

# NLP – HW4

یاسمن لطف اللهی

## ۱. سوال تئوری

در RNN ها گرادیان وزن ها به صورت زیر محاسبه می شوند:

$$\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E_t}{\partial W}$$

گرادیان تمام استپ ها با هم جمع می شوند. اگر تمام گرادیان ها در حالت vanish و explode شدن باشند، گرادیان نهایی نیز این مشکل را دارد.

در LSTM، حاصل گیت forget نیز در محاسبه گرادیان ها تاثیر دارد. حتی اگر گرادیان در استپ  $k$ ، مشکل vanish یا explode شدن داشته باشد، شبکه می توان با تعیین پارامترهای مناسب، حاصل این گیت را طوری تعیین کند که گرادیان استپ  $k + 1$  دارای این مشکل نباشد.

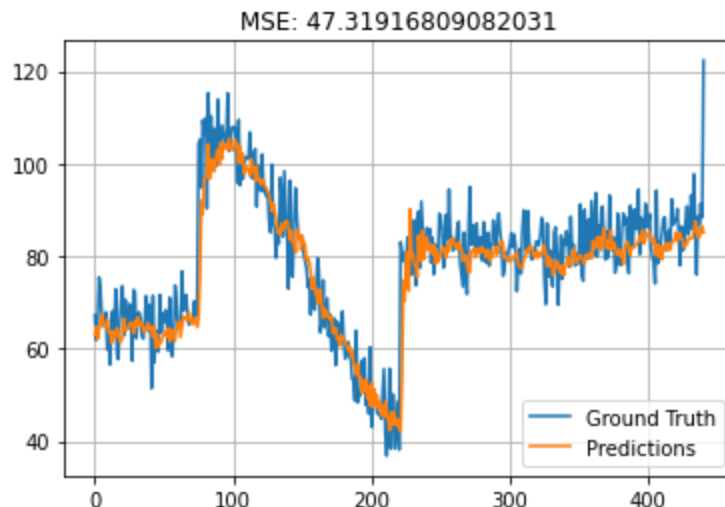
در نتیجه در شبکه LSTM، با وجود گیت forget کنترل بیشتری روی گرادیان ها و جلوگیری از vanish یا explode نشدن آن ها داریم.

## ۲. سوال عملی

این سوال در قالب یک نوت بوک پیاده سازی شده است. ابتدا به کمک تابع sliding window از لایبری numpy، پنجره های ۲۰ روزه از دیتاست استخراج می شود. این داده، داده ورودی شبکه است. داده خروجی، داده روز ۲۱ ام هر پنجره است که با اسلایس دیتاست بدست آمده است.

به خاطر کوچک بودن دیتاست، با یک شبکه کوچک و ساده این مسئله قابل حل است. معماری پیاده سازی شده، یک RNN با ۱۶ یونیت، یک لایه FC با ۸ یونیت و یک لایه FC با یک یونیت به عنوان لایه خروجی پیاده سازی شده است. تابع فعالسازی لایه ها، به جز لایه خروجی، ReLU تعیین شده. این مدل با MSE به عنوان تابع Loss و RMSProps به عنوان Optimizer کامپایل شده و سپس روی دیتاست ساخته شده، با سایز batch ۶۴ و در طول ۲۰۰ اپیاک آموزش می بیند.

مقدار MSE بدست آمده از حاصل پیش بینی روی داده تست، ۴۷/۳۱۹۲ است. مقدار پیش بینی و مقدار صحیح را در نمودار زیر می توان مشاهده کرد:



مدل نسبتاً خوب عمل کرده و توانسته روند کلی را یاد بگیرد. اما نتوانسته که نویز آن را یاد بگیرد.

### ۳. سوال عملی

قسمت های خواسته شده طبق توضیحات سوال پیاده سازی شده است.

• نتیجه بدون Attention : Train loss : 0.710 | Val loss : 1.247

• نتیجه با Attention : Train loss : 0.077 | Val loss : 0.312

واضح است که دیکودر به کمک Attention عملکرد بسیار بهتری دارد. آموزش نیز سریع تر است. برای مثال در ایپاک ۳۰، دیکودر با Attention توانسته که loss آموزش را تا ۰/۱۷۶ کاهش دهد. در صورتی که دیکودر عادی در پایان ایپاک ۱۰۰ نیز به این مقدار نمی رسد و در ایپاک ۳۰ این مقدار به ۰/۹۸۶ رسیده است.