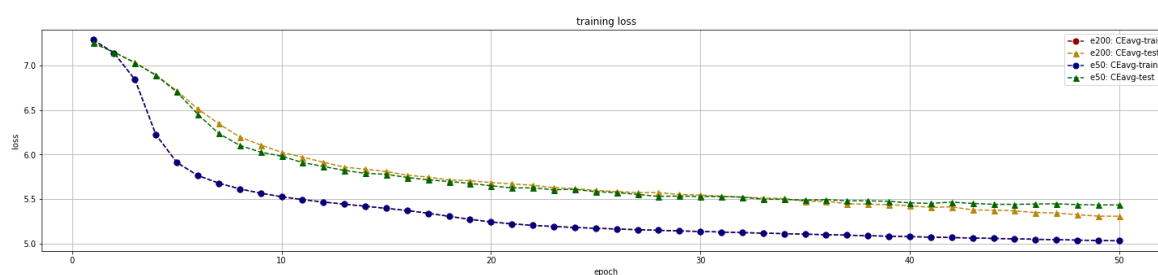
* مدل S1 در مقایسه با S5 جملات دقیق تر و با معنی تری تولید میکند.

تاثیر *Embedding Dim* در یادگیری مدل زبانی

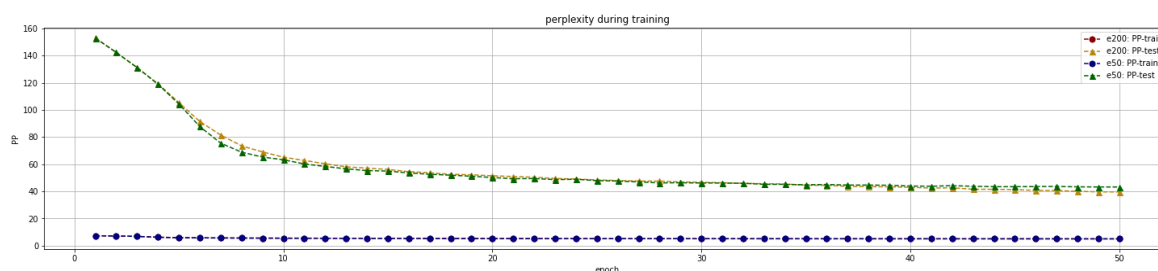
در جدول ۵ مشخصات دو مدلی که مقایسه شده اند، آورده شده است. همچنین در شکل ۷ و ۸ روند کاهش هزینه و سرگشتگی برای داده های آموزش و آزمایش آورده شده است. همچنین جملات تولید شده توسط هر یک از مدل ها در جدول ۶ آورده شده است.

Model Name	E50	E200
Embedding Size	50	200
Hidden Layer Size	10	10
Number of Layers	2	2
Batch Size	256	256
Epochs	50	50
Learning Rate	0.1	0.1

جدول ۵. خلاصه پارامتر های مدل های مقایسه شده



شکل ۷. مقدار هزینه در آموزش شبکه ها



شکل ۸. مقدار سرگشتگی در آموزش شبکه ها

Model Name	Generated Sentence
E50	she had lost part of truck for the distance walk
E200	she had lost part of numbers out and reached self only

جدول ۶. جملات تولیدی توسط مدل ها

* سرعت کاهش سرگشتگی در مقایسه دو مدل متفاوت نبوده است.

* سرعت رسیدن به نقطه همگرایی در E50 نسبت به E200 بیشتر است .

* مدل E200 نسبت به مدل E50 عملکرد بهتری بر داده های test دارد. (مربوط به overfitting)

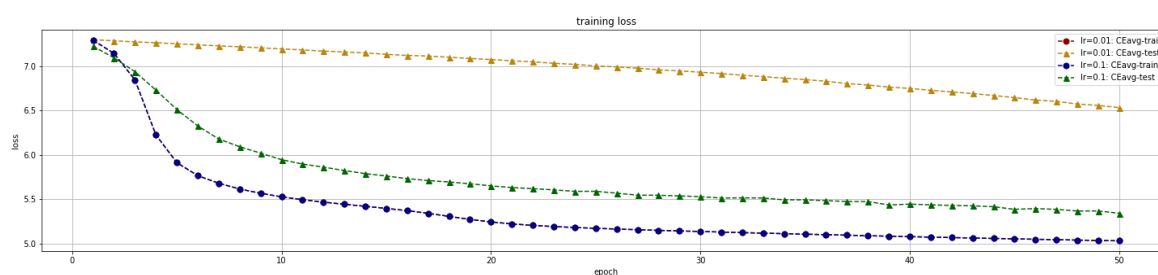
* مدل E50 در مقایسه با E200 جملات دقیق تر و با معنی تری تولید میکند.

تاثیر Learning Rate در یادگیری مدل زبانی

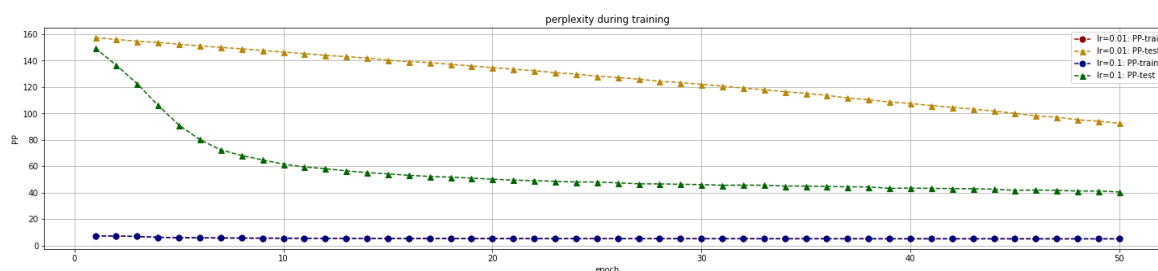
در جدول ۷ مشخصات دو مدلی که مقایسه شده اند، آورده شده است. همچنین در شکل ۹ و ۱۰ روند کاهش هزینه و سرگشتگی برای داده های آموزش و آزمایش آورده شده است. همچنین جملات تولید شده توسط هر یک از مدل ها در جدول ۸ آورده شده است.

Model Name	LR0.1	LR0.01
Embedding Size	50	50
Hidden Layer Size	10	10
Number of Layers	2	2
Batch Size	256	256
Epochs	50	50
Learning Rate	0.1	0.01

جدول ۷. خلاصه پارامتر های مدل های مقایسه شده



شکل ۹. مقدار هزینه در آموزش شبکه ها



شکل ۱۰. مقدار سرگشتگی در آموزش شبکه ها

Model Name	Generated Sentence
LR0.1	she had lost part of other to your fella
LR0.01	she had lost part of stephen lie fish watched to split

جدول ۸. جملات تولیدی توسط مدل ها

* سرعت کاهش سرگشتگی در مدل LR0.1 نسبت به LR0.01 بیشتر است.

* سرعت رسیدن به نقطه همگرایی در مدل LR0.1 نسبت به LR0.01 بیشتر است.

* مدل LR0.1 نسبت به مدل LR0.01 عملکرد بهتری بر داده های test دارد. (مربوط به overfitting)

* مدل LR0.1 نسبت به مدل LR0.01 جملات دقیق تر و با معنی تری تولید میکند.

مشکل تکرار **stopping word** تا حدی در مدل ها وجود داشت. علت آن این است که تعداد این کلمات در متن زیاد است، در نتیجه اگر جهت گرایان به سمتی باشد که این کلمات بیشتر پیش بینی/تولید شوند، مقدار هزینه کاهش پیدا میکند. این مشکل با مشکل **imbalanced classes** که در مسئله های پایه ML مشاهده می شود یکسان است و راهکار های مشابهی نیز میتوان برای آن اتخاذ کرد. برای مثال میتوان تابع هزینه ها به صورت وزندار محاسبه کرد به صورتی که ضریب **stopping words** کمتر از کلمات معمول شود.

برای تولید جملات در این قسمت نیز از روش بخش های قبل، نمونه برداری با شانس ها متناسب با احتمال پیش بینی شده، استفاده شده است.

در این قسمت، مطابق خواسته سوال چهار کتاب انتخاب شده و مدل های زبانی بر روی آن ها آموزش داده شده است. تمامی مدل های زبانی پارامتر هایی که در جدول ۹ آورده شده را دارند.

Model Parameters	
Embedding Size	50
Hidden Layer Size	10
Number of Layers	2
Batch Size	256
Epochs	50
Learning Rate	0.1

جدول ۹. پارامتر های مدل ها

کتاب های انتخاب شده به ترتیب در زیر آورده شده است.

`selected_book1 = 'crocodiles-spirit.epub.txt'`

`selected_book2 = 'zombies-bite.epub.txt'`

`selected_book3 = 'your-guide-to-succeed-after-graduation.epub.txt'`

`selected_book4 = 'towards-a-sacred-sexuality.epub.txt'`

در زیر جملات تولیدی توسط هر یک از مدل های آموزش داده شده، آورده می شود.

`model_book1 -> 'finding evidence in inside university his it'`

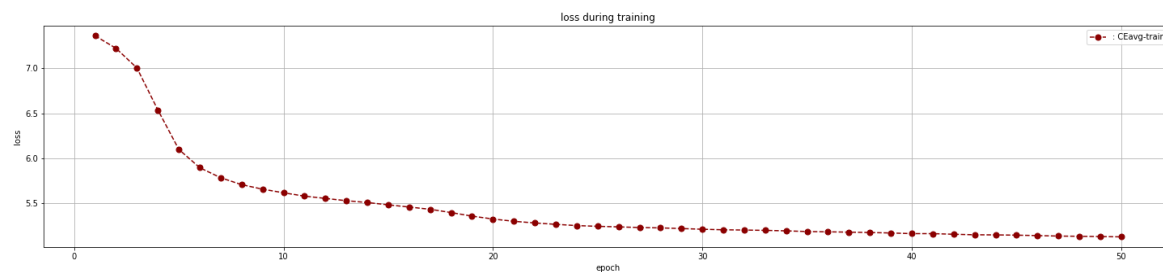
`model_book2 -> 'coming back over and my vamp a'`

`model_book3 -> 'your professional provide idea'`

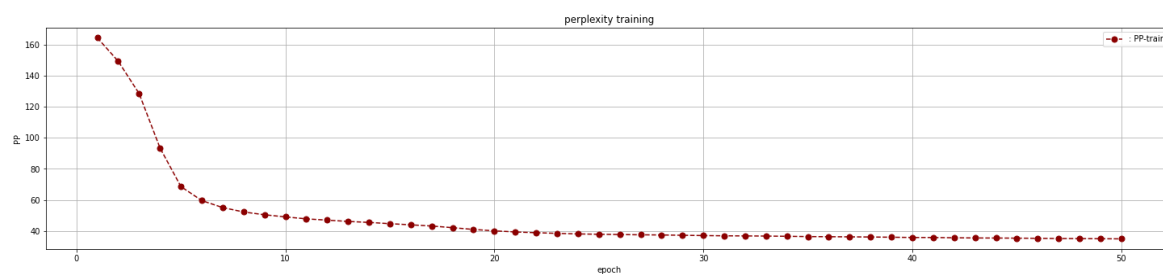
`model_book4 -> 'fact energy were a prison or solitary and of other feminine why you'`

از مقایسه جملات تولید شده، متوجه میشویم که سبک نوشتاری کتاب و ژانر آن در جملات تولید شده نیز موجود است. برای مثال جمله سوم و چهارم را مقایسه کنید.

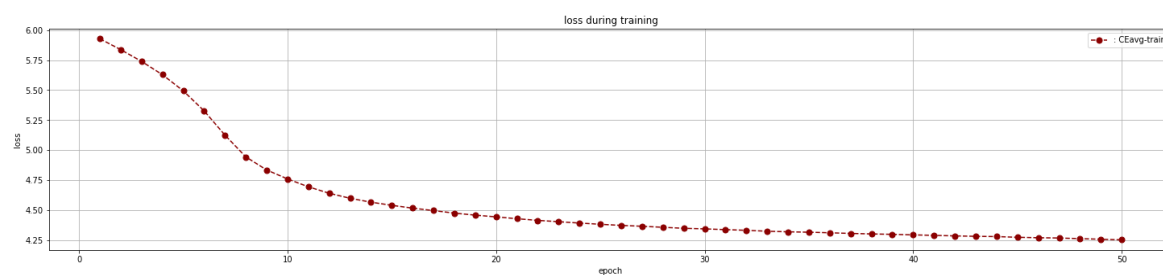
در ادامه نمودار های مربوط به مدل کتاب ۱ در شکل های ۱۱ و ۱۲، نمودار های مربوط به کتاب ۲ در شکل های ۱۳ و ۱۴ آورده شده است.



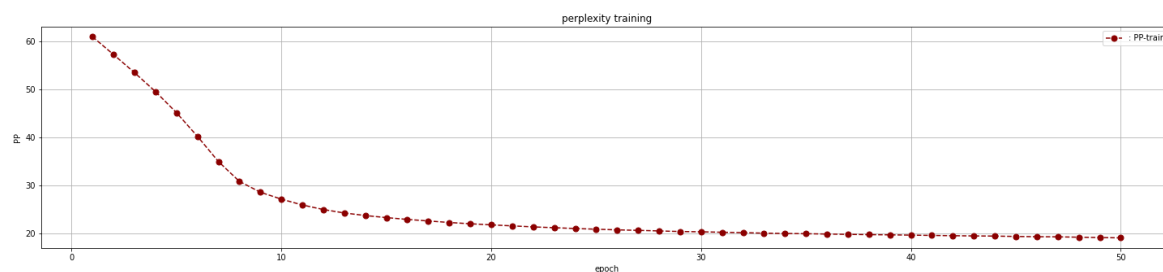
شکل ۱۱. مقدار هزینه در آموزش شبکه کتاب ۱



شکل ۱۲. مقدار سرگشتگی در آموزش شبکه کتاب ۱

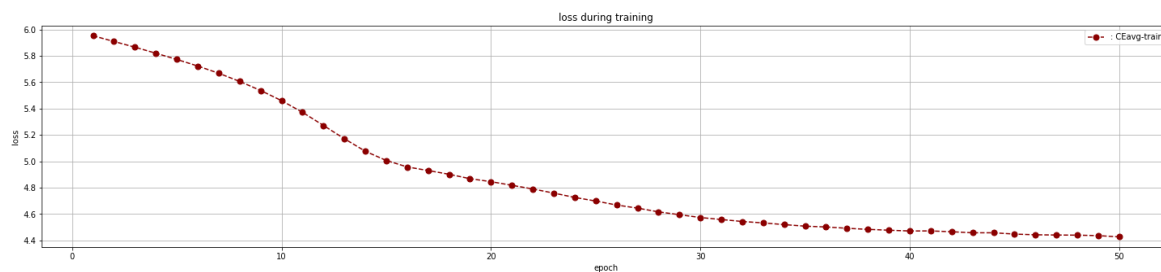


شکل ۱۳. مقدار هزینه در آموزش شبکه کتاب ۲

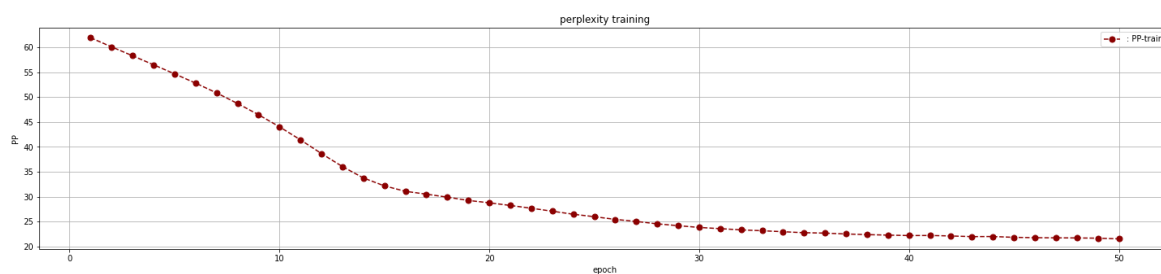


شکل ۱۴. مقدار سرگشتگی در آموزش شبکه کتاب ۲

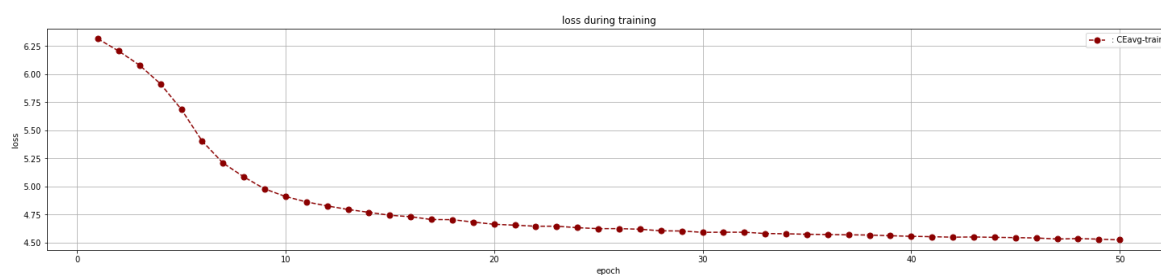
در ادامه نمودار های مربوط به مدل کتاب ۳ در شکل های ۱۵ و ۱۶، نمودار های مربوط به کتاب ۴ در شکل های ۱۷ و ۱۸ آورده شده است.



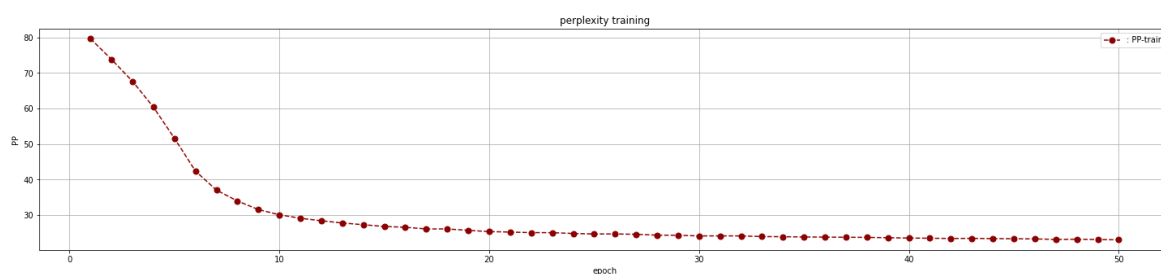
شکل ۱۵. مقدار هزینه در آموزش شبکه کتاب ۳



شکل ۱۶. مقدار سرگشتگی در آموزش شبکه کتاب ۳



شکل ۱۷. مقدار هزینه در آموزش شبکه کتاب ۴



شکل ۱۸. مقدار سرگشتگی در آموزش شبکه کتاب ۴

اجرای کد ها

پیاده سازی های مربوط به قسمت دست‌گرمی در فایل `NLP_CA3_FFNNLM.inpy` و پیاده سازی های مربوط به قسمت اصلی در فایل `NLP_CA3_LSTM.inpy` موجود است.