یادگیری عمیق

گزارش مینی پروژه سوم

نام استاد : دکتر ستوده

نام دانشجو : حمزه قائدی

شماره دانشجویی: **9831419**

**هدف**: در این پروژه، با نوع خاصی از شبکهای عصبی به نام RNN آشنا میشویم. ویژگی اصلی این شبکه ها، حافظه دارد بودن انهاست. در شبکه های عصبی عادی، ورودی و خروجی هر نمونه مستقل از یکدیگر میباشد درحالی که در RNN ها خروجیهای هر مرحله به خروجی مراحل قبل وابسته است، به دلیل وجود این ویژگی، RNN ها، قادرند پترن های موجود در دنباله ها و سری های زمانی و.. را بهتر یاد بگیرند دراین پروژه به کمک RNN ها یک مدل برای پیش بینی تعداد لکه های خورشیدی بر اساس مشاهدات قبلی، توسعه میدهیم. با توجه به این که تعداد این لکه ها میتواند هر عددی باشد (البته اگر دوره ای بودن آن را در نظر نگیریم) این مسئله یک مسئله Regression است.

**ملاحظات کلی:**

به منظور مقایسه بهتر مدلهای مطرح شده، در تمامی مدلها از تنظیمات یکسانی که در زیر امده است استفاده شده است

Epochs = 50، تقریبا تمامی مدلهای ساخته شده در این پروژه در epoch =20 شروع به overfitting میکنند

Batch\_size = 128 ، اعداد مختلفی به عنوان اندازه دسته ها بررسی شده و از عدد 32 batch\_size= به بعد تفاوت قابل ملاحظه ای مشاهده نشد. اما بنا به احتیاط از عدد 128 استفاده شده است

**بهینه ساز**: در تمامی مدلهای مطرح شده از RMSProp به عنوان بهینه ساز استفاده شده است

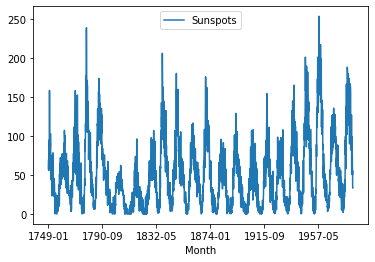
**تابع هزینه** : خروجی پیش بینی یک عدد پیوسته است لذا مسئله ما یک مسئله regression میباشد. تابع هزینه تمامی مدلها، حداقل قدر مطلق خطا (MAE) با رابطه زیر میباشد:

در رابطه فوق h(x) مقدار تخمین زده شده و Y مقدار واقعی نظیر ورودی X است و N نیز تعداد نمونه هاست

**\*\***) داده ها به صورت z-score نرمالیزه شده اند یعنی میانگین داده ها از تک تک داده ها کم شده و سپس حاصل بر واریانس داده ها تقسیم شده این امر منجر استقلال داده ها از یک واحد خاص شده و همچنین پراکندگی داده ها (یعنی وجود اعداد با تفاوت قابل ملاحظه در مجموعه) را کاهش میدهد که در نهایت منجر به افزایش پایداری و سرعت فرایند یادگیری میشود.

**سوال اول:**

نمودار داده ها (مربوط به قسمت c):



**سوال دوم:**

ساختار و تنظیمات تمامی مدلها یکسان بوده و تنها تفاوت آنها دل نوع لایه RNN است به عنوان مثال ساختار مدل SImpleRNN به صورت زیر است. هر دو مدل دیگر ساختاری مشابه دارند:

**Model: "sequential\_80"**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Layer (type) Output Shape Param #**

**=================================================================**

**simple\_rnn\_56 (SimpleRNN) (None, 16) 288**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**dense\_127 (Dense) (None, 1) 17**

**=================================================================**

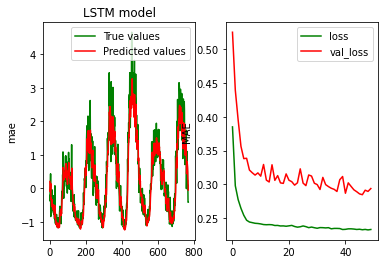
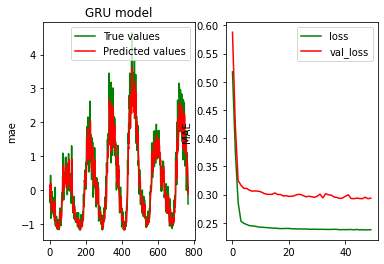
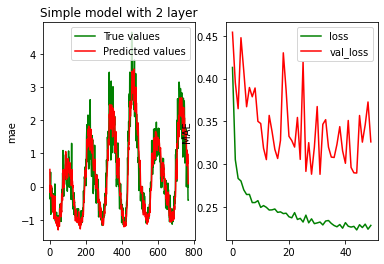
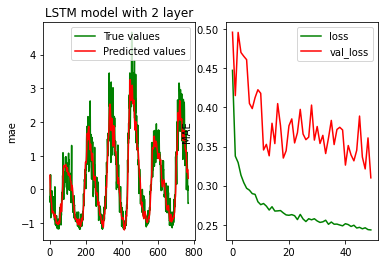
**Total params: 305**

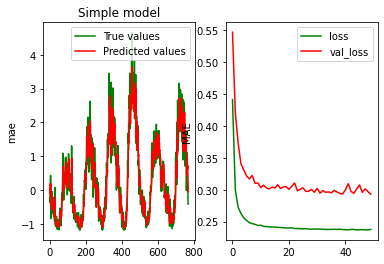
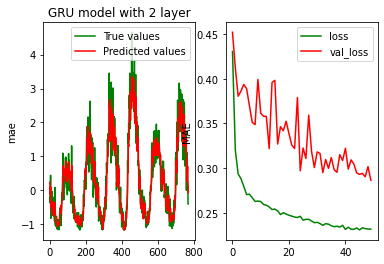
**Trainable params: 305**

**Non-trainable params: 0**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

در تصاویر زیر، نمودارهای مربوط به داده های واقعی و داده های تخمین زده شده و همچنین نمودار خطای آموزش و خطای validation قابل مشاهده است.





**سوال سوم**

**سوال دوم**

**مقایسه نتایج:**

با توجه نمودار داده ها در سوال اول، به نظر میرسد که پیدایش لکه های خورشیدی یک فرایند دوره ای[[1]](#footnote-1) است

نوسانات شدید موجود در نمودار میتواند به علت وجود نویز و یا ذاتا مربوط به خود پدیده باشد . لذا با بررسی پدیده و در صورتی که نویز عامل نوسانات شدید نمودار تشخیص داده شود، بهتر است پیش از آموزش دادن مدل، راه حلی به منظور کاهش نویز اتخاذ شود زیرا در غیر این صورت با افزایش پیچیدگی مدلها، این مدلها الگوهای نویز را یاد میگیرند. در هر صورت، RNN ها قادرند پترنهای دوره ای موجود در داده ها را یاد بگیرند و هرچه پیچیدگی مدلهای RNN ( تعداد لایه های استفاده شده) بیشتر باشند قابلیت یادگیری پترن های بیشتری را خواهند داشت

میانگین خطای تعمیم برای سه نوع شبکه RNN استفاده شده به شرح زیر است:

Simple RNN : val\_loss = 0.31

LSTM: val\_loss = 0.288

GRU : val\_loss = 0.29

با توجه به اینکه مسئله ما از نوع regression است، خطای تعمیم هرسه مدل قابل قبول است . اما مدلهای LSTM و GRU پیچدگی بیشتری نسبت به مدل ساده دارند درحالی که در این مسئله به ازای این افزایش پیچیدگی تفاوت قابل ملاحظه ای در کارایی (با معیار MAE) ایجاد نکرده اند. اما با توجه به نمودارهای مربوطه، به نظر میرسد که مدلهای LSTM و GRU بهتر نوسانات داده ها را یاد گرفته است که علت این امر نیز، بهبود مشکل محو شدگی گرادیان [[2]](#footnote-2)در دو مدل فوق نسبت به مدل ساده است.

در سوال سوم مدلها مطرح شده را با افزایش یک لایه به هرکدام پیچیده کرده ایم. دراین حالت نیز تفاوت معناداری بین معیار MAE این سه مدل وجود ندارد اما با توجه به نمودارهای رسم شده، به نظر میرسد مدل ساده و GRU هردو عملکرد بهتری نسبت به LSTM دارند. علت این امر میتواند overfitting سریعتر مدل LSTM به دلیل پیچیدگی بیشتر این مدل نسبت به دو مدل دیگر باشد.

1. periodic [↑](#footnote-ref-1)
2. Vanishing gradient problem [↑](#footnote-ref-2)