**سوال یک**

الف)

1. :Overfitting استفاده از نرخ یادگیری بسیار بالا می‌تواند منجر به بیش‌برازش مدل شود. که به داد های اموزشی دقت بالایی دارد ولی در داده های تست خوب عمل نمی کند.
2. عدم همگرایی: نرخ یادگیری بسیار بالا ممکن است باعث عدم همگرایی مدل شود. به عبارت دیگر، مدل ممکن است نتواند به طور کامل یاد بگیره و هی مدام در خال تغییر باشه.
3. پرش شدن از مینیمم محلی: با استفاده از نرخ یادگیری بالا، ممکن است مدل از مینیمم محلی خارج شود و به جای آن در جستجوی نقاط بهینه دیگر به سر ببرد. این می‌تواند باعث بهبود کیفیت مدل شود، اما همزمان ممکن است منجر به عدم پیدا کردن بهینه‌ی واقعی شود.

برای تشخیص مشکلات ناشی از استفاده از نرخ یادگیری بالا، می‌توانید به روش‌های زیر رجوع کنید:

1. نمودار تابع هزینه: بررسی تغییرات تابع هزینه در طول زمان آموزش می‌تواند نشان دهنده وجود مشکلات باشد. اگر تابع هزینه به سرعت کاهش یابد و سپس به طور ناگهانی افزایش یابد، احتمالاً نرخ یادگیری بیش از حد بالاست.

عملکرد در داده‌های تست: اگر رو داده های اموزشی دقت خوبی داریم ولی رو داده های تست نه ممکن یکی از دلایل نرخ یادگیری باشد.

تحلیل نمودار دقت: بررسی نمودار دقت مدل در طول زمان آموزش می‌تواند راهنمایی کننده باشد. اگر دقت در داده‌های آموزشی به سرعت افزایش یابد ولی دقت در داده‌های ارزیابی یا داده‌های تست به شدت کاهش یابد، نشانه‌ای از وجود مشکل با نرخ یادگیری بالا است.

ب)

استفاده از نرخ یادگیری بسیار پایین میتونه باعث underfitting بشه و هم چنین میتونه باعث مشکلات زیر نیز بشود:

1. عدم تطبیق کافی با داده‌های آموزشی: اگر نرخ یادگیری بسیار پایین باشد، مدل نمی‌تواند به طور کامل با داده‌های آموزشی تطبیق پیدا کند. این موجب می‌شود تا مدل به طور ناکافی آموزش ببیند و نتواند دقت قابل قبولی بر روی داده‌های آموزشی داشته باشد.
2. عملکرد ضعیف در داده‌های آموزشی و تست: استفاده از نرخ یادگیری بسیار پایین می‌تواند منجر به عملکرد ضعیف در داده‌های آموزشی و همچنین داده‌های تست شود. مدل نمی‌تواند به طور کافی یاد بگیرد و از توانایی تعمیم به داده‌های جدید کاسته می‌شود.
3. سرعت همگرایی بسیار پایین

برای تشخیص مشکلات ناشی از استفاده از نرخ یادگیری بسیار پایین، می‌توانید به روش‌های زیر رجوع کنید:

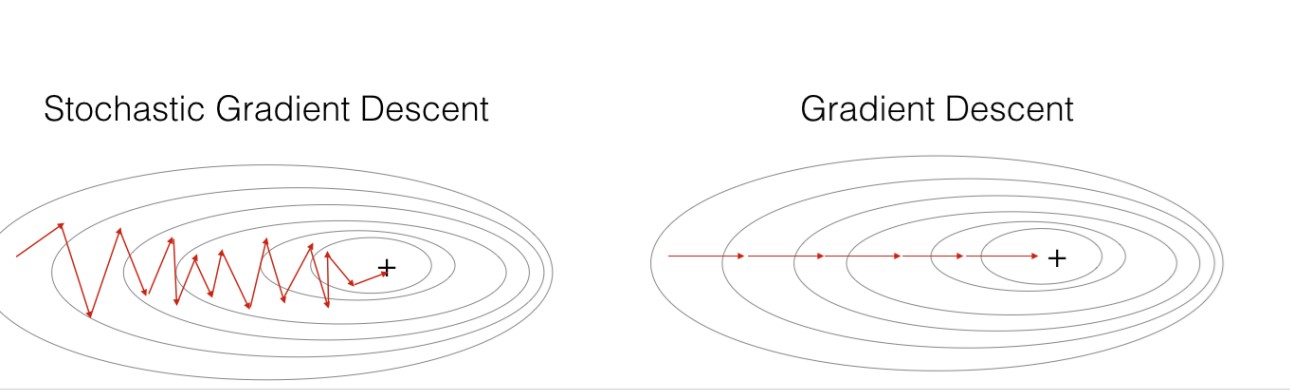
1. نمودار تابع هزینه: اگر تابع هزینه مقدارش در داده های اموزشی زیاد است وبه کندی کاهش میابد یکی از دلایلش میتواند نرخ یادگیری پایین باشد.
2. تحلیل نمودار دقت: بررسی نمودار دقت مدل در طول زمان آموزش می‌تواند مفید باشد. اگر دقت در داده‌های آموزشی به سرعت افزایش یافته و سپس به حالت ماندگاری رسید یا به حالتی کاهش یافته، ممکن است نرخ یادگیری بسیار پایین عاملی در این مشکل باشد.

پ)

نقطه زینی یا مینیمم محلی به معنای نقطه‌ای است که نسبت به نقط اطاف خود کمینه هست. در مسائل بهینه‌سازی، یافتن نقاط زینی می‌تواند مشکلاتی را ایجاد کند زیرا ممکن است با استفاده از الگوریتم بهینه‌سازی مناسب، به جای کمینه‌ی سراسری، در نقاط زینی محلی گیر کنیم و نتوانیم به جواب بهینه دست پیدا کنیم.

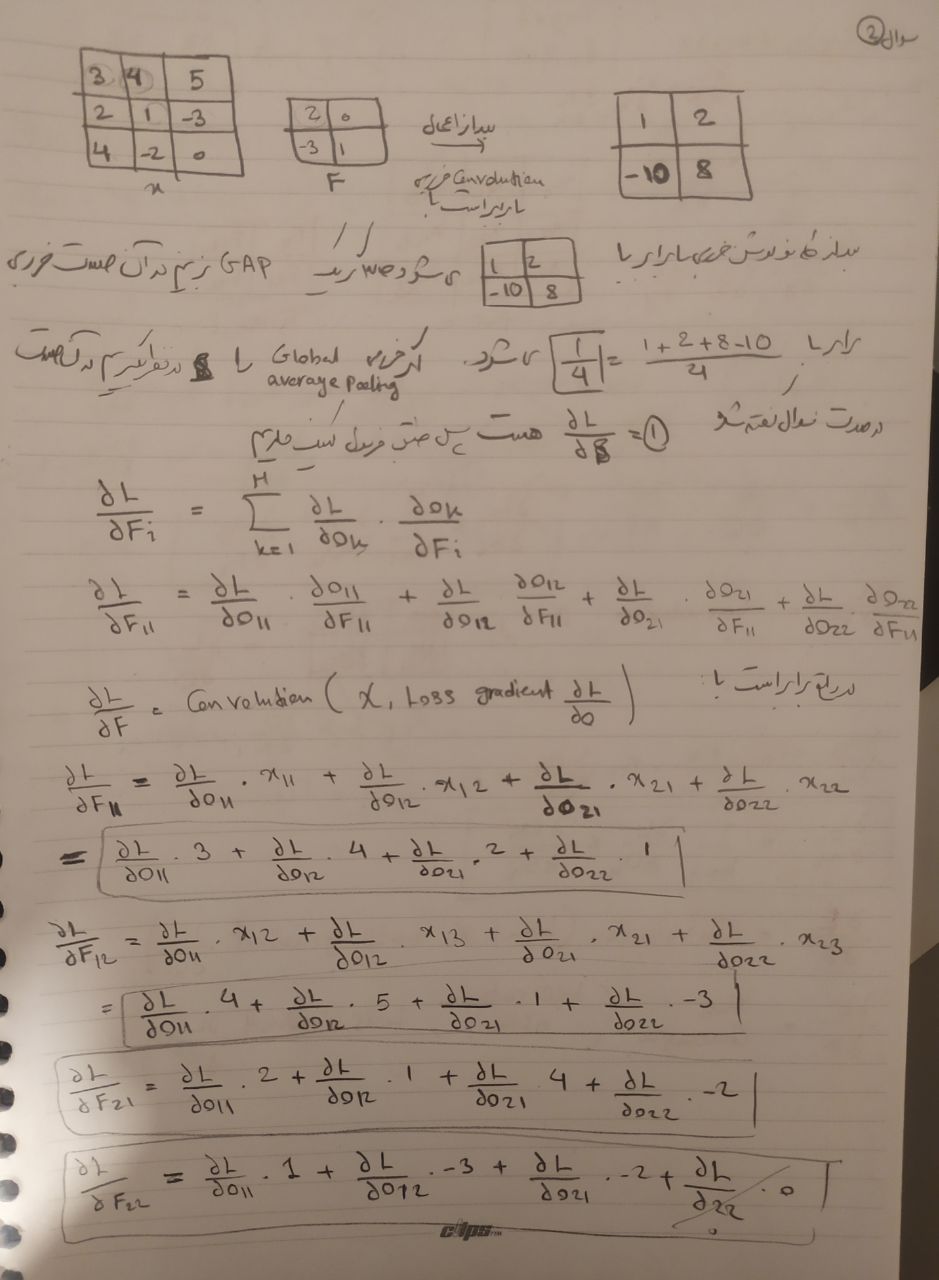
ت)

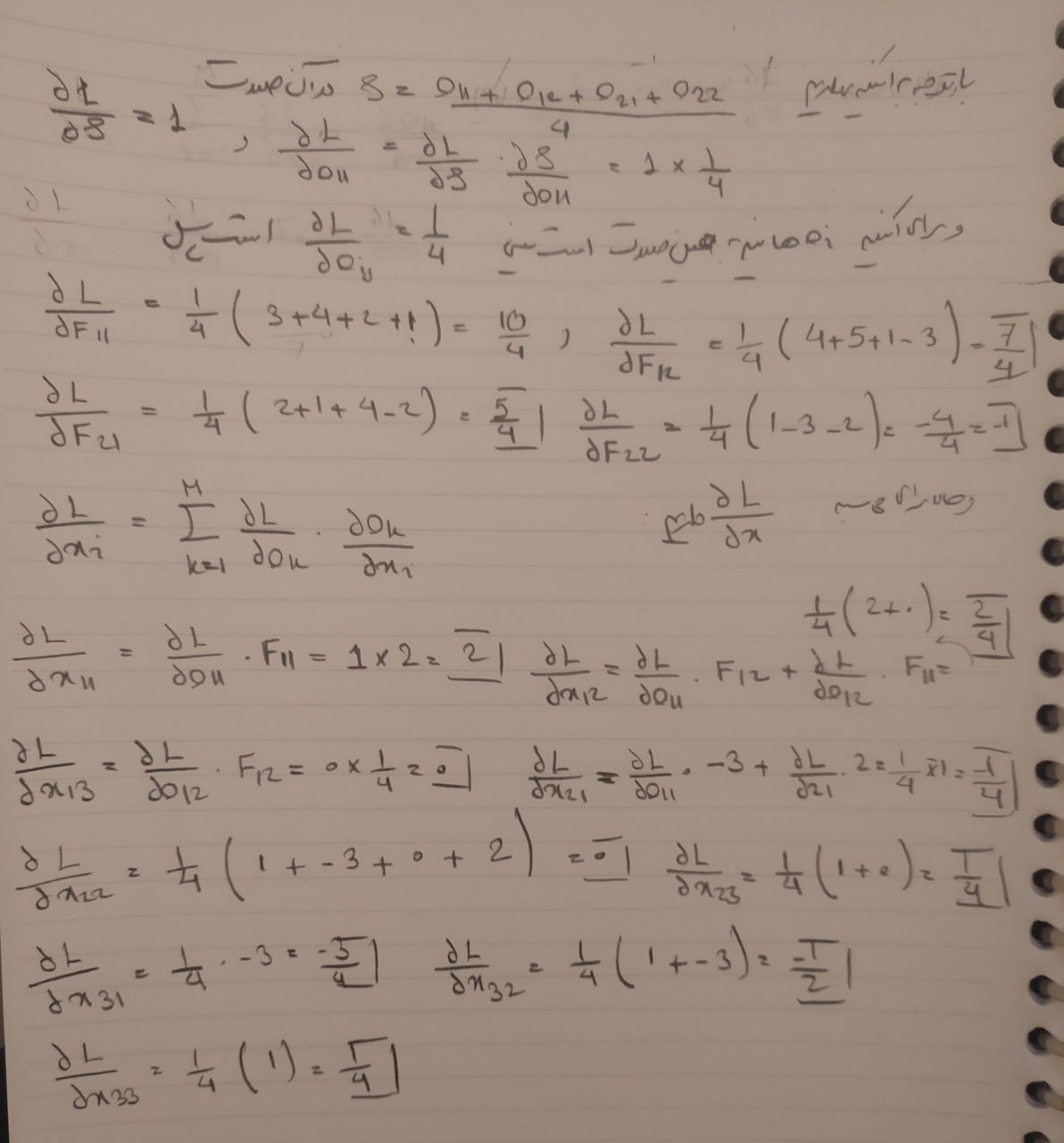
نمودار سمت چپ یک نمودار smooth هست و نمودار سمت راست یک نمودار نویزی هست وقتی از minibatch استفاده می کنیم نمودار تابع ضرر ممکن هست نویزی بشه ولی در نهایت در مسیر درستی پیش میریم به این دلیل نویزی میشه که دیتاست را به batch های ک.چک تقسیم میکنیم و برای هر کدام گرادیان حساب می کنیم که این نسبت به حالتی برای کل دیتاست گرادیان حساب میکنیم کمی باعث نویز میشود به همین علت نمودار سمت راست متعلق به minibatch هست.شکل پایین دید خوبی از این قضیه میدهد.



**سوال دو**

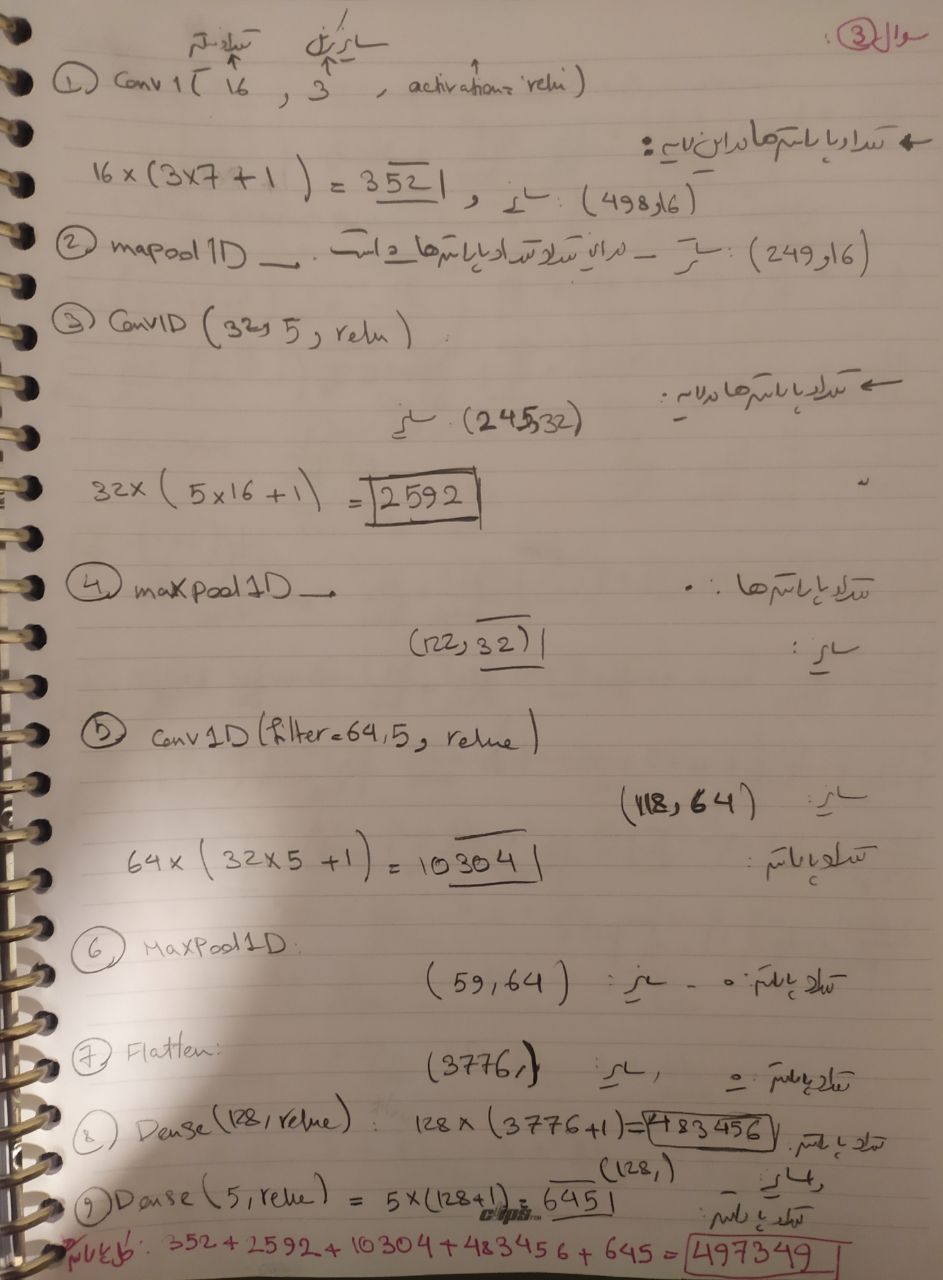
**پاسخ این سوال در صفحه بعد نوشته شده است.**

****

****

**سوال سه**

الف)

****ب)

لايه همگشتي دو بعدي و لايه همگشتي سه بعدي هر دو از لايه‌هاي همگشتي در شبکه‌هاي عصبي هستند، اما تفاوت اصلي بين آنها در ابعاد ورودي و عملکرد آنها است.

لايه همگشتي دو بعدي:

* ورودي و خروجي اين لايه داراي ابعاد دو بعدي است، به عنوان مثال (ارتفاع، عرض).
* اين لايه از فلترهاي دو بعدي براي استخراج ويژگي‌هاي مکاني از تصاوير استفاده مي‌کند.
* در برنامه‌هاي بينايي ماشين، معمولاً در شبکه‌هاي پيچشي (Convolutional Neural Networks) استفاده مي‌شود و به خوبي براي تشخيص الگوها و ويژگي‌هاي مکاني در تصاوير عمل مي‌کند.

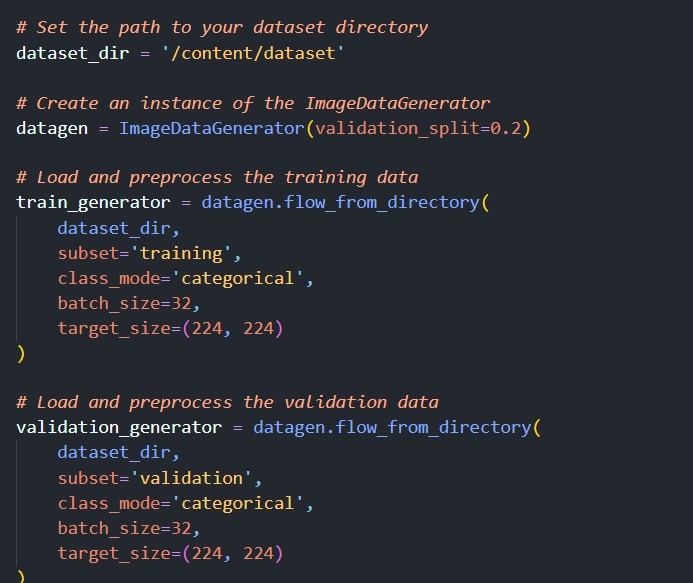
لايه همگشتي سه بعدي :

* ورودي و خروجي اين لايه داراي ابعاد سه بعدي است، به عنوان مثال (عمق، ارتفاع، عرض).
* اين لايه براي استخراج ويژگي‌هاي مکاني و زماني در داده‌هاي سه بعدي مانند ويديوها، سيگنال‌هاي صوتي و داده‌هاي مکعبي (volumetric data) استفاده مي‌شود.
* با استفاده از فلترهاي سه بعدي، داده‌هاي سه بعدي را در سه جهت (عمق، ارتفاع، عرض) همگشتي مي‌کند و ويژگي‌هاي مکاني و زماني را استخراج مي‌کند.
* در برنامه‌هاي مربوط به پردازش سيگنال‌هاي صوتي، تحليل و پردازش وديوها، تشخيص حرکت در ويديوها و پردازش داده‌هاي سه بعدي مانند تصاوير MRI و CT استفاده مي‌شود.

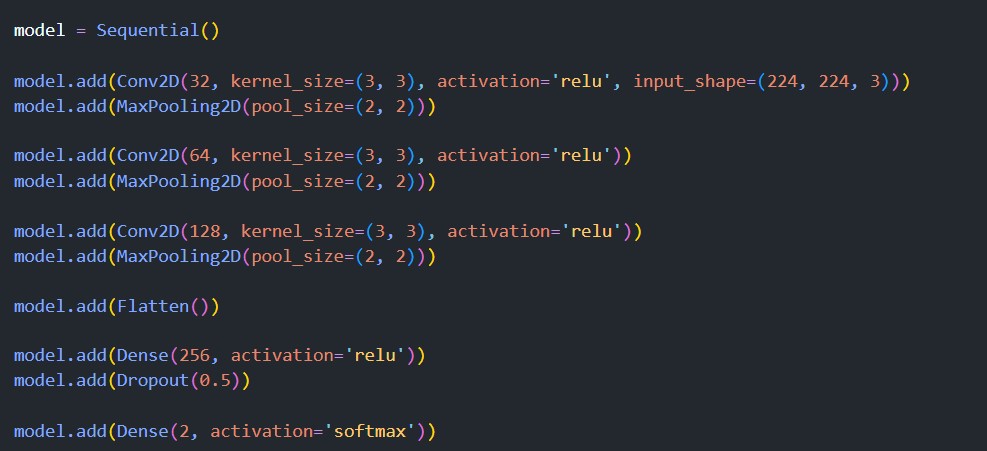
بنابراين، لايه همگشتي دو بعدي براي استخراج ويژگي‌هاي مکاني از تصاوير دو بعدي مناسب است، در حالي که لايه همگشتي سه بعدي براي استخراج ويژگي‌هاي مکاني و زماني از داده‌هاي سه بعدي مناسب است.

**سوال چهار**

در ابتدا با اسفاده از کد زیر train ,test جدا میکنیم و با استفاده از imagedataloader به صورت بچ بچ در می اوریم:



و سپس با استفاده از مدلی که به صورت زیر تعریف کردیم دیتاست را train میکنیم:

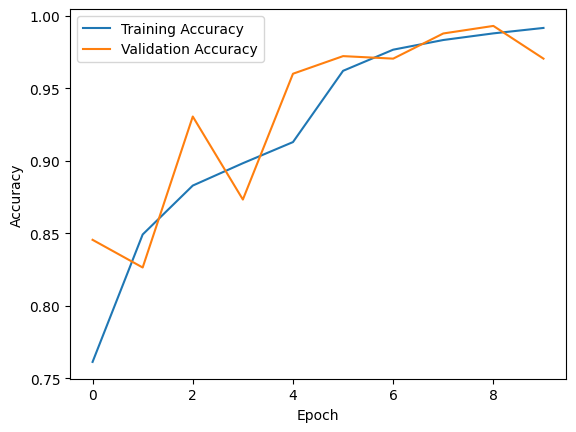
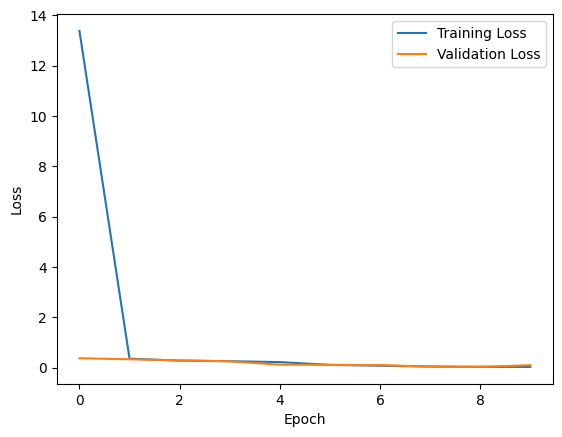


نتایج برای sequential بعد از پایان train رو داده های تست به صورت زیر شد:

دقت: 0.971666693687 0.9716666936874 0.971666693687439439

0.971666693687439

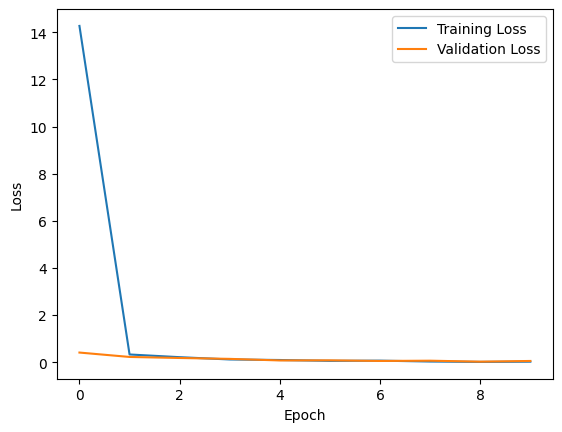
نمودار دقت و تابع ضرر:

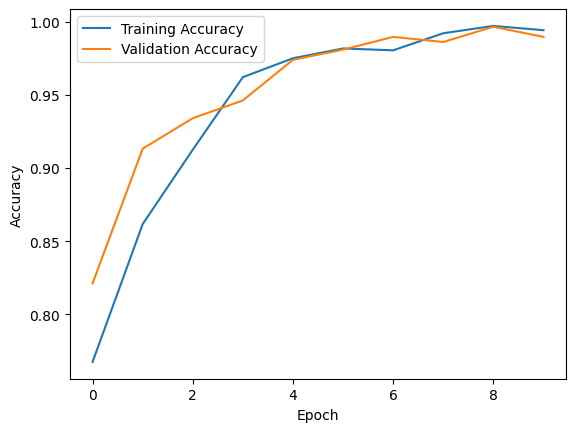


و سپس مدل را به صورت functional پیاده سازی کردیم که نتایج ان به صورت زیر شد:

دقت: 0.9900000095367432

و نمودار ها به صورت زیر شد:



****

تفاوت کمی که در نتایج sequential وfunctional هست این هست که ما داده ها را به صورت بچ بچ میدهیم وبه خاطر randomness کمی نتایج متفاوت است.

**سوال پنج**

فرض کنید می‌خواهیم یک مدل دسته‌بندی تصاویر بسازیم که توانایی تشخیص اشیاء در تصاویر را داشته باشد. برای این کار، می‌توانیم از لایه‌های همگشتی استفاده کنیم. این لایه‌ها به خاطر ویژگی‌های خاص خود، مزایایی در دسته‌بندی تصاویر ایجاد می‌کنند.

1. تشخیص ویژگی‌های محلی: لایه‌های همگشتی قادر به تشخیص ویژگی‌های محلی در تصاویر هستند. به عنوان مثال، لایه‌های اولیه می‌توانند لبه‌ها، خطوط و نقاط را تشخیص دهند. لایه‌های عمیق‌تر می‌توانند ویژگی‌های بیشتری را شناسایی کنند، مانند شکل‌ها، الگوها و اشیاء مخصوصا در مسائل ناحیه بندی این تشخیص ویژگی های محلی خیلی اهمیت داره.
2. مقاومت در برابر تغییرات محلی: لایه‌های همگشتی مقاومت قابل توجهی در برابر تغییرات محلی در تصاویر دارند. به عنوان مثال، اگر یک شیفت کوچک در تصویر اتفاق بیفتد، لایه‌های همگشتی همچنان قادر به تشخیص ویژگی‌ها و الگوها خواهند بود.
3. اشتراک پارامتر: لایه‌های همگشتی از اشتراک پارامتر بهره می‌برند. این به معنای استفاده مجدد از وزن‌ها در سراسر تصاویر است که به کاهش تعداد پارامترها و افزایش کارایی مدل منجر می‌شود.
4. کاهش تعداد پارامترها

به طور معکوس، یک حالت تصور کنید که ویژگی‌های لایه‌های همگشتی چالش‌ها ایجاد می‌کنند:

1. نیاز به محاسبات سنگین: لایه‌های همگشتی نیازمند محاسبات سنگین هستند، به ویژه در مواقعی که تصاویر بزرگ را پردازش می‌کنند. این موضوع می‌تواند باعث افزایش زمان آموزش و پیش‌بینی مدل شود.
2. نیاز به داده برچسب‌دار بزرگ: لایه‌های همگشتی برای آموزش به داده برچسب‌دار بزرگ نیاز دارند.
3. مشکل گرادیان ناپدید شونده: در شبکه‌های عصبی عمیق با لایه‌های هم‌گشتی، مشکل گرادیان ناپدید شونده همچنان وجود دارد. این مشکل به این معنی است که گرادیان‌ها در طول زمان به صورت نادرست به لایه‌همگشتی CNN منتقل می‌شوند. این موضوع می‌تواند باعث کاهش سرعت و کیفیت آموزش شبکه شود و ممکن است باعث بهبود ناچیز یا عدم پیشرفت در آموزش شبکه شود.

در توتبوک q5.ipynb من رو دیتاست mnist یکبار بدون CNN وبار دیگر با CNN دیتاست را train کردم نتایج بدون CNN دقت رو داده های تست برابر با 96.67 شد و با CNN برابر با 98.6 شد که دقت افزایش یافت البته mnist دیتاست پیچیده های نیست رو دیتاست های پیچیده تاثیر این CNN خیلی زیاد هست.

**سوال شش**

الف)

استفاده از فیلترهای ۱×۱ در شبکه‌های عصبی هم‌گشتی برای دو هدف اصلی استفاده می‌شود: کاهش تعداد نقشه‌های ویژگی در واقع همان کاهش عمق و حفظ ویژگی‌های مهم.

۱. کاهش تعداد نقشه‌های ویژگی: در شبکه‌های عصبی هم‌گشتی، تعداد نقشه‌های ویژگی مستقل افزایش می‌یابد هرچند که این موضوع می‌تواند باعث افزایش پیچیدگی محاسباتی و مصرف منابع شود. با استفاده از فیلترهای ۱×۱ می‌توان تعداد نقشه‌های ویژگی را کاهش داد. فیلتر ۱×۱ در واقع یک عملگر تک‌پیکسلی است که بر روی نقشه‌های ویژگی اعمال می‌شود. با ترکیب وزن‌های مختلف در فیلتر ۱×۱ با نقشه‌های ویژگی، می‌توان نقشه‌های ویژگی جدیدی تولید کرد که تعداد کانال‌ها را کاهش می‌دهد. این عمل باعث کاهش پیچیدگی محاسباتی و مصرف منابع شبکه می‌شود.

۲. حفظ ویژگی‌های مهم: استفاده از فیلترهای ۱×۱ به شبکه امکان می‌دهد که ویژگی‌های مهم را در نقشه‌های ویژگی حفظ کند. با ترکیب وزن‌های مختلف در فیلتر ۱×۱، شبکه قادر است ویژگی‌های مهم برای تشخیص الگوها و اشیاء را استخراج کند و در نقشه‌های ویژگی جدید حفظ کند. به عبارت دیگر، این فیلترها به شبکه امکان می‌دهند تا به طور همزمان اطلاعات مکانی و اطلاعات کانال‌های ویژگی را ترکیب کنند تا ویژگی‌های مهم را استخراج کنند و در نقشه‌های ویژگی کمتری نمایش دهند.

ب)

فیلتر ۱×۱ پس از اعمال، یک نقشه ویژگی جدید تولید می‌کند. این نقشه ویژگی اطلاعاتی را ارائه می‌دهد که ترکیبی از اطلاعات مکانی و اطلاعات کانال‌های ویژگی قبلی است.

هر پیکسل در نقشه ویژگی جدید، نمایانگر یک ترکیب خطی از ویژگی‌های موجود در نقشه‌های ویژگی قبلی است. وزن‌های مختلف در فیلتر ۱×۱ برای هر کانال ویژگی مشخص می‌کنند که چه میزان از ویژگی‌ها در نقشه ویژگی جدید نمایان شود. با ترکیب این وزن‌ها و ویژگی‌های موجود، نقشه ویژگی جدید به صورت خطی با ترکیبی از این ویژگی‌ها تولید می‌شود.

اطلاعات موجود در نقشه ویژگی جدید شامل ویژگی‌های مهمی است که توسط فیلتر ۱×۱ استخراج شده‌اند. این ویژگی‌ها می‌توانند الگوها، اشیاء، و یا ویژگی‌های مهم دیگری در داده‌ها باشند، که به شبکه امکان می‌دهند الگوها را تشخیص دهد و ویژگی‌های مورد نیاز را استخراج کند.

به طور خلاصه، نقشه ویژگی جدید که با استفاده از فیلتر ۱×۱ تولید می‌شود، اطلاعات ترکیبی از ویژگی‌های موجود در نقشه‌های ویژگی قبلی را ارائه می‌دهد و می‌تواند ویژگی‌های مهم و الگوهای موجود در داده را نمایش دهد.

پ)

فیلترهای n×n که n بزرگتر از ۱ است، نقشه ویژگی جدیدی با ابعاد مختلف تولید می‌شود و تفاوت‌های مهمی با فیلتر ۱×۱ دارد.وقتی از یک فیلتر n×n استفاده می‌کنیم، فیلتر بر روی نقشه ویژگی اصلی (ورودی شبکه) یا نقشه ویژگی قبلی اعمال می‌شود. این فیلتر با استفاده از یک پنجره مربعی به اندازه n×n، ویژگی‌های محلی را در نقشه ویژگی تشخیص می‌دهد و باعث ایجاد تغییرات مکانی در نقشه ویژگی می‌شود.تفاوت اصلی با فیلتر ۱×۱ این است که فیلتر n×n قادر به تشخیص الگوهای بزرگتر و پیچیده تر است. با افزایش اندازه فیلتر، محدوده مکانی که توسط فیلتر در نظر گرفته می‌شود بزرگتر می‌شود و اطلاعات مکانی گسترده‌تری را در نقشه ویژگی جدید نمایش می‌دهد.به طور کلی، فیلترهای n×n می‌توانند ویژگی‌های متنوعی را در نقشه ویژگی جدید تشخیص دهند، مانند لبه‌ها، نقاط کلیدی، الگوها و ساختارهای پیچیده‌تر. این فیلترها توانایی استخراج اطلاعات مکانی و ساختاری را بهبود می‌بخشند و به شبکه عصبی امکان می‌دهند برچسب‌ها و ویژگی‌های مهم را با دقت بیشتری تشخیص دهند.

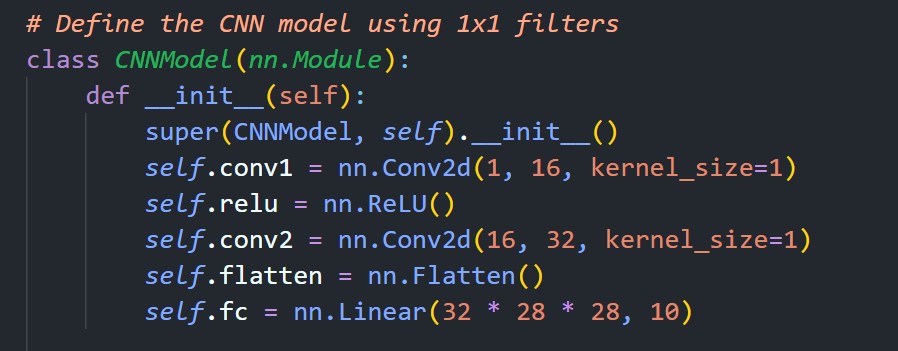
ت)

مدل‌هایی که از فیلترهای 1×1 استفاده می‌کنند، عبارتند از:

1. شبکه عصبی کانولوشنال :در شبکه‌های CNN، لایه‌های کانولوشنال 1×1 برای ترکیب ویژگی‌ها و کاهش ابعاد کانال‌ها (نیز به عنوان کانال داده می‌شود) استفاده می‌شود. این لایه‌ها می‌توانند به شبکه امکان بدهند تعداد کانال‌ها را کاهش دهد یا از ویژگی‌های غیرمهم جدا شده و فقط ویژگی‌های مهم را انتخاب کند.
2. شبکه عصبی (ResNet) معماری ResNet از بلوک‌های اصلی استفاده می‌کند که شامل لایه‌های کانولوشنال 1×1 و 3×3 است. لایه‌های کانولوشنال 1×1 در این بلوک‌ها برای کاهش ابعاد کانال‌ها و ترکیب ویژگی‌ها استفاده می‌شوند.
3. شبکه عصبی (Inception Network) معماری Inception نیز از فیلترهای 1×1 در لایه‌های کانولوشنال خود استفاده می‌کند. این فیلترها برای ترکیب ویژگی‌ها و کاهش ابعاد کانال‌ها در هر بلوک استفاده می‌شوند.

ج)

در نوتبوک q6.ipynb اومدم و یک CNN ساده با استفاده از فیلترهای 1 در 1 مطابق شکل زیر پیاده کردم:



بر اساس خروجی چاپ شده، ابعاد تصویر ورودی و خروجی برای مدل CNN با فیلترهای یک در یک به شرح زیر است:

* ابعاد پس از لایه Conv1: [1, 16, 28, 28]
* ابعاد پس از لایه Conv2: [1, 32, 28, 28]
* ابعاد پس از لایه FC: [1, 10]

همچنین، ابعاد تصویر ورودی اصلی از مجموعه داده MNIST به شکل زیر است:

* ابعاد تصویر ورودی: [1, 1, 28, 28]

با مقایسه ابعاد تصویر ورودی و خروجی، می‌توانیم تأثیر فیلترهای 1در 1 را بر روی تغییر ابعاد مشاهده کنیم. در این مدل، با استفاده از لایه Conv1 با 16 فیلتر 1 در 1، تعداد کانال‌ها از 1 به 16 تغییر می‌کند و ابعاد تصویر همان ابعاد ورودی (28x28) باقی می‌ماند. سپس، با استفاده از لایه Conv2 با 32 فیلتر 1 در 1، تعداد کانال‌ها از 16 به 32 تغییر می‌کند و ابعاد تصویر نیز بدون تغییر باقی می‌ماند. در نهایت، با استفاده از لایه FC، تصویر به یک بردار با ابعاد 10 تغییر شکل می‌یابد.

با مقایسه ابعاد ورودی و خروجی، می‌توانیم ببینیم که فیلترهای 1در 1 توانستند تعداد کانال‌ها را تغییر دهند ولی ابعاد تصویر را تغییر ندادند.

**سوال هفت**

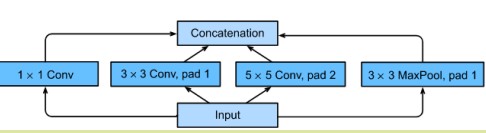
در ابتدا سوال ها را پاسخ میدهم و در ادامه پیاده سازی خودم را توضیح میدهم

الف)

هدف اصلی معماری Inception، استفاده بهینه از اطلاعات موجود در تصاویر است. برای این منظور، از ماژول Inception استفاده می‌شود که شامل لایه‌های پیچشی با اندازه‌های مختلف است. به این ترتیب، ماژول Inception قادر است الگوهای مختلفی را با اندازه‌های مختلف در تصاویر تشخیص دهد.

ساختار Inception شامل لایه‌های پیچشی ۱x۱، ۳x۳ و ۵x۵ است. لایه‌های پیچشی ۱x۱ با هدف کاهش تعداد کانال‌های ورودی، اطلاعات را ترکیب می‌کنند و تعداد پارامترها را کاهش می‌دهند. لایه‌های پیچشی ۳x۳ و ۵x۵، الگوهای مختلف را در تصاویر تشخیص می‌دهند. در نهایت، خروجی این لایه‌ها ترکیب شده و به لایه‌های بعدی ارسال می‌شود.

که شکل ای ماژول به صورت زیر است:



ب)

اندازه پارامتر گام (stride) در لایه‌های هم‌گشتی (convolutional layers) در شبکه‌های عصبی کانولوشنی تأثیر زیادی بر ابعاد فضایی نگاشت و ویژگی‌ها دارد. اندازه گام تعیین می‌کند که فیلترها چه فاصله‌ای از هم در طول و عرض تصویر حرکت کنند. در نتیجه، این پارامتر بر روی ابعاد فضایی نگاشت تأثیر می‌گذارد.

وقتی گام بزرگتر است، فیلترها با فاصله بیشتری حرکت می‌کنند و اندازه نگاشت کوچکتر می‌شود. به عبارت دیگر، با افزایش گام، تعداد پیکسل‌هایی که هر فیلتر را پوشش می‌دهد، کاهش می‌یابد. این می‌تواند منجر به کاهش دقت مدل در تشخیص ویژگی‌ها و اطلاعات مهم در تصویر شود. از طرف دیگر، استفاده از گام بزرگتر می‌تواند به کاهش پیچیدگی محاسباتی و افزایش سرعت آموزش و پیش‌بینی کمک کند.

در مقابل، با استفاده از گام کوچکتر، فیلترها با فاصله کمتری حرکت می‌کنند و اندازه نگاشت بزرگتر می‌شود. این باعث می‌شود که مدل بتواند ویژگی‌های جزئی‌تر را در تصویر تشخیص دهد و اطلاعات دقیق‌تری را در نگاشت حفظ کند. با این حال، استفاده از گام کوچکتر ممکن است منجر به افزایش پیچیدگی محاسباتی و کاهش سرعت آموزش و پیش‌بینی شود.

بنابراین، انتخاب مناسب برای اندازه گام بستگی به مسئله مورد نظر، اندازه تصاویر و پیچیدگی محاسباتی مورد نیاز دارد. برای مسائلی که نیاز به تشخیص ویژگی‌های دقیق و جزئی دارند، استفاده از گام کوچکتر مفید است. اما اگر سرعت محاسباتی یا سرعت آموزش و پیش‌بینی مهم باشد، ممکن است استفاده از گام بزرگتر مناسب باشد. همچنین، می‌توانید به صورت تجربی با اندازه‌های مختلف گام آزمایش کنید و اثر آن را روی دقت مدل موردنظر بررسی کنید.

ج)

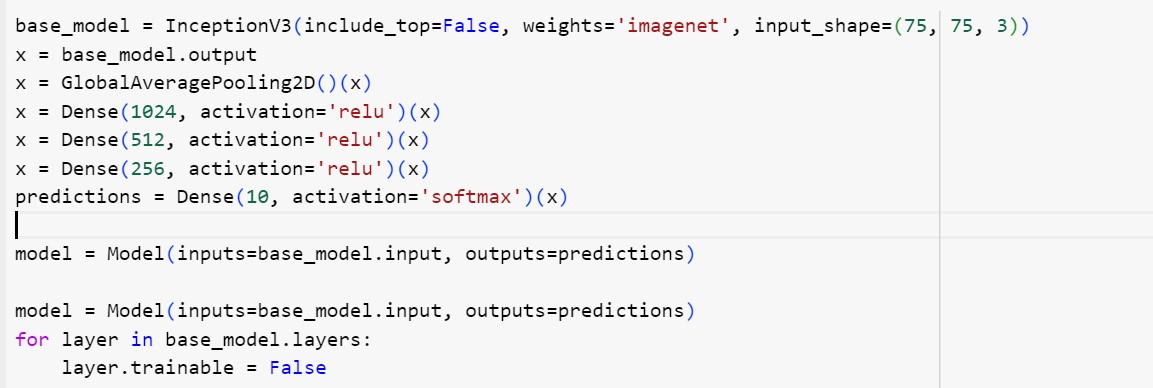
لایه‌های همگشتی در شبکه InceptionV3 از تکنیک پولینگ (pooling) استفاده می‌کنند. در کد شبکه ای که زدم، پس از لایه GlobalAveragePooling2D، لایه‌ی Dense قرار داده شده است. این لایه با میانگین‌گیری از ویژگی‌های استخراج شده از لایه‌های قبلی، ابعاد ویژگی‌ها را کاهش می‌دهد و ویژگی‌های مهم را استخراج می‌کند.

ویژگی‌های استخراج شده توسط لایه‌های همگشتی در شبکه اغلب ویژگی‌های محلی از تصویر هستند، مانند نقاط قابل توجه، الگوها و اشکال مختلف. این ویژگی‌ها معمولاً اطلاعات مهم و تمایزدهنده‌ای در مورد تصویر و شی‌های موجود در آن را ارائه می‌دهند. این ویژگی‌ها سپس توسط لایه‌های Dense به عنوان ورودی در نظر گرفته می‌شوند و در نتیجه، شبکه قادر خواهد بود دسته‌بندی صحیح را انجام دهد.

بنابراین، لایه‌های همگشتی در شبکه InceptionV3 وظیفه استخراج ویژگی‌های تصویر را دارند و با استفاده از فیلترها و ترکیب نتایج، ویژگی‌های مختلفی را از تصویر استخراج می‌کنند. این ویژگی‌ها سپس توسط لایه‌های Dense در شبکه برای دسته‌بندی استفاده می‌شوند.

توضیح مدل:

کد شبکه های که طراحی کردم همانطور که دربالا اشاره کردم به صور زیر است:



و تا اپک 23 شبکه رو train کردم. کد ها در نوتبوک q7 موجود است.