

یادگیری عمیق

گزارش مینی پروژه سوم

نام استاد: دکتر ستوده

نام دانشجو: حمزه قادی

شماره دانشجویی: 9831419

هدف: در این پروژه، با نوع خاصی از شبکه‌های عصبی به نام RNN آشنا می‌شویم. ویژگی اصلی این شبکه‌ها، حافظه دارد بودن آنهاست. در شبکه‌های عصبی عادی، ورودی و خروجی هر نمونه مستقل از یکدیگر می‌باشد در حالی که در RNN ها خروجی‌های هر مرحله به خروجی مراحل قبل وابسته است، به دلیل وجود این ویژگی، RNN ها، قادرند پترن‌های موجود در دنباله‌ها و سری‌های زمانی و... را بهتر یاد بگیرند در این پروژه به کمک RNN ها یک مدل برای پیش‌بینی تعداد لکه‌های خورشیدی بر اساس مشاهدات قبلی، توسعه می‌دهیم. با توجه به این که تعداد این لکه‌ها می‌تواند هر عددی باشد (البته اگر دوره‌ای 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 آمده است استفاده شده است

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

Batch_size = 128، اعداد مختلفی به عنوان اندازه دسته‌ها بررسی شده و از عدد batch_size = 32 به بعد تفاوت قابل ملاحظه‌ای مشاهده نشد. اما بنا به احتیاط از عدد 128 استفاده شده است

بهینه‌ساز: در تمامی مدل‌های مطرح شده از RMSProp به عنوان بهینه‌ساز استفاده شده است

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

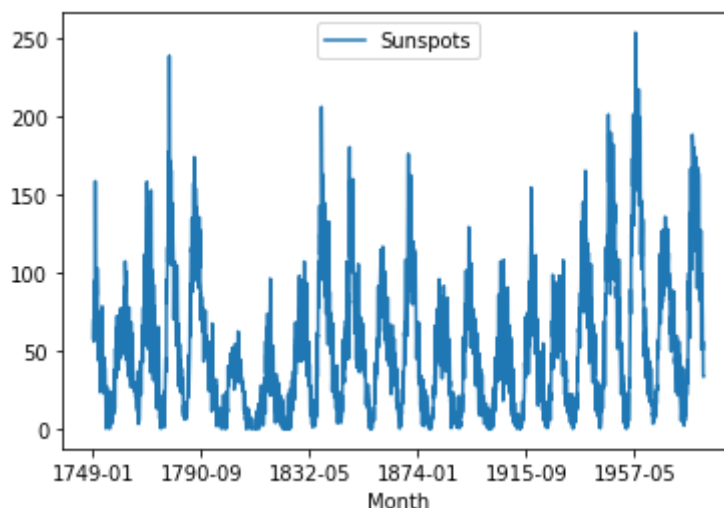
$$loss = \sum_{i=1}^N |h(X_i) - Y_i|$$

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

****** داده‌ها به صورت Z-score نرمالیزه شده‌اند. یعنی میانگین داده‌ها از تک‌تک داده‌ها کم شده و سپس حاصل بر واریانس داده‌ها تقسیم شده این امر منجر استقلال داده‌ها از یک واحد خاص شده و همچنین پراکندگی داده‌ها (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):

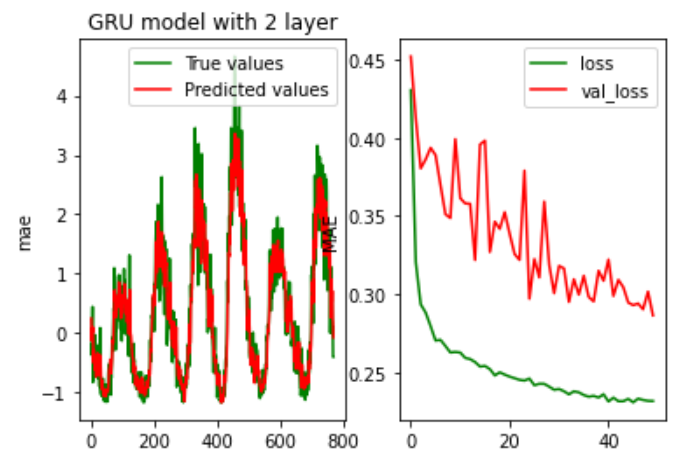
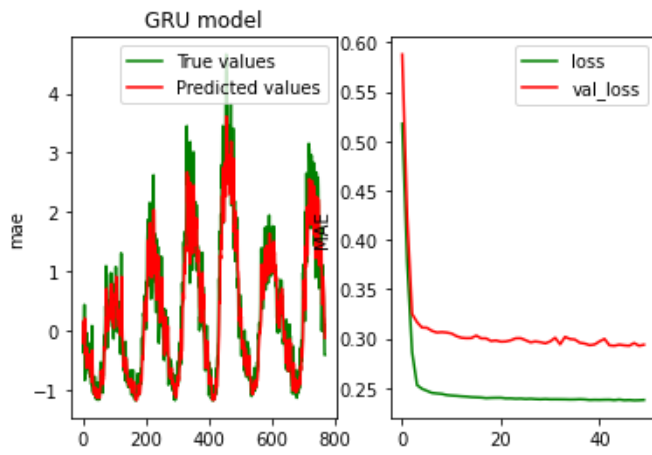
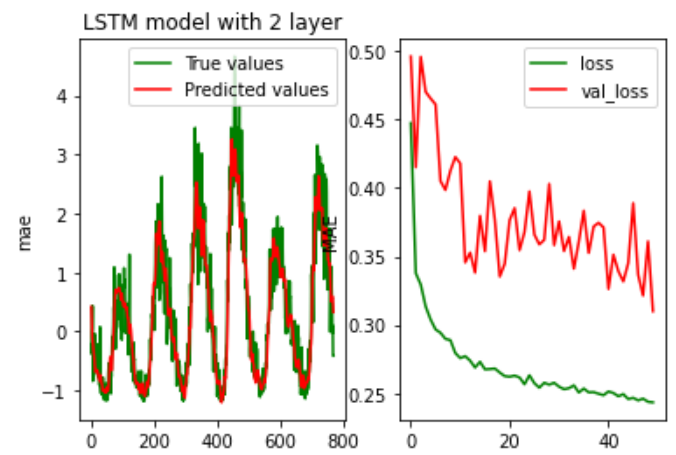
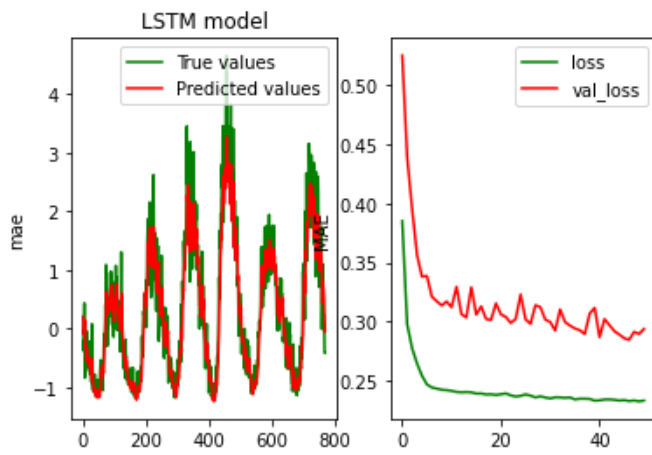
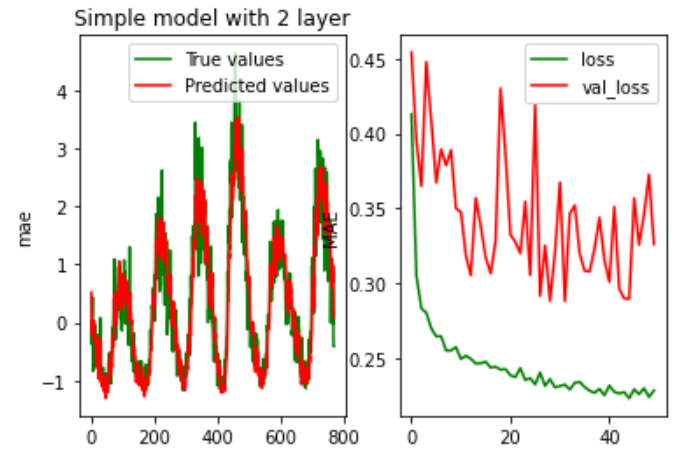
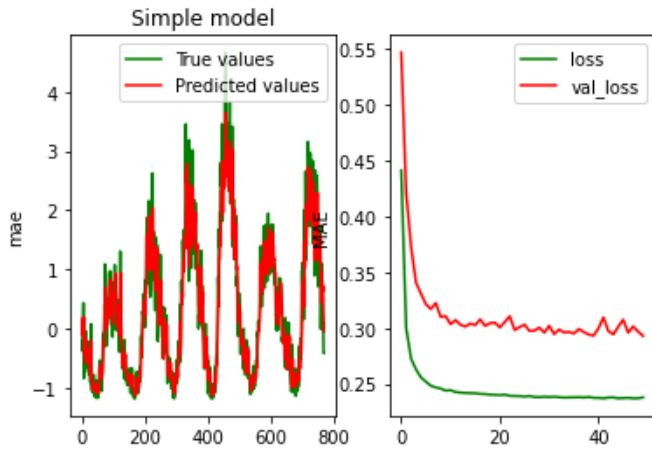


سوال دوم:

ساختار و تنظیمات تمامی مدلها یکسان بوده و تنها تفاوت آنها در 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 دوم

سوال سوم

با توجه نمودار داده ها در سوال اول، به نظر میرسد که پیدایش لکه های خورشیدی یک فرایند دوره ای^۱ 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 هر چه 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 هر سه مدل قابل قبول است. اما مدلهای LSTM و GRU پیچیدگی 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^۲ در دو مدل فوق نسبت به مدل ساده است.

در سوال سوم مدلها مطرح شده را با افزایش یک لایه به هر کدام پیچیده کرده ایم. در این 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 هر دو 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

² Vanishing gradient problem