



گزارش تمرین سری ۶

Yasmin Madani-97532265

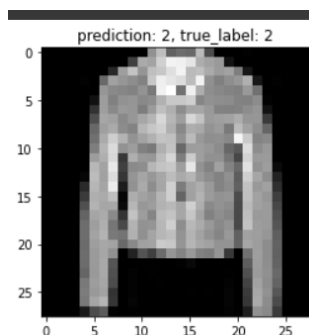


Fashion MNIST

این مجموعه داده ای از ۶۰۰۰۰ تصویر ۲۸*۲۸ در مقیاس خاکستری از ۱۰ دسته مد، همراه با مجموعه آزمایشی ۱۰۰۰۰ تصویر است. که کلاس های آن به شرح زیر است.

0	T-shirt/top
1	Trouser
2	Pullover
3	Dress
4	Coat
5	Sandal
6	Shirt
7	Sneaker
8	Bag
9	Ankle boot

برای مثال نمونه ای از اعضای این دیتاست را در تصویر زیر می بینیم.



توضیحات

با کمی تحقیق و مطالعه می توان دریافت که اگر تعداد کل داده ها به مقدار کافی بزرگ باشد یک پیشنهاد احتمالی در راستای تقسیم بندی داده های تست و آموزش و اعتبارسنجی به نسبت زیر خواهد بود.

TRAIN	VALIDATION	TEST
80%	10%	10%

همان طور که می دانیم نیاز است تا بیشتر داده های ما به عنوان داده آموزش به سیستم داده شود تا بتواند ویژگی های حداکثری را تخمین بزند اما داده های اعتبارسنجی اطلاعات مفیدی برای بهینه سازی فرآیندها ارائه دهند، که بر نحوه ارزیابی داده ها توسط مدل تأثیر می گذارد. اما اگر داده های اعتبارسنجی بیشتر از داده های آموزش باشند یا تعداد بیشتری از داده ها را به اعتبارسنجی اختصاص دهیم طبیعی است که از داده های آموزش کاسته شده و مدل به خوبی آموزش نمی بیند.

در این مورد مقدار فاکتور `validation_split` را برابر با ۰.۱۵ داده آموزش قرار دادیم که تقریباً ۹۰۰۰ داده را در اختیار ما قرار می دهد. که حدوداً با داده های تست برابری می کند که ۱۲ درصد از کل داده ها را شامل می شود.

برای آزمایش این موضوع ما مدل را سه بار با فاکتور `validation_split` متفاوت تست می کنیم و نتایج حاصل که مربوط به دقت و `loss` مرحله تست است، در جدول زیر قابل مشاهده است.

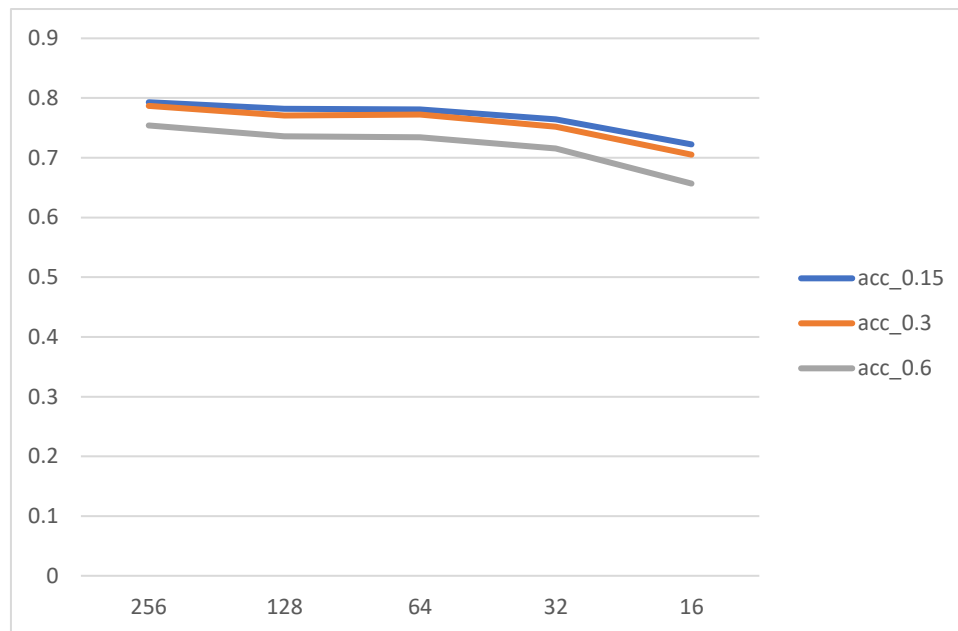
<code>validation_split</code>	<code>loss</code>	<code>accuracy</code>
۰/۱۵	0.6286	0.7930
۰/۳	0.6568	0.7869
۰/۶	0.7553	0.7541

همان طور که مشاهده شد دقت به دست آمده سخنان پیشین را تایید می کنند.

بدیهی است که مناسب نبودن داده های اعتبارسنجی می تواند موجب پدیده هایی مانند اورفیت یا آندرفیت شود. برای مثال اگر یک سیستم حل سوالات تستی ریاضی به جای یادگیری شروع به حفظ کردن پاسخ های موجود در صفحات یا پاسخ ها برحسب شماره سوال کند و اعتبار سنجی درستی از آن به عمل نیامده باشد در مرحله تست قادر به پاسخگویی نخواهد بود.

جدول زیر دقت فاز تست در سه حالت فوق با تعداد نورون های متفاوت است.

N-no	acc_0.15	acc_0.3	acc_0.6
256	0.793	0.7869	0.7541
128	0.7819	0.7708	0.736
64	0.781	0.7726	0.7341
32	0.7642	0.7521	0.7158
16	0.7224	0.7052	0.6568



تعداد نورون های به دست آمده در حالت ۲۵۶ نورون بهترین نتیجه را داشته است.

به طور کلی پیشنهاد می شود که

- تعداد نورون های پنهان بین اندازه لایه ورودی و اندازه لایه خروجی باشد.
- تعداد نورون های پنهان $2/3$ اندازه لایه ورودی به اضافه اندازه لایه خروجی باشد.
- تعداد نورون های پنهان کمتر از دو برابر اندازه لایه ورودی باشد.

که از شرایط فوق دو مورد در مورد ۲۵۶ صدق می کند و از طرفی مقدار ۲۵۶ به $2/3$ اندازه لایه ورودی نزدیکتر است.

تا اینجا بهترین دقت و نتیجه مختص به مدل با ۲۵۶ نورون میانی و ۹۰۰۰ داده اعتبار سنجی بوده است که برای این شرایط جدول زیر را داریم.

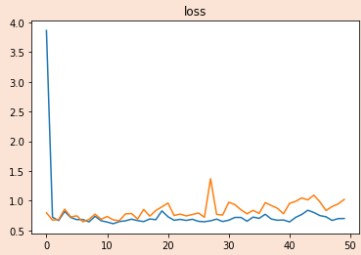
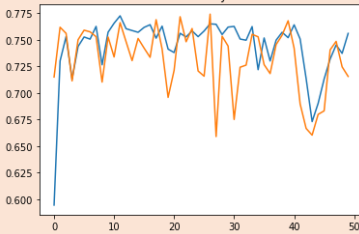
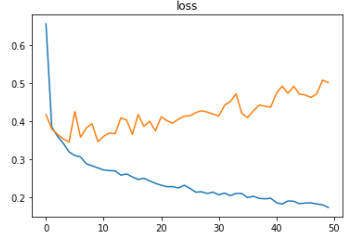
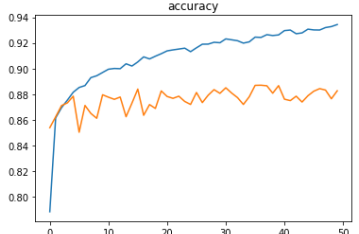
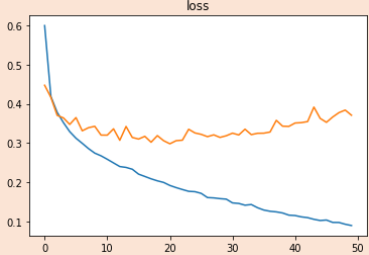
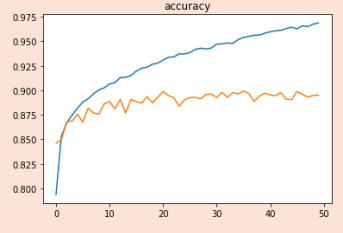
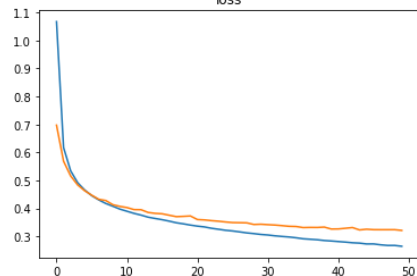
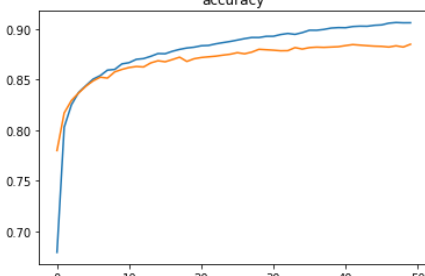
آموزش		تست	
accuracy	loss	accuracy	loss
0.8055	0.6094	0.7930	0.6286

طبیعی است که دقت فاز تست کمتر می شود و هدر رفت آن بیشتر چراکه بر روی داده هایی که تا به حال توسط سیستم دیده نشده اند، در حال تست هستیم و ممکن است داده هایی وجود داشته باشد که در داده های تست ما مشابه آن نبوده یا سیستم در مورد آن ها آموزشی ندیده است، اما این مقدار ناچیز است و بازده خوبی را شاهد هستیم.

نتایج به دست آمده برای ۴ optimizer به صورت زیر است. که با توجه به هدف ما کاهش loss یا accuracy بیشتر می تواند متفاوت باشد اما عامل مهم دیگری که در این بخش اهمیت دارد زمان اجرا و رسیدن به پاسخ است که در مجموع پارامترهای ذکر شده ADAM مورد مناسبی می تواند تلقی شود. چرا که در زمان کمتر به دقت قابل قبولی دست پیدا می کند.

	loss	acc
adam	0.4194	0.8739
sgd	0.6286	0.793
adagrad	0.606	0.799
RMSprop	0.451	0.875

با انتخاب Adam به آزمون تغییر نرخ یادگیری می پردازیم.

Learning rate	Loss	Accuracy
0.1		
0.01		
0.001		
0.0001		

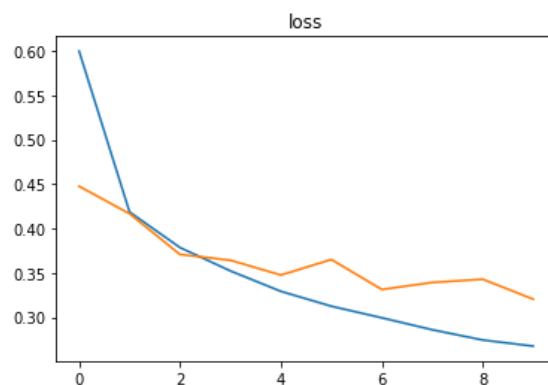
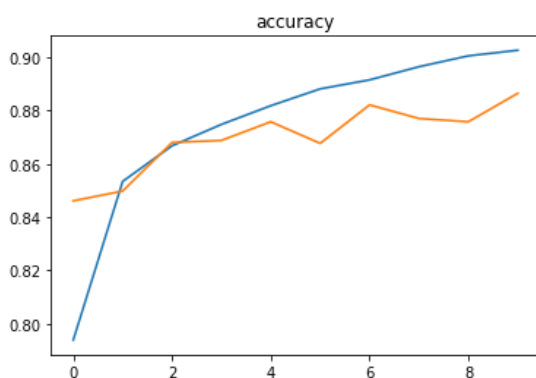
Learning rate	loss	accuracy
0.1	0.9331	0.7066
0.01	0.5338	0.8778
0.001	0.3964	0.8890
0.0001	0.3455	0.8766

همان طور که از دو جدول بالا می توان مشاهده کرد با نرخ 0.0001 نمودار ما هموار تر یکنواخت تر است و این مورد نسبت به نمودار هایی که دارای تغییرات ناگهانی اند مناسب تر است چرا که در نمودارهای غیر یکنواخت هرچند ممکن است سریع تر به مقدار نسبی از مینیم دست پیدا کنیم اما باید توجه کنیم که این مقدار می تواند مینیمم نسبی بوده باشد و مینیمم حقیقی در بخشی هایی است که مشاهده نشده اند.

با کاهش تعداد تکرار ها از ۵۰ به ۱۰ برای دو نرخ یادگیری 0.0001 و 0.0001 شاهد نتایج زیر هستیم .

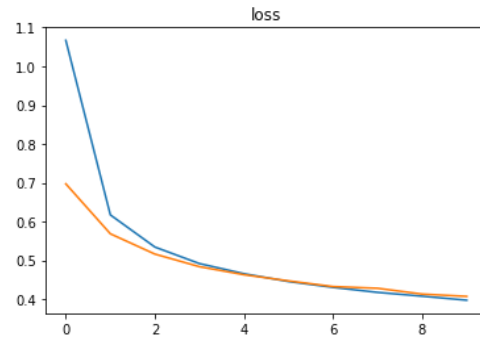
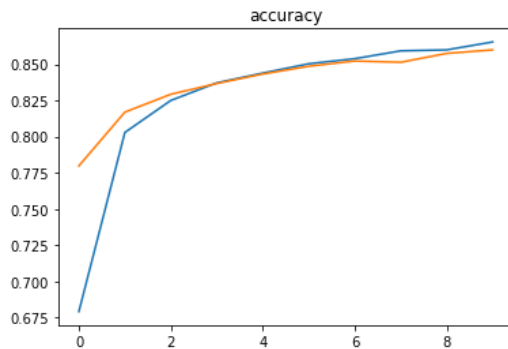
Learning rate=0.001

loss: 0.3420 - accuracy: 0.8771



Learning rate=0.0001

loss: 0.4308 - accuracy: 0.8503



تکرار ۱۰ بار برای سیستم با نرخ ۰.۰۰۱ به هم گرایی نمی انجامد اما با نرخ یادگیری ۰.۰۰۰۱ شاهد همگرا شدن خواهیم بود.