

«*In The Name Of GOD*»



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

[HW-02-Report]

[NEURAL COMPUTING AND DEEP LEARNING]

Hasan Masroor | [403131030] | April 4, 2025

"فهرست مطالب تمرین 02"

Question 1	2
1)	2
2)	9
Question 2	11
1)	11
2)	13
Question 3	15
1)	15
2)	18
3)	20

Problem 2: Multilayer perceptron neural networks

:Question 1 

1.

در این سوال ما قصد داریم که دو مدل شبکه عمیق و شبکه عریض را برای دسته‌بندی تصاویر 28×28 از لباس‌ها، کفش‌ها و ... که در 10 کلاس مختلف است آموزش دهیم و سپس عملکرد هر دو مدل را با استفاده از نمودارهای خطای آموزشی و اعتبارسنجی، ماتریس درهم‌ریختگی و معیارهای ارزیابی مثل دقت و ... مقایسه و بررسی کنیم.

در ابتدا کتابخانه‌های مورد نظر را می‌آوریم و سپس با استفاده از قطعه کد تمرین دیتاست FashionMNIST را دانلود می‌کنیم و اندازه مجموعه تست و آموزش را نمایش می‌دهیم:

```
Trainset size: 60000 samples  
Testset size: 10000 samples
```

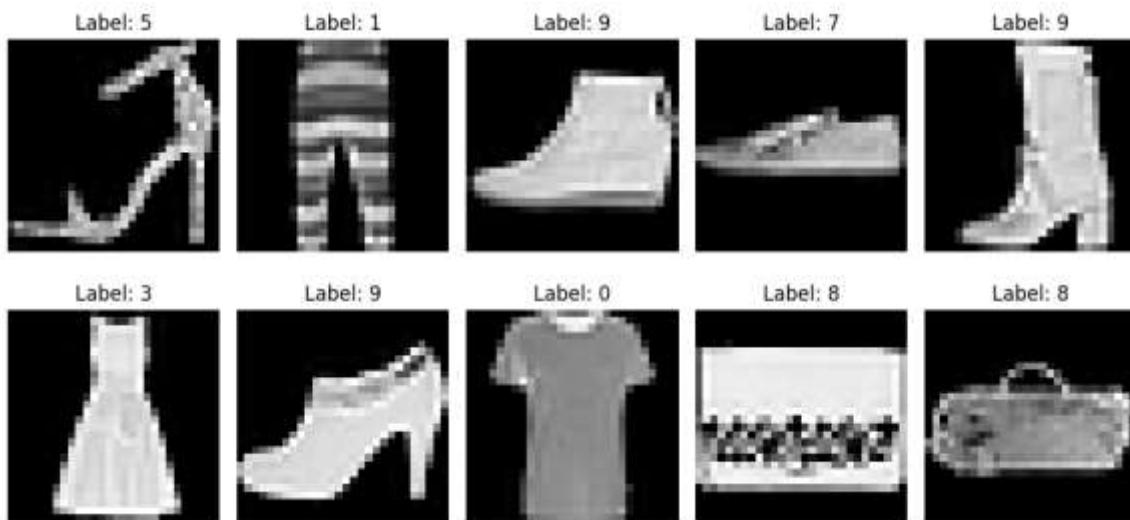
همانطور که می‌بینیم داده‌های آموزش حدود 85 درصد و داده‌های تست نیز حدود 15 درصد از مجموع کل داده‌ها هستند و حالا به نمونه‌های مجموعه تست دست نمی‌زنیم و سراغ مجموعه آموزش می‌رویم و از مجموعه آموزش، 12.5 درصد را برای اعتبارسنجی در نظر می‌گیریم و به تبع 87.5 درصد هم برای مجموعه آموزش می‌شود. برای این کار از `random_split()` استفاده می‌کنیم و در نهایت مجدد تعداد نمونه‌های این سه مجموعه را در خروجی نمایش می‌دهیم:

```
Trainset size: 52500 samples  
Validation set size: 7500 samples  
Testset size: 10000 samples
```

با این کار مجموعه آموزش شامل 75 درصد، اعتبارسنجی حدود 10 درصد و تست حدود 15 کل دیتاست را شامل می‌شوند. در مرحله بعد اسامی هر کدام از این 10 کلاس مختلف را نشان می‌دهیم و سپس `fashion_mnist_classes` را تعریف می‌کنیم که بعداً برای تجزیه و تحلیل بهتر ماتریس‌های در هم ریختگی از قسمت برای لیبل‌های 0 تا 9 استفاده کنیم و این 10 کلاس عبارت‌اند از:

```
[ 'T-shirt/top',
  'Trouser',
  'Pullover',
  'Dress',
  'Coat',
  'Sandal',
  'Shirt',
  'Sneaker',
  'Bag',
  'Ankle boot' ]
```

برای درک بهتر هم یک تابع تعریف می‌کنیم و به عنوان مثال 10 نمونه از داده‌های این کلاس‌ها را در خروجی نمایش دادیم:



بعد از اینکه کارهای اولیه و ضروری را برای این تمرین انجام دادیم حالا می‌خواهیم که مدل شبکه عمیق یا Deep Neural Network Model (DNN) را پیاده‌سازی کنیم. در این بخش یک شبکه عصبی عمیق 5 لایه‌ای با معماری Feed-Forward طراحی کردیم که شامل چهار لایه پنهان به ترتیب با ابعاد 512، 512، 256 و 128 نوروں و یک لایه خروجی با 10 نوروں که متناظر با 10 کلاس موجود در دیتابست می‌باشد. همچنین از تابع فعال‌سازی ReLU در لایه‌های پنهان استفاده کردیم که هم غیرخطی بودن به شبکه اضافه می‌کند و هم از مشکل محوشدنگی گرادیان جلوگیری نماید، برای بهبود عملکرد مدل نیز از Dropout با نرخ 0.3 برای دو لایه اول و نرخ 0.4 برای لایه سوم و چهارم برای کاهش overfitting استفاده کردیم. لایه‌های Batch Normalization نیز به معماری شبکه اضافه گردیده‌اند تا فرآیند آموزش را پایدارتر کنند. این معماری شبکه با تبدیل تصاویر

28×28 پیکسلی به بردارهای یک بعدی و عبور از لایه‌های متوالی، قادر به یادگیری ویژگی‌های سلسله‌مراتبی و انجام طبقه‌بندی با دقت مناسب می‌باشد.

در ادامه مدلی که تعریف کردیم را آموزش می‌دهیم و برای این کارتابع هزینه CrossEntropyLoss و بهینه‌ساز Adam با نرخ یادگیری 0.001 در نظر گرفتیم، همچنین فرآیند آموزش در 20 دوره تکرار انجام شد و پس از هر بار محاسبه خروجی مدل، خطای محاسبه شده و با استفاده از الگوریتم انتشار برگشتی، گرادیان‌ها به دست آمدند که منجر به بهروزرسانی پارامترهای مدل شد. هم‌زمان با پیشرفت آموزش، در پایان هر دوره عملکرد مدل روی داده‌های اعتبارسنجی را نیز بررسی کردیک که این کار به ما کمک کرد تا هم از یادگیری مناسب مدل مطمئن شویم و هم از بروز overfitting جلوگیری کنیم. در نهایت نیز خروجی آموزش مدل شبکه عمیق که به صورت زیر حاصل شد را در خروجی نمایش دادیم:

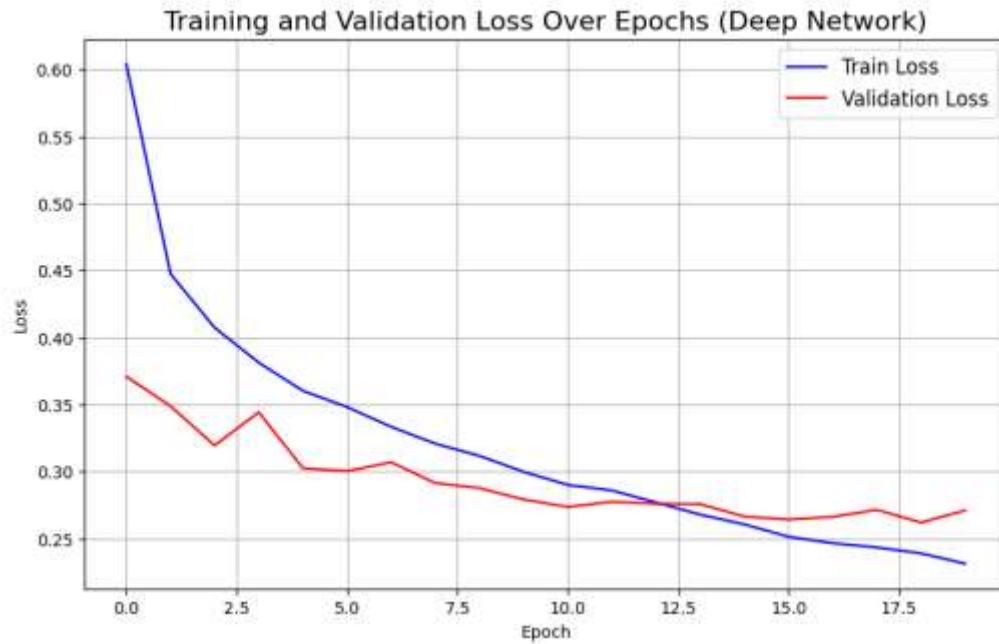
```
Epoch [1/20], Loss: 0.6040, Validation Loss: 0.3709
Epoch [2/20], Loss: 0.4475, Validation Loss: 0.3488
Epoch [3/20], Loss: 0.4073, Validation Loss: 0.3194
Epoch [4/20], Loss: 0.3813, Validation Loss: 0.3442
Epoch [5/20], Loss: 0.3602, Validation Loss: 0.3024
Epoch [6/20], Loss: 0.3481, Validation Loss: 0.3003
Epoch [7/20], Loss: 0.3333, Validation Loss: 0.3069
Epoch [8/20], Loss: 0.3208, Validation Loss: 0.2913
Epoch [9/20], Loss: 0.3116, Validation Loss: 0.2876
Epoch [10/20], Loss: 0.2998, Validation Loss: 0.2791
Epoch [11/20], Loss: 0.2900, Validation Loss: 0.2736
Epoch [12/20], Loss: 0.2859, Validation Loss: 0.2774
Epoch [13/20], Loss: 0.2769, Validation Loss: 0.2759
Epoch [14/20], Loss: 0.2678, Validation Loss: 0.2757
Epoch [15/20], Loss: 0.2605, Validation Loss: 0.2664
Epoch [16/20], Loss: 0.2513, Validation Loss: 0.2641
Epoch [17/20], Loss: 0.2465, Validation Loss: 0.2662
Epoch [18/20], Loss: 0.2433, Validation Loss: 0.2714
Epoch [19/20], Loss: 0.2389, Validation Loss: 0.2619
Epoch [20/20], Loss: 0.2313, Validation Loss: 0.2709
```

همانطور در تصویر بالا هم مشخص است در ابتدای آموزش یا Epoch 1 خطای آموزش و اعتبارسنجی به ترتیب 0.6040 و 0.3709 بود که با پیشرفت فرآیند آموزش، هر دو معیار به صورت پایدار کاهش یافته‌ند و در انتهای آموزش یا Epoch 20 نیز خطای آموزش به 0.2313 و خطای اعتبارسنجی به 0.2709 رسید که نشان‌دهنده بهبود تدریجی و همگرایی مناسب مدل است و در صورت ادامه با دوره‌ای بیشتر ممکن است شاهد بهبود بیشتر مدل خواهیم بود و اینکه مدل به سمت همگرایی پیش خواهد رفت.

دقت نهایی مدل شبکه عمیق را نیز در زیر مشاهده می‌کنیم:

Test Accuracy: 90.04%

در ادامه برای اینکه بهتر بتوانیم این خطای آموزش و اعتبارسنجی را تجزیه و تحلیل کنیم و همچنین همگرایی در این مدل را بررسی کنیم نمودار مربوطه را رسم می‌کنیم:



همان‌طور که از تصویر هم مشخص است به مرور و با افزایش Epoch مقدار خطا یا Loss هم برای آموزشی و هم اعتبارسنجی کاهش پیدا می‌کند و نمودار شبیه نزولی دارد و یعنی به مرور خطا کمتر می‌شود و مدل یاد می‌گیرد که بهتر عمل کند و دسته‌بندی درست و بهتری برای این دسته‌بندی داشته باشد (البته خطای اعتبارسنجی دارای صعود و نزول‌های مختلفی است). همچنین از روی نمودار می‌بینیم که هنوز مدل همگرا نشده است و احتمالاً بعد از چند epoch، در نقطه‌ای مقدار خطا تقریباً ثابت می‌شود و دیگر کاهش محسوسی ندارد و می‌توانیم روند را متوقف کنیم و دیگر بیشتر ادامه ندهیم.

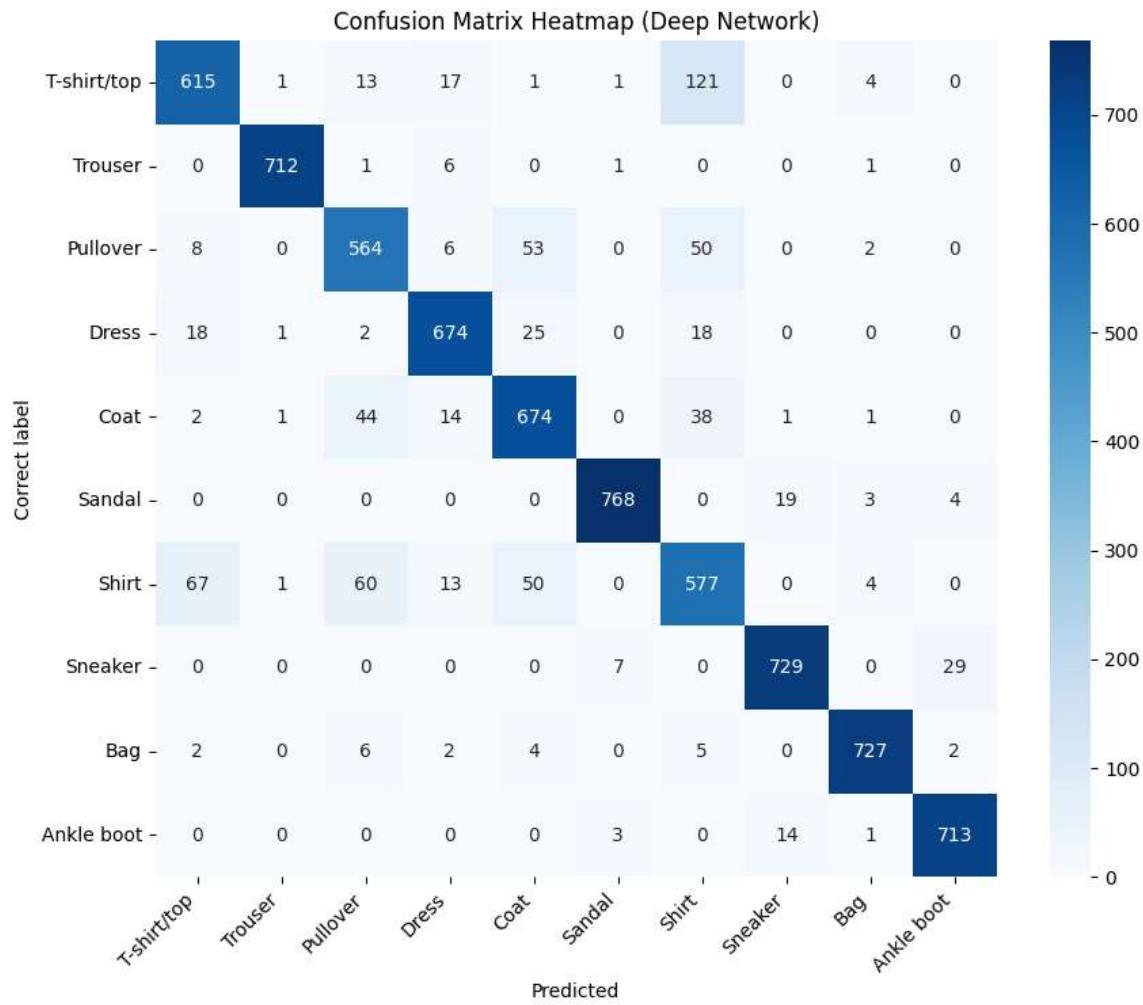
حالا می‌آییم برای ارزیابی دقیق‌تر عملکرد مدل، ماتریس درهم‌ریختگی را رسم می‌کنیم که نمایش‌دهنده تعداد پیش‌بینی‌های صحیح و غلط برای هر کلاس است و ستون لیبل‌های درست ما را نشان می‌دهد و ردیف نیز پیش‌بینی‌های ما از این کلاس‌ها هستند. همان‌طور که در تمرین قبل هم اشاره کردیم می‌توانیم از روی این ماتریس درهم‌ریختگی recall، precision، accuracy را حساب کنیم:

$$\square \text{ Accuracy} = \frac{\#TP + \#TN}{\#P + \#N} = \frac{\#TP + \#TN}{\#TP + \#FN + \#TN + \#FP}$$

$$\square \text{ Precision} = \frac{TP}{TP + FP} \quad \text{و} \quad TPR = \frac{TP}{TP + FN} = \text{Sensitivity} = \text{Recall}$$

به عنوان مثال در ماتریس زیر مشاهده می‌کنیم که 615 نمونه را به درستی پیش‌بینی کردیم که جزو کلاس T-SHIRT/TOP می‌باشند و به همین ترتیب مواردی که روی قطر اصلی ماتریس قرار دارند همان TP های ما هستند؟

مثل همین مثالی که زدیم می‌تواینم موارد بالا را برای هر کلاس بررسی کنیم اما موردنی که از ماتریس مشهود است این است که مدل توانسته تقریباً پیش‌بینی‌های درست و قابل قبولی برای هر کلاس داشته باشد:



در این بخش، مدل شبکه عصبی عریض یا (WNN) Wide Neural Network Model را پیاده‌سازی کردیم. این مدل شامل دو لایه اصلی است، یک لایه پنهان با 1024 نورون و یک لایه خروجی با 10 نورون که به 10 کلاس موجود در دیتاست مربوط می‌شود. لایه پنهان از تابع فعال‌سازی ReLU برای افزودن غیرخطی بودن به شبکه استفاده می‌کند و از Dropout با نرخ 0.3 به منظور کاهش احتمال overfitting استفاده کردیم. لایه‌های نیز به معماری شبکه اضافه گردیده‌اند تا فرآیند آموزش را پایدارتر کنند.

در ادامه مدلی که تعریف کردیم را آموزش می‌دهیم و برای این کار از تابع هزینه CrossEntropyLoss و بهینه‌ساز Adam با نرخ یادگیری 0.001 استفاده کردیم. فرآیند آموزش در 20 دوره تکرار انجام می‌شود و در هر دوره ابتدا مدل در فاز آموزش قرار می‌گیرد. پس از محاسبه خروجی مدل، خطای محسوب شده و با استفاده از الگوریتم Backpropagation گرادیان‌ها به دست می‌آیند و پارامترهای مدل به روزرسانی می‌شوند؛ همچنین پس از هر

دوره، مدل در فاز اعتبارسنجی قرار می‌گیرد و عملکرد آن روی داده‌های اعتبارسنجی بررسی می‌شود. این مرحله به ما کمک می‌کند تا اطمینان حاصل کنیم که مدل در حال یادگیری بهینه است و از بروز overfitting جلوگیری کنیم. در نهایت نیز خروجی آموزش مدل شبکه عریض که به صورت زیر حاصل شد را در خروجی نمایش دادیم:

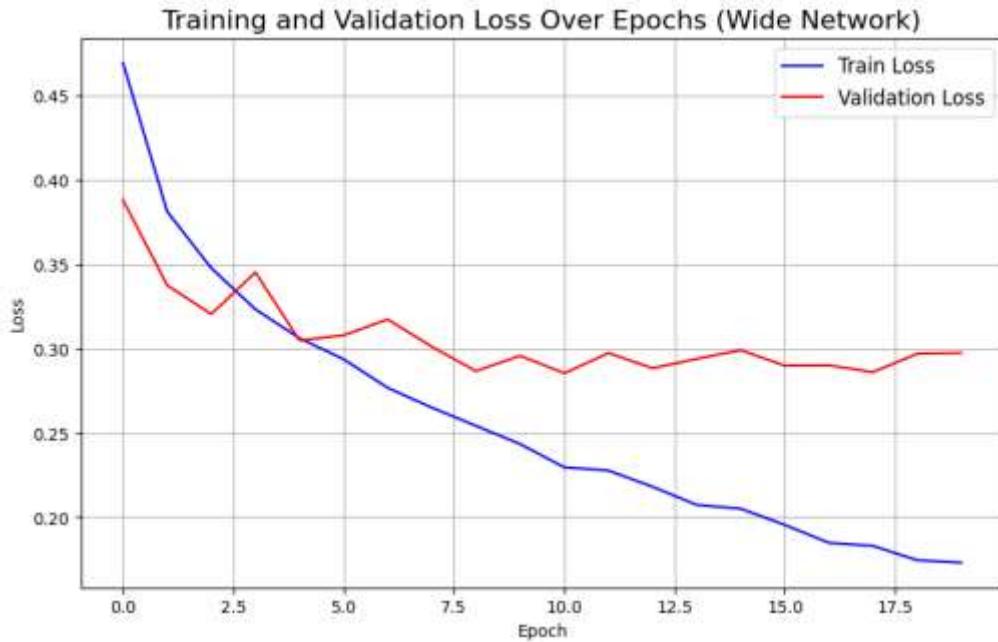
```
Epoch [1/20], Loss: 0.4695, Validation Loss: 0.3882
Epoch [2/20], Loss: 0.3813, Validation Loss: 0.3376
Epoch [3/20], Loss: 0.3478, Validation Loss: 0.3204
Epoch [4/20], Loss: 0.3233, Validation Loss: 0.3453
Epoch [5/20], Loss: 0.3061, Validation Loss: 0.3049
Epoch [6/20], Loss: 0.2937, Validation Loss: 0.3079
Epoch [7/20], Loss: 0.2767, Validation Loss: 0.3173
Epoch [8/20], Loss: 0.2650, Validation Loss: 0.3012
Epoch [9/20], Loss: 0.2542, Validation Loss: 0.2866
Epoch [10/20], Loss: 0.2434, Validation Loss: 0.2958
Epoch [11/20], Loss: 0.2295, Validation Loss: 0.2854
Epoch [12/20], Loss: 0.2276, Validation Loss: 0.2974
Epoch [13/20], Loss: 0.2181, Validation Loss: 0.2884
Epoch [14/20], Loss: 0.2072, Validation Loss: 0.2939
Epoch [15/20], Loss: 0.2050, Validation Loss: 0.2991
Epoch [16/20], Loss: 0.1954, Validation Loss: 0.2899
Epoch [17/20], Loss: 0.1847, Validation Loss: 0.2900
Epoch [18/20], Loss: 0.1829, Validation Loss: 0.2861
Epoch [19/20], Loss: 0.1744, Validation Loss: 0.2969
Epoch [20/20], Loss: 0.1730, Validation Loss: 0.2974
```

همانطور که در نتایج آموزش مدل مشاهده می‌شود، در ابتدای فرآیند آموزش یا Epoch 1، خطای آموزش 0.4695 و خطای اعتبارسنجی 0.3882 بود، با پیشرفت آموزش هر دو معیار به طور بیوسته کاهش یافته‌ند و در پایان دوره‌های آموزش یا Epoch 20، خطای آموزش به 0.1730 و خطای اعتبارسنجی به 0.2974 رسید که نشان‌دهنده بهبود تدریجی مدل است و در صورت ادامه با دوره‌ای بیشتر احتمالاً شاهد بهبود بیشتر مدل خواهیم بود.

دقت نهایی مدل شبکه عریض را نیز در زیر مشاهده می‌کنیم و همانطور که می‌بینیم نسبت به مدل شبکه عمیق دقต بیشتری را این شبکه بدست آورده است:

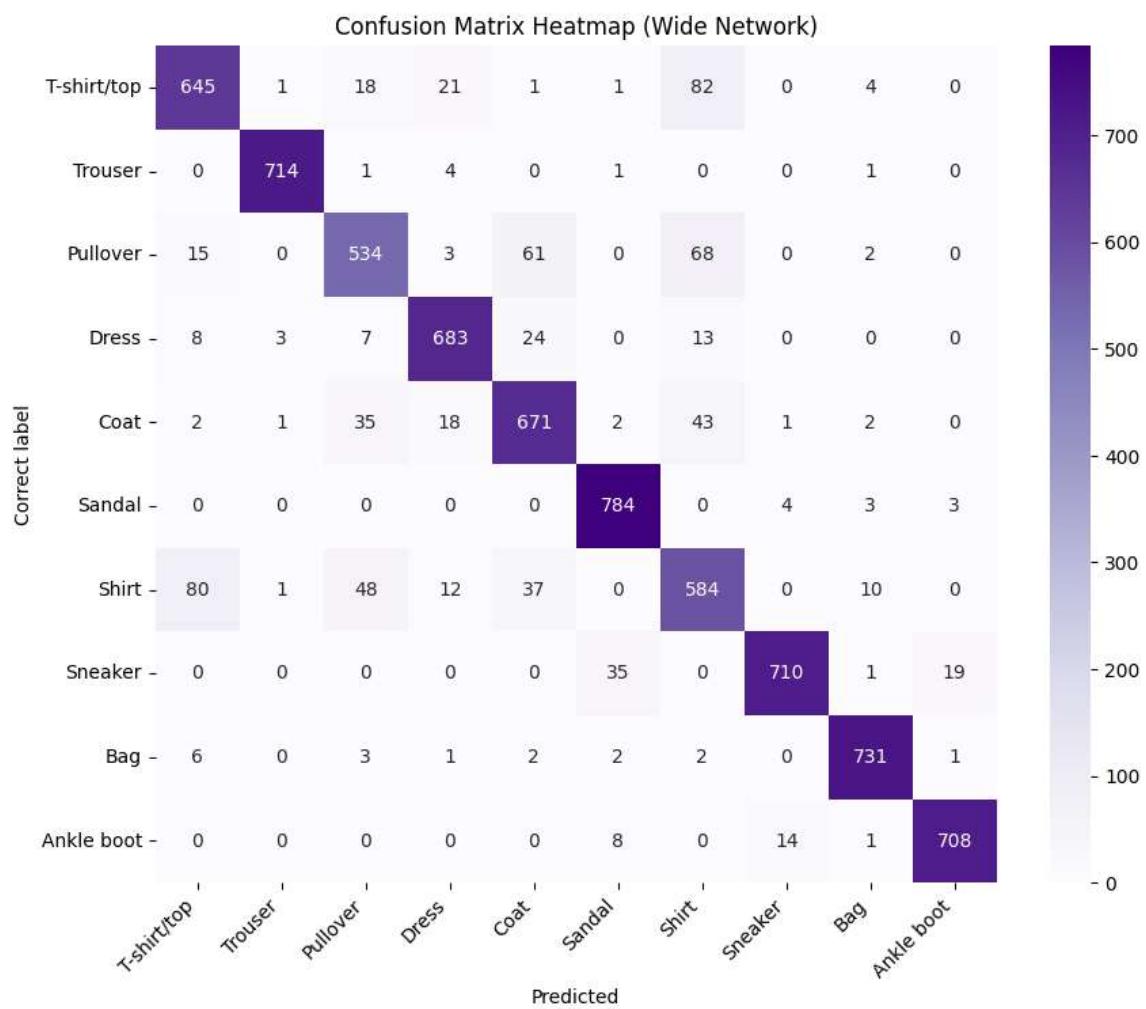
Test Accuracy: 90.19%

در ادامه برای اینکه بهتر بتوانیم این خطای آموزش و اعتبارسنجی را تجزیه و تحلیل کنیم و همچنین همگرایی در این مدل را بررسی کنیم نمودار مربوطه را رسم می‌کنیم:



همان طور که از تصویر هم مشخص است به مرور و با افزایش Epoch مقدار خطا یا Loss برای آموزشی کاهش پیدا می کند و نمودار شبیه نزولی دارد و یعنی به مرور خطای کمتر می شود و مدل یاد می گیرد که بهتر عمل کند و دسته بندی درست و بهتری برای این دسته بندی داشته باشد اما برای اعتبارسنجی دارای صعود و نزول های مختلف متعددی است و هنوز مدل ما همگرا نشده است و احتمالاً اگر تعداد epoch ها را افزایش دهیم، به مرور روند کاهش خطای ادامه یابد و همچنین دقیق نیز افزایش یابد.

حالا می آییم برای ارزیابی دقیق تر عملکرد مدل، ماتریس درهم ریختگی را رسم می کنیم که نمایش دهنده تعداد پیش بینی های صحیح و غلط برای هر کلاس است و طبق توضیحاتی که برای بخش قبل یعنی مدل شبکه عمیق دادیم می توانیم اینجا هم از همان فرمول ها برای محاسبه دقیق و ... استفاده کنیم و یا سطر و ستون های مختلف را تجزیه و تحلیل کنیم و به عنوان مثال در ماتریس زیر مشاهده می کنیم که 645 نمونه را به درستی پیش بینی کردیم که جزو کلاس TROUSER می باشند و به همین ترتیب مواردی که روی قطر اصلی ماتریس قرار دارند همان TP های هر کلاس هستند و مجموع آنها نیز به ما TP کل را می دهد و می توانیم به همین صورت FP، FN و ... را هم روی ماتریس بررسی کنیم و به تجزیه و تحلیل بپردازیم:



.2

همانطور که از نتایج بخش قبل هم متوجه شدیم شبکه عریض بهتر عمل کرد و در ادامه به تحلیل و بررسی عملکرد دقیق این دو مدل می پردازیم.

Test Accuracy: 90.04%

Test Accuracy: 90.19%

شبکه عمیق که دارای چهار لایه‌ی پنهان با تعداد نورون‌های 512، 512، 256 و 128 است، در نهایت به دقت 90.04 درصد روی داده‌های FashionMNIST دست پیدا کرد؛ در مقابل نیز شبکه عریض که تنها یک لایه‌ی

پنهان با 1024 نورون دارد، توانست دقت 90.19 درصد را به دست آورد و می بینیم که مدل شبکه عریض 0.15 درصد دقیق‌تر از شبکه عمیق شده است و توانسته دقت بالاتری را کسب کند.

همانطور که مشخص است برای این تمرین با این مقداردهی هایی که برای مدل در نظر گرفتیم مدل شبکه عریض عملکرد بهتری نسبت به شبکه عمیق داشته است و از لحاظ ساختار شبکه عمیق سعی می کند ویژگی ها را در چندین مرحله به صورت سلسله مراتبی استخراج کند و ابتدا ویژگی های ساده را یاد می گیرد و سپس در لایه های بعدی آنها را با هم ترکیب می کند تا به ویژگی های پیچیده تری برسد. این نوع ساختار معمولاً برای داده هایی که دارای ساختارهای پیچیده تری هستند بهتر عمل می کند اما در این مسئله که پیچیدگی آن چنانی نداشتیم (ماهیت داده های FashionMNIST اگرچه از MNIST پیچیده تر است اما همچنان آنقدر پیچیدگی ندارد که نیاز به استخراج ویژگی های سلسله مراتبی عمیق داشته باشد)؛ این عمق بالا چندان ضروری ندارد و حتی ممکن است باعث مشکلاتی مثل محوشدگی گرایان یا حساسیت زیاد به تنظیمات مدل شود. در مقابل شبکه عریض با وجود داشتن فقط یک لایه پنهان و تعداد زیاد نورون ها در همان لایه، توانایی مدل سازی مناسبی از داده ها دارد. علاوه بر ساده تر بودن مدل، به دلیل کمتر بودن تعداد لایه ها زمان آموزش کمتری نیاز دارد و کمتر دچار مشکلات رایج در شبکه های عمیق می شود و معمولاً پایداری بیشتری در فرآیند آموزش از خود نشان می دهد.

با توجه به دقت نهایی بیشتر، سادگی در طراحی و آموزش و تطابق بهتر با نوع داده های این مسئله می توانیم نتیجه بگیریم که برای دسته بندی داده های FashionMNIST استفاده از معماری شبکه عریض نسبت به معماری شبکه عمیق مناسب تر است و شبکه عریض با ارائه عملکرد بهتر و پیچیدگی کمتر، انتخاب بهتری برای این مسئله محسوب می شود.

:Question 2

.1

در این آزمایش، سه روش مقداردهی اولیه مختلف He، Xavier و تصادفی را بر روی مدل‌های DNN و WNN پیاده‌سازی کردیم. مقداردهی He با تکنیک Kaiming برای توابع ReLU بهینه شده و با تنظیم واریانس وزن‌ها بر اساس fan_in، همگرایی پایدارتری ایجاد می‌کند. مقداردهی Xavier هم با در نظر گرفتن fan_in و fan_out، برای حفظ واریانس در تمام لایه‌ها طراحی شده است. روش تصادفی ساده با توزیع نرمال (std=0.01) نیز به عنوان پایه مقایسه استفاده کردیم. در نهایت هم نتایج آموختش، دقت و ... را خروجی نشان دادیم که تصاویر سمت چپ برای شبکه عمیق و تصاویر سمت راست برای شبکه عریض می‌باشند.

همانطور که در تصاویر پایین هم مشاهده می‌کنیم بهترین عملکرد را مقداردهی Xavier بین این سه الگوریتم گفته شده برای شبکه عریض به دقت 89.99 درصد و برای شبکه عمیق نیز با مقداردهی تصادفی به دقت 89.52 درصد رسانده است. برای شبکه عمیق ترتیب دقت در این سه مدل به صورت Xavier < He < Random است و برای شبکه عریض هم به صورت Random < Xavier < He است اما تفاوت دقت‌ها خیلی کم است و مثلاً مقداردهی تصادفی و He در شبکه عریض 0.01 درصد با هم تفاوت دارند (در ضمن آموختش مدل را روی همان مدل پیش فرض اولیه که در سوال اول تمرین پیاده‌سازی کرده بودیم اجرا کردیم).

نتایج هر کدام از این سه مقداردهی روی مدل شبکه عمیق (سمت چپ‌ها) و مدل شبکه عریض (سمت راست‌ها) را می‌توانیم در خروجی‌های زیر هم مشاهده کنیم:

- **He Initialization**

```
Epoch [1/20], Loss: 0.8082, Validation Loss: 0.4458
Epoch [2/20], Loss: 0.5433, Validation Loss: 0.4022
Epoch [3/20], Loss: 0.4902, Validation Loss: 0.3935
Epoch [4/20], Loss: 0.4680, Validation Loss: 0.3719
Epoch [5/20], Loss: 0.4477, Validation Loss: 0.3628
Epoch [6/20], Loss: 0.4339, Validation Loss: 0.3618
Epoch [7/20], Loss: 0.4162, Validation Loss: 0.3354
Epoch [8/20], Loss: 0.4055, Validation Loss: 0.3428
Epoch [9/20], Loss: 0.3960, Validation Loss: 0.3280
Epoch [10/20], Loss: 0.3892, Validation Loss: 0.3329
Epoch [11/20], Loss: 0.3791, Validation Loss: 0.3250
Epoch [12/20], Loss: 0.3753, Validation Loss: 0.3322
Epoch [13/20], Loss: 0.3686, Validation Loss: 0.3208
Epoch [14/20], Loss: 0.3620, Validation Loss: 0.3250
Epoch [15/20], Loss: 0.3592, Validation Loss: 0.3342
Epoch [16/20], Loss: 0.3593, Validation Loss: 0.3241
Epoch [17/20], Loss: 0.3516, Validation Loss: 0.3138
Epoch [18/20], Loss: 0.3490, Validation Loss: 0.3110
Epoch [19/20], Loss: 0.3475, Validation Loss: 0.3081
Epoch [20/20], Loss: 0.3435, Validation Loss: 0.3228
Test Accuracy: 88.51%
```

```
Epoch [1/20], Loss: 0.5321, Validation Loss: 0.3959
Epoch [2/20], Loss: 0.4124, Validation Loss: 0.3608
Epoch [3/20], Loss: 0.3826, Validation Loss: 0.3348
Epoch [4/20], Loss: 0.3617, Validation Loss: 0.3227
Epoch [5/20], Loss: 0.3456, Validation Loss: 0.3232
Epoch [6/20], Loss: 0.3314, Validation Loss: 0.3156
Epoch [7/20], Loss: 0.3211, Validation Loss: 0.3159
Epoch [8/20], Loss: 0.3080, Validation Loss: 0.3055
Epoch [9/20], Loss: 0.2979, Validation Loss: 0.3048
Epoch [10/20], Loss: 0.2932, Validation Loss: 0.2970
Epoch [11/20], Loss: 0.2831, Validation Loss: 0.3083
Epoch [12/20], Loss: 0.2813, Validation Loss: 0.2972
Epoch [13/20], Loss: 0.2706, Validation Loss: 0.3029
Epoch [14/20], Loss: 0.2654, Validation Loss: 0.3021
Epoch [15/20], Loss: 0.2535, Validation Loss: 0.3069
Epoch [16/20], Loss: 0.2532, Validation Loss: 0.3066
Epoch [17/20], Loss: 0.2476, Validation Loss: 0.3102
Epoch [18/20], Loss: 0.2406, Validation Loss: 0.3107
Epoch [19/20], Loss: 0.2386, Validation Loss: 0.3180
Epoch [20/20], Loss: 0.2336, Validation Loss: 0.3184
Test Accuracy: 89.87%
```

- Xavier Initialization

```
Epoch [1/20], Loss: 0.7297, Validation Loss: 0.4243
Epoch [2/20], Loss: 0.5230, Validation Loss: 0.4028
Epoch [3/20], Loss: 0.4735, Validation Loss: 0.3928
Epoch [4/20], Loss: 0.4545, Validation Loss: 0.3654
Epoch [5/20], Loss: 0.4323, Validation Loss: 0.3692
Epoch [6/20], Loss: 0.4188, Validation Loss: 0.3501
Epoch [7/20], Loss: 0.4040, Validation Loss: 0.3397
Epoch [8/20], Loss: 0.3946, Validation Loss: 0.3360
Epoch [9/20], Loss: 0.3851, Validation Loss: 0.3341
Epoch [10/20], Loss: 0.3734, Validation Loss: 0.3231
Epoch [11/20], Loss: 0.3693, Validation Loss: 0.3134
Epoch [12/20], Loss: 0.3630, Validation Loss: 0.3293
Epoch [13/20], Loss: 0.3617, Validation Loss: 0.3267
Epoch [14/20], Loss: 0.3542, Validation Loss: 0.3160
Epoch [15/20], Loss: 0.3525, Validation Loss: 0.3194
Epoch [16/20], Loss: 0.3454, Validation Loss: 0.3015
Epoch [17/20], Loss: 0.3430, Validation Loss: 0.3217
Epoch [18/20], Loss: 0.3371, Validation Loss: 0.3084
Epoch [19/20], Loss: 0.3334, Validation Loss: 0.3111
Epoch [20/20], Loss: 0.3288, Validation Loss: 0.3136
```

Test Accuracy: 88.44%

```
Epoch [1/20], Loss: 0.5258, Validation Loss: 0.3848
Epoch [2/20], Loss: 0.4140, Validation Loss: 0.3746
Epoch [3/20], Loss: 0.3855, Validation Loss: 0.3471
Epoch [4/20], Loss: 0.3601, Validation Loss: 0.3339
Epoch [5/20], Loss: 0.3460, Validation Loss: 0.3287
Epoch [6/20], Loss: 0.3346, Validation Loss: 0.3103
Epoch [7/20], Loss: 0.3180, Validation Loss: 0.3155
Epoch [8/20], Loss: 0.3110, Validation Loss: 0.3242
Epoch [9/20], Loss: 0.3005, Validation Loss: 0.3062
Epoch [10/20], Loss: 0.2947, Validation Loss: 0.3250
Epoch [11/20], Loss: 0.2878, Validation Loss: 0.3090
Epoch [12/20], Loss: 0.2768, Validation Loss: 0.3103
Epoch [13/20], Loss: 0.2684, Validation Loss: 0.3052
Epoch [14/20], Loss: 0.2669, Validation Loss: 0.3269
Epoch [15/20], Loss: 0.2586, Validation Loss: 0.3174
Epoch [16/20], Loss: 0.2535, Validation Loss: 0.2955
Epoch [17/20], Loss: 0.2521, Validation Loss: 0.3105
Epoch [18/20], Loss: 0.2428, Validation Loss: 0.3083
Epoch [19/20], Loss: 0.2383, Validation Loss: 0.3290
Epoch [20/20], Loss: 0.2351, Validation Loss: 0.3066
```

Test Accuracy: 89.99%

- Random initialization

```
Epoch [1/20], Loss: 0.7838, Validation Loss: 0.5338
Epoch [2/20], Loss: 0.5395, Validation Loss: 0.4156
Epoch [3/20], Loss: 0.4713, Validation Loss: 0.3975
Epoch [4/20], Loss: 0.4387, Validation Loss: 0.3896
Epoch [5/20], Loss: 0.4188, Validation Loss: 0.3622
Epoch [6/20], Loss: 0.4059, Validation Loss: 0.3861
Epoch [7/20], Loss: 0.3888, Validation Loss: 0.3687
Epoch [8/20], Loss: 0.3807, Validation Loss: 0.3558
Epoch [9/20], Loss: 0.3672, Validation Loss: 0.3419
Epoch [10/20], Loss: 0.3650, Validation Loss: 0.3204
Epoch [11/20], Loss: 0.3557, Validation Loss: 0.3303
Epoch [12/20], Loss: 0.3509, Validation Loss: 0.3339
Epoch [13/20], Loss: 0.3440, Validation Loss: 0.3210
Epoch [14/20], Loss: 0.3388, Validation Loss: 0.3344
Epoch [15/20], Loss: 0.3330, Validation Loss: 0.3259
Epoch [16/20], Loss: 0.3315, Validation Loss: 0.3264
Epoch [17/20], Loss: 0.3232, Validation Loss: 0.3146
Epoch [18/20], Loss: 0.3217, Validation Loss: 0.3157
Epoch [19/20], Loss: 0.3209, Validation Loss: 0.3070
Epoch [20/20], Loss: 0.3198, Validation Loss: 0.3105
```

Test Accuracy: 89.52%

```
Epoch [1/20], Loss: 0.5248, Validation Loss: 0.4137
Epoch [2/20], Loss: 0.4141, Validation Loss: 0.3731
Epoch [3/20], Loss: 0.3810, Validation Loss: 0.3711
Epoch [4/20], Loss: 0.3575, Validation Loss: 0.3459
Epoch [5/20], Loss: 0.3399, Validation Loss: 0.3357
Epoch [6/20], Loss: 0.3263, Validation Loss: 0.3169
Epoch [7/20], Loss: 0.3166, Validation Loss: 0.3421
Epoch [8/20], Loss: 0.3047, Validation Loss: 0.3183
Epoch [9/20], Loss: 0.2942, Validation Loss: 0.3112
Epoch [10/20], Loss: 0.2900, Validation Loss: 0.3320
Epoch [11/20], Loss: 0.2796, Validation Loss: 0.3177
Epoch [12/20], Loss: 0.2775, Validation Loss: 0.3137
Epoch [13/20], Loss: 0.2690, Validation Loss: 0.3061
Epoch [14/20], Loss: 0.2604, Validation Loss: 0.3091
Epoch [15/20], Loss: 0.2578, Validation Loss: 0.3243
Epoch [16/20], Loss: 0.2493, Validation Loss: 0.3180
Epoch [17/20], Loss: 0.2455, Validation Loss: 0.3154
Epoch [18/20], Loss: 0.2393, Validation Loss: 0.3072
Epoch [19/20], Loss: 0.2381, Validation Loss: 0.3264
Epoch [20/20], Loss: 0.2297, Validation Loss: 0.3097
```

Test Accuracy: 89.88%

۲

در این بخش باید این سه الگوریتم را بررسی کنیم و به نقاط ضعف و قوت آنها پردازیم.

برای شبکه عمیق ۵ لایه‌ای، روش تصادفی با دقت ۸۹.۵۲ درصد بهترین نتیجه را نشان داد. این نتیجه نشان‌دهنده مقاومت خوب این معماری در برابر مقداردهی‌های مختلف است. روش He با دقت ۸۸.۵۱ درصد و Xavier با دقت ۸۸.۴۴ درصد در جایگاه‌های بعدی قرار گرفتند. این نتایج بهوضوح نشان می‌دهد که اگرچه He و Xavier برای توابع ReLU و معماری‌های عمیق بهینه شده‌اند، اما در این اجرای خاص، مقداردهی اولیه تصادفی به یک نقطه بهینه محلی بهتر رسیده و عملکرد قوی‌تری از خود نشان داده است.

در شبکه عریض، روش‌های Xavier و تصادفی عملکرد مشابهی از خود نشان دادند، با این تفاوت که روش Xavier با دقت ۸۹.۹۹ درصد بهترین نتیجه را کسب کرد و پس از آن He با دقت ۸۹.۸۷ درصد و تصادفی با دقت ۸۹.۸۸ درصد قرار گرفتند. این نتایج حاکی از آن است که در معماری شبکه عریض، انتخاب روش مقداردهی اولیه تأثیر کمتری بر عملکرد نهایی مدل دارد و خود ساختار لایه‌های گستره کمک می‌کند تا فرآیند یادگیری تثبیت شود. در نهایت می‌توانیم نتیجه بگیریم که برای شبکه عریض هم Xavier بهترین نتیجه را داشته است.

▪ روشن: He

برای این روش هم می‌توانیم از توزیع یکنواخت استفاده کنیم و هم توزیع نرمال که ما از توزیع نرمال با میانگین صفر و واریانس معین (که به تعداد نورون‌ها بستگی دارد) استفاده کردیم. $\mathcal{N}(0, \frac{2}{\text{fan_in}})$

- این مقداردهی کمک می‌کند تا مشکلاتی مانند محوشدنگی یا انفجار گرادیان در شبکه‌های عمیق بطریف شوند و باعث بهبود سرعت همگرایی مدل‌ها می‌شود. این روش بیشتر برای فعال‌سازی‌های ReLU طراحی شده است؛ همچنین میزان همگرایی سریع‌تر از مقداردهی تصادفی و Xavier در شبکه‌هایی با تعداد لایه زیاد دارد. اما در مقابل نسبت به روش‌های دیگر ممکن است محاسبات ریاضیاتی بیشتر و پیچیده‌تری داشته باشد و در صورت استفاده با توابع فعال‌سازی غیر از ReLU عملکرد ضعیفی داشته باشد و در شبکه‌های بسیار عمیق‌تر ممکن است مشکلاتی مانند exploding gradient ایجاد شود.

▪ روشن: Xavier

این روش بیشتر برای تابع فعال‌سازی‌هایی مانند tanh و sigmoid طراحی شده اما می‌توان برای ReLU هم استفاده کرد و با استفاده از توزیع نرمال به طوری مقداردهی می‌کند که واریانس وزن‌ها با تعداد ورودی‌ها و خروجی‌ها متناسب باشد. $\mathcal{N}(0, \frac{2}{\text{fan_in} + \text{fan_out}})$

این مقداردهی به کاهش مشکلات مشابه در شبکه‌های عمیق کمک می‌کند و باعث بهینه‌سازی همگرایی مدل‌ها می‌شود و واریانس وزن‌ها به گونه‌ای تنظیم می‌شود که شبکه به خوبی یادگیری کند. این روش به طور کلی برای مدل‌هایی که از توابع فعال‌سازی غیر از ReLU استفاده می‌کنند مثل tanh یا sigmoid، گزینه‌ی مناسبی است،

همچنین می‌تواند به کاهش مشکل محوشدگی گرادیان در شبکه‌های عمیق کمک کند اما در مقابل در مدل‌های بسیار عمیق یا پیچیده‌تر ممکن است مشکلات مشابه He در رابطه با exploding gradient مشاهده شود؛ همچنین ممکن است در شبکه‌هایی که از ReLU به عنوان تابع فعال‌سازی استفاده می‌کنند بازده خوبی نداشته باشد که در اینجا توانسته بازده خوبی از خود نشان دهد.

▪ روشن: Random

در این روش از توزیع نرمال با میانگین صفر و انحراف معیار ثابت برای مقداردهی به وزن‌ها استفاده می‌شود و ساده‌ترین روش مقداردهی است و نیاز به تنظیمات خاصی ندارد. این روش معمولاً برای مدل‌های ساده‌تر یا زمانی که می‌خواهیم کنترل بیشتری روی مقداردهی داشته باشیم مناسب است اما در مقایسه با He یا Xavier ممکن است در شبکه‌های عمیق‌تر مشکلاتی مثل محوشدگی گرادیان بیشتر نمایان شود (در شبکه‌های ساده و کم‌عمق، به خوبی عمل می‌کند و ممکن است باعث همگرایی سریع‌تری در مقایسه با روش‌های پیچیده‌تر شود)؛ همچنین این روش در شبکه‌های با تعداد لایه زیاد معمولاً منجر به یادگیری ضعیف‌تر و کندر خواهد شد، زیرا وزن‌ها به طور تصادفی و بدون در نظر گرفتن ویژگی‌های معماری شبکه مقداردهی می‌شوند. این روش بسیار وابسته به مقدایر اولیه انتخاب شده است و همگرایی کندری دارد، همچنین تنظیم پارامتر std باید درست و دقیق انجام شود.

:Question 3

۱.

نرخ یادگیری یکی از مهمترین ابزارهای آموزش شبکه‌های عصبی است که تأثیر زیادی بر سرعت همگرایی و کیفیت مدل نهایی دارد. در این پارت به بررسی تأثیر نرخ‌های یادگیری مختلف بر دقت مدل‌ها می‌پردازیم و نرخ‌های یادگیری مختلف را برای مدل‌های شبکه عمیق و عریض اعمال کردیم و در نهایت هم نتایج آموزش، دقت و ... را خروجی نشان دادیم که تصاویر سمت چپ برای شبکه عمیق و تصاویر سمت راست برای شبکه عریض می‌باشند.

همانطور که از تصاویر پایین هم مشخص است برای شبکه عمیق مقایسه نرخ یادگیری بر اساس دقت بدست آمده به صورت $lr=0.001 < lr=0.01 < lr=0.1 < lr=1$ است و بهترین عملکرد را با نرخ 0.001 بدست آورده و وقتی نرخ یادگیری را 10 برابر بزرگ کردیم دقت 0.81 درصد کم شد اما مجدد که نرخ را 10 برابر بزرگ کردیم و به مقدار 0.01 رسید شبکه کاهش دقت کمی بیشتر شد و این دفعه حدود 1.28 درصد دقت مدل کم شد ولی وقتی نرخ یادگیری را برابر 1 گذاشتیم این کاهش دقت شبکه خیلی زیادی گرفت و دقت به یکباره به حدود 10 درصد رسید.

برای شبکه عریض هم به همین صورت می‌توانیم تحلیل کنیم و مقایسه نرخ یادگیری بر اساس دقت بدست آمده مثل شبکه عمیق به صورت $lr=0.001 < lr=0.01 < lr=0.1 < lr=1$ است و بهترین عملکرد را با نرخ 0.001 بدست آورده و وقتی نرخ یادگیری را 10 برابر بزرگ کردیم دقت 0.16 درصد کم شد اما مجدد که نرخ را 10 برابر بزرگ کردیم و به مقدار 0.01 رسید شبکه کاهش دقت کمی بیشتر شد و این دفعه حدود 4.94 درصد دقت مدل کم شد ولی وقتی نرخ یادگیری را برابر 1 گذاشتیم این کاهش دقت شبکه خیلی زیادی گرفت و دقت به یکباره به حدود 10 درصد رسید.

نتایج هر کدام از این مقداردهی‌های نرخ یادگیری روی مدل شبکه عمیق (سمت چپ‌ها) و مدل شبکه عریض (سمت راست‌ها) را می‌توانیم در خروجی‌های زیر هم مشاهده کنیم:

- Learning Rate = 0.001

```
Epoch [1/20], Loss: 0.6040, Validation Loss: 0.3709
Epoch [2/20], Loss: 0.4475, Validation Loss: 0.3488
Epoch [3/20], Loss: 0.4073, Validation Loss: 0.3194
Epoch [4/20], Loss: 0.3813, Validation Loss: 0.3442
Epoch [5/20], Loss: 0.3602, Validation Loss: 0.3024
Epoch [6/20], Loss: 0.3481, Validation Loss: 0.3003
Epoch [7/20], Loss: 0.3333, Validation Loss: 0.3069
Epoch [8/20], Loss: 0.3208, Validation Loss: 0.2913
Epoch [9/20], Loss: 0.3116, Validation Loss: 0.2876
Epoch [10/20], Loss: 0.2998, Validation Loss: 0.2791
Epoch [11/20], Loss: 0.2900, Validation Loss: 0.2736
Epoch [12/20], Loss: 0.2859, Validation Loss: 0.2774
Epoch [13/20], Loss: 0.2769, Validation Loss: 0.2759
Epoch [14/20], Loss: 0.2678, Validation Loss: 0.2757
Epoch [15/20], Loss: 0.2605, Validation Loss: 0.2664
Epoch [16/20], Loss: 0.2513, Validation Loss: 0.2641
Epoch [17/20], Loss: 0.2465, Validation Loss: 0.2662
Epoch [18/20], Loss: 0.2433, Validation Loss: 0.2714
Epoch [19/20], Loss: 0.2389, Validation Loss: 0.2619
Epoch [20/20], Loss: 0.2313, Validation Loss: 0.2709
```

Test Accuracy: 90.04%

```
Epoch [1/20], Loss: 0.4695, Validation Loss: 0.3882
Epoch [2/20], Loss: 0.3813, Validation Loss: 0.3376
Epoch [3/20], Loss: 0.3478, Validation Loss: 0.3204
Epoch [4/20], Loss: 0.3233, Validation Loss: 0.3453
Epoch [5/20], Loss: 0.3061, Validation Loss: 0.3049
Epoch [6/20], Loss: 0.2937, Validation Loss: 0.3079
Epoch [7/20], Loss: 0.2767, Validation Loss: 0.3173
Epoch [8/20], Loss: 0.2650, Validation Loss: 0.3012
Epoch [9/20], Loss: 0.2542, Validation Loss: 0.2866
Epoch [10/20], Loss: 0.2434, Validation Loss: 0.2958
Epoch [11/20], Loss: 0.2295, Validation Loss: 0.2854
Epoch [12/20], Loss: 0.2276, Validation Loss: 0.2974
Epoch [13/20], Loss: 0.2181, Validation Loss: 0.2884
Epoch [14/20], Loss: 0.2072, Validation Loss: 0.2939
Epoch [15/20], Loss: 0.2050, Validation Loss: 0.2991
Epoch [16/20], Loss: 0.1954, Validation Loss: 0.2899
Epoch [17/20], Loss: 0.1847, Validation Loss: 0.2900
Epoch [18/20], Loss: 0.1829, Validation Loss: 0.2861
Epoch [19/20], Loss: 0.1744, Validation Loss: 0.2969
Epoch [20/20], Loss: 0.1730, Validation Loss: 0.2974
```

Test Accuracy: 90.19%

- Learning Rate = 0.01

```
Epoch [1/20], Loss: 0.6265, Validation Loss: 0.4392
Epoch [2/20], Loss: 0.4823, Validation Loss: 0.3653
Epoch [3/20], Loss: 0.4457, Validation Loss: 0.3653
Epoch [4/20], Loss: 0.4138, Validation Loss: 0.3326
Epoch [5/20], Loss: 0.3974, Validation Loss: 0.3184
Epoch [6/20], Loss: 0.3815, Validation Loss: 0.3213
Epoch [7/20], Loss: 0.3661, Validation Loss: 0.3122
Epoch [8/20], Loss: 0.3579, Validation Loss: 0.3105
Epoch [9/20], Loss: 0.3454, Validation Loss: 0.2952
Epoch [10/20], Loss: 0.3315, Validation Loss: 0.3084
Epoch [11/20], Loss: 0.3235, Validation Loss: 0.3110
Epoch [12/20], Loss: 0.3175, Validation Loss: 0.2834
Epoch [13/20], Loss: 0.3100, Validation Loss: 0.3085
Epoch [14/20], Loss: 0.3020, Validation Loss: 0.2865
Epoch [15/20], Loss: 0.2923, Validation Loss: 0.2836
Epoch [16/20], Loss: 0.2895, Validation Loss: 0.2863
Epoch [17/20], Loss: 0.2805, Validation Loss: 0.2933
Epoch [18/20], Loss: 0.2771, Validation Loss: 0.2736
Epoch [19/20], Loss: 0.2727, Validation Loss: 0.2741
Epoch [20/20], Loss: 0.2653, Validation Loss: 0.2865
```

Test Accuracy: 89.23%

```
Epoch [1/20], Loss: 0.5461, Validation Loss: 0.4017
Epoch [2/20], Loss: 0.4097, Validation Loss: 0.3905
Epoch [3/20], Loss: 0.3788, Validation Loss: 0.3307
Epoch [4/20], Loss: 0.3600, Validation Loss: 0.3405
Epoch [5/20], Loss: 0.3471, Validation Loss: 0.3683
Epoch [6/20], Loss: 0.3343, Validation Loss: 0.3216
Epoch [7/20], Loss: 0.3210, Validation Loss: 0.3205
Epoch [8/20], Loss: 0.3107, Validation Loss: 0.3067
Epoch [9/20], Loss: 0.3010, Validation Loss: 0.3321
Epoch [10/20], Loss: 0.2939, Validation Loss: 0.3209
Epoch [11/20], Loss: 0.2843, Validation Loss: 0.3161
Epoch [12/20], Loss: 0.2793, Validation Loss: 0.3050
Epoch [13/20], Loss: 0.2693, Validation Loss: 0.3065
Epoch [14/20], Loss: 0.2634, Validation Loss: 0.3023
Epoch [15/20], Loss: 0.2632, Validation Loss: 0.3145
Epoch [16/20], Loss: 0.2551, Validation Loss: 0.3152
Epoch [17/20], Loss: 0.2530, Validation Loss: 0.3011
Epoch [18/20], Loss: 0.2430, Validation Loss: 0.3074
Epoch [19/20], Loss: 0.2404, Validation Loss: 0.3127
Epoch [20/20], Loss: 0.2389, Validation Loss: 0.3065
```

Test Accuracy: 90.03%

- Learning Rate = 0.1

```
Epoch [1/20], Loss: 0.9094, Validation Loss: 0.5456
Epoch [2/20], Loss: 0.7870, Validation Loss: 0.4998
Epoch [3/20], Loss: 0.7417, Validation Loss: 0.5360
Epoch [4/20], Loss: 0.7310, Validation Loss: 0.5006
Epoch [5/20], Loss: 0.7067, Validation Loss: 0.4328
Epoch [6/20], Loss: 0.6903, Validation Loss: 0.4506
Epoch [7/20], Loss: 0.6804, Validation Loss: 0.4548
Epoch [8/20], Loss: 0.6536, Validation Loss: 0.4441
Epoch [9/20], Loss: 0.6384, Validation Loss: 0.4389
Epoch [10/20], Loss: 0.6337, Validation Loss: 0.4401
Epoch [11/20], Loss: 0.6260, Validation Loss: 0.4120
Epoch [12/20], Loss: 0.6200, Validation Loss: 0.4117
Epoch [13/20], Loss: 0.6046, Validation Loss: 0.4518
Epoch [14/20], Loss: 0.6103, Validation Loss: 0.3929
Epoch [15/20], Loss: 0.5958, Validation Loss: 0.3677
Epoch [16/20], Loss: 0.5914, Validation Loss: 0.3811
Epoch [17/20], Loss: 0.5916, Validation Loss: 0.3981
Epoch [18/20], Loss: 0.5800, Validation Loss: 0.3845
Epoch [19/20], Loss: 0.5831, Validation Loss: 0.3810
Epoch [20/20], Loss: 0.5744, Validation Loss: 0.3860
```

Test Accuracy: 87.95%

```
Epoch [1/20], Loss: 0.8232, Validation Loss: 0.4733
Epoch [2/20], Loss: 0.6421, Validation Loss: 0.4899
Epoch [3/20], Loss: 0.6434, Validation Loss: 0.4755
Epoch [4/20], Loss: 0.6218, Validation Loss: 0.4527
Epoch [5/20], Loss: 0.6199, Validation Loss: 0.5008
Epoch [6/20], Loss: 0.6131, Validation Loss: 0.4517
Epoch [7/20], Loss: 0.5994, Validation Loss: 0.4403
Epoch [8/20], Loss: 0.5999, Validation Loss: 0.4168
Epoch [9/20], Loss: 0.5845, Validation Loss: 0.4559
Epoch [10/20], Loss: 0.5900, Validation Loss: 0.4189
Epoch [11/20], Loss: 0.5842, Validation Loss: 0.4445
Epoch [12/20], Loss: 0.5768, Validation Loss: 0.4581
Epoch [13/20], Loss: 0.5755, Validation Loss: 0.4558
Epoch [14/20], Loss: 0.5834, Validation Loss: 0.4684
Epoch [15/20], Loss: 0.5743, Validation Loss: 0.4598
Epoch [16/20], Loss: 0.5993, Validation Loss: 0.4449
Epoch [17/20], Loss: 0.5960, Validation Loss: 0.4520
Epoch [18/20], Loss: 0.5907, Validation Loss: 0.4583
Epoch [19/20], Loss: 0.6173, Validation Loss: 0.4607
Epoch [20/20], Loss: 0.6887, Validation Loss: 0.4750
```

Test Accuracy: 85.09%

- Learning Rate = 1.0

```
Epoch [1/20], Loss: 2.6837, Validation Loss: 2.3391
Epoch [2/20], Loss: 2.4237, Validation Loss: 2.3764
Epoch [3/20], Loss: 2.3752, Validation Loss: 2.3612
Epoch [4/20], Loss: 2.3742, Validation Loss: 2.4558
Epoch [5/20], Loss: 2.3963, Validation Loss: 2.3611
Epoch [6/20], Loss: 2.3683, Validation Loss: 2.3382
Epoch [7/20], Loss: 2.3754, Validation Loss: 2.3730
Epoch [8/20], Loss: 2.3758, Validation Loss: 2.4564
Epoch [9/20], Loss: 2.3752, Validation Loss: 2.3321
Epoch [10/20], Loss: 2.3758, Validation Loss: 2.4023
Epoch [11/20], Loss: 2.3737, Validation Loss: 2.3845
Epoch [12/20], Loss: 2.3755, Validation Loss: 2.3768
Epoch [13/20], Loss: 2.3705, Validation Loss: 2.3759
Epoch [14/20], Loss: 2.3818, Validation Loss: 2.3690
Epoch [15/20], Loss: 2.3774, Validation Loss: 2.3387
Epoch [16/20], Loss: 2.3745, Validation Loss: 2.4382
Epoch [17/20], Loss: 2.3754, Validation Loss: 2.3964
Epoch [18/20], Loss: 2.3736, Validation Loss: 2.4025
Epoch [19/20], Loss: 2.3755, Validation Loss: 2.3529
Epoch [20/20], Loss: 2.3757, Validation Loss: 2.3489
```

Test Accuracy: 9.87%

```
Epoch [1/20], Loss: 5.0250, Validation Loss: 2.3764
Epoch [2/20], Loss: 2.4194, Validation Loss: 2.3856
Epoch [3/20], Loss: 2.4113, Validation Loss: 2.3348
Epoch [4/20], Loss: 2.3910, Validation Loss: 2.3747
Epoch [5/20], Loss: 2.3942, Validation Loss: 2.4247
Epoch [6/20], Loss: 2.4032, Validation Loss: 2.4011
Epoch [7/20], Loss: 2.4023, Validation Loss: 2.3618
Epoch [8/20], Loss: 2.3789, Validation Loss: 2.4311
Epoch [9/20], Loss: 2.3864, Validation Loss: 2.3683
Epoch [10/20], Loss: 2.3805, Validation Loss: 2.6004
Epoch [11/20], Loss: 2.3825, Validation Loss: 2.3704
Epoch [12/20], Loss: 2.3776, Validation Loss: 2.3655
Epoch [13/20], Loss: 2.3970, Validation Loss: 2.3495
Epoch [14/20], Loss: 2.3862, Validation Loss: 2.3790
Epoch [15/20], Loss: 2.4074, Validation Loss: 2.3417
Epoch [16/20], Loss: 2.3702, Validation Loss: 2.4140
Epoch [17/20], Loss: 2.3890, Validation Loss: 2.3395
Epoch [18/20], Loss: 2.3716, Validation Loss: 2.3431
Epoch [19/20], Loss: 2.3826, Validation Loss: 2.3830
Epoch [20/20], Loss: 2.3835, Validation Loss: 2.3539
```

Test Accuracy: 10.47%

اندازه دسته آموزشی یا batch size یکی از پارامترهای مهم در آموزش شبکه‌های عصبی است که می‌تواند بر سرعت همگرایی، دقت نهایی و مصرف منابع محاسباتی تأثیر بگذارد. در این بخش نیز تأثیر اندازه‌های مختلف دسته‌های آموزشی بر دقت شبکه‌های عمیق و عریض را بررسی می‌کنیم. همانطور که از نتایج هم مشخص است، برای شبکه عمیق مقایسه اندازه دسته‌های آموزشی بر اساس دقت به صورت $64 > 256 > 128 > 32$ است و بهترین عملکرد را با اندازه 32 به دست آورده که بیشترین دقت برای مدل شبکه عمیق در این آزمایش 90.57 درصد می‌باشد. همچنین کمترین دقت بین این اندازه‌های دسته‌های آموزشی مربوط به اندازه 64 است که دقت آن 90.04 درصد به دست آمد. با این وجود، می‌توان گفت که همه این اندازه‌ها عملکرد خیلی خوبی داشتند و همگی دقتی بین 90.04 تا 90.57 را به دست آورده و تقریباً عملکرد مشابهی را از خود نشان دادند.

برای شبکه عریض هم به همین منوال می‌توانیم تجزیه و تحلیل کنیم. مقایسه اندازه دسته‌های آموزشی بر اساس دقت به دست آمده به صورت $256 > 128 > 64 > 32$ حاصل شده است. کمترین دقت با اندازه 256 حاصل شد که مقدار 89.95 درصد می‌باشد و بیشترین دقت هم مربوط به اندازه 32 می‌باشد که دقت 90.41 درصد را نشان می‌دهد. به طور کلی می‌توانیم بگوییم که شبکه عمیق با اندازه دسته 32 بیشترین دقت را برای دسته‌بندی داده‌های مجموعه داده Fashion-MNIST با مقدار 90.57 درصد به دست آورده است.

نتایج هر کدام از این اندازه‌های دسته‌های آموزشی روی مدل شبکه عمیق (سمت چپ‌ها) و مدل شبکه عریض (سمت راست‌ها) را می‌توانیم در خروجی‌های زیر هم مشاهده کنیم:

- **batch_size = 32**

```
Epoch [1/20], Loss: 0.6257, Validation Loss: 0.4070
Epoch [2/20], Loss: 0.4761, Validation Loss: 0.3590
Epoch [3/20], Loss: 0.4379, Validation Loss: 0.3454
Epoch [4/20], Loss: 0.4056, Validation Loss: 0.3340
Epoch [5/20], Loss: 0.3872, Validation Loss: 0.3292
Epoch [6/20], Loss: 0.3694, Validation Loss: 0.3003
Epoch [7/20], Loss: 0.3502, Validation Loss: 0.3089
Epoch [8/20], Loss: 0.3442, Validation Loss: 0.3071
Epoch [9/20], Loss: 0.3298, Validation Loss: 0.2911
Epoch [10/20], Loss: 0.3206, Validation Loss: 0.2907
Epoch [11/20], Loss: 0.3131, Validation Loss: 0.2796
Epoch [12/20], Loss: 0.3027, Validation Loss: 0.2755
Epoch [13/20], Loss: 0.2930, Validation Loss: 0.2754
Epoch [14/20], Loss: 0.2859, Validation Loss: 0.2838
Epoch [15/20], Loss: 0.2809, Validation Loss: 0.2762
Epoch [16/20], Loss: 0.2736, Validation Loss: 0.2709
Epoch [17/20], Loss: 0.2674, Validation Loss: 0.2745
Epoch [18/20], Loss: 0.2675, Validation Loss: 0.2783
Epoch [19/20], Loss: 0.2568, Validation Loss: 0.2785
Epoch [20/20], Loss: 0.2519, Validation Loss: 0.2698

Test Accuracy: 90.57%
```

```
Epoch [1/20], Loss: 0.4991, Validation Loss: 0.4069
Epoch [2/20], Loss: 0.4066, Validation Loss: 0.3503
Epoch [3/20], Loss: 0.3718, Validation Loss: 0.3361
Epoch [4/20], Loss: 0.3435, Validation Loss: 0.3077
Epoch [5/20], Loss: 0.3263, Validation Loss: 0.3207
Epoch [6/20], Loss: 0.3083, Validation Loss: 0.3344
Epoch [7/20], Loss: 0.2957, Validation Loss: 0.2916
Epoch [8/20], Loss: 0.2817, Validation Loss: 0.2914
Epoch [9/20], Loss: 0.2699, Validation Loss: 0.2826
Epoch [10/20], Loss: 0.2605, Validation Loss: 0.2812
Epoch [11/20], Loss: 0.2482, Validation Loss: 0.2936
Epoch [12/20], Loss: 0.2418, Validation Loss: 0.2877
Epoch [13/20], Loss: 0.2347, Validation Loss: 0.2920
Epoch [14/20], Loss: 0.2291, Validation Loss: 0.2998
Epoch [15/20], Loss: 0.2166, Validation Loss: 0.2797
Epoch [16/20], Loss: 0.2125, Validation Loss: 0.2762
Epoch [17/20], Loss: 0.2037, Validation Loss: 0.2983
Epoch [18/20], Loss: 0.1973, Validation Loss: 0.2874
Epoch [19/20], Loss: 0.1963, Validation Loss: 0.2927
Epoch [20/20], Loss: 0.1872, Validation Loss: 0.2904

Test Accuracy: 90.41%
```

- batch_size = 64

```
Epoch [1/20], Loss: 0.6040, Validation Loss: 0.3709
Epoch [2/20], Loss: 0.4475, Validation Loss: 0.3488
Epoch [3/20], Loss: 0.4073, Validation Loss: 0.3194
Epoch [4/20], Loss: 0.3813, Validation Loss: 0.3442
Epoch [5/20], Loss: 0.3602, Validation Loss: 0.3024
Epoch [6/20], Loss: 0.3481, Validation Loss: 0.3003
Epoch [7/20], Loss: 0.3333, Validation Loss: 0.3069
Epoch [8/20], Loss: 0.3208, Validation Loss: 0.2913
Epoch [9/20], Loss: 0.3116, Validation Loss: 0.2876
Epoch [10/20], Loss: 0.2998, Validation Loss: 0.2791
Epoch [11/20], Loss: 0.2900, Validation Loss: 0.2736
Epoch [12/20], Loss: 0.2859, Validation Loss: 0.2774
Epoch [13/20], Loss: 0.2769, Validation Loss: 0.2759
Epoch [14/20], Loss: 0.2678, Validation Loss: 0.2757
Epoch [15/20], Loss: 0.2605, Validation Loss: 0.2664
Epoch [16/20], Loss: 0.2513, Validation Loss: 0.2641
Epoch [17/20], Loss: 0.2465, Validation Loss: 0.2662
Epoch [18/20], Loss: 0.2433, Validation Loss: 0.2714
Epoch [19/20], Loss: 0.2389, Validation Loss: 0.2619
Epoch [20/20], Loss: 0.2313, Validation Loss: 0.2709
```

Test Accuracy: 90.04%

```
Epoch [1/20], Loss: 0.4695, Validation Loss: 0.3882
Epoch [2/20], Loss: 0.3813, Validation Loss: 0.3376
Epoch [3/20], Loss: 0.3478, Validation Loss: 0.3204
Epoch [4/20], Loss: 0.3233, Validation Loss: 0.3453
Epoch [5/20], Loss: 0.3061, Validation Loss: 0.3049
Epoch [6/20], Loss: 0.2937, Validation Loss: 0.3079
Epoch [7/20], Loss: 0.2767, Validation Loss: 0.3173
Epoch [8/20], Loss: 0.2650, Validation Loss: 0.3012
Epoch [9/20], Loss: 0.2542, Validation Loss: 0.2866
Epoch [10/20], Loss: 0.2434, Validation Loss: 0.2958
Epoch [11/20], Loss: 0.2295, Validation Loss: 0.2854
Epoch [12/20], Loss: 0.2276, Validation Loss: 0.2974
Epoch [13/20], Loss: 0.2181, Validation Loss: 0.2884
Epoch [14/20], Loss: 0.2072, Validation Loss: 0.2939
Epoch [15/20], Loss: 0.2050, Validation Loss: 0.2991
Epoch [16/20], Loss: 0.1954, Validation Loss: 0.2899
Epoch [17/20], Loss: 0.1847, Validation Loss: 0.2900
Epoch [18/20], Loss: 0.1829, Validation Loss: 0.2861
Epoch [19/20], Loss: 0.1744, Validation Loss: 0.2969
Epoch [20/20], Loss: 0.1730, Validation Loss: 0.2974
```

Test Accuracy: 90.19%

- batch_size = 128

```
Epoch [1/20], Loss: 0.6074, Validation Loss: 0.3893
Epoch [2/20], Loss: 0.4341, Validation Loss: 0.3548
Epoch [3/20], Loss: 0.3901, Validation Loss: 0.3618
Epoch [4/20], Loss: 0.3665, Validation Loss: 0.3190
Epoch [5/20], Loss: 0.3514, Validation Loss: 0.3240
Epoch [6/20], Loss: 0.3352, Validation Loss: 0.3109
Epoch [7/20], Loss: 0.3213, Validation Loss: 0.3200
Epoch [8/20], Loss: 0.3079, Validation Loss: 0.3063
Epoch [9/20], Loss: 0.2961, Validation Loss: 0.2894
Epoch [10/20], Loss: 0.2895, Validation Loss: 0.3036
Epoch [11/20], Loss: 0.2795, Validation Loss: 0.2931
Epoch [12/20], Loss: 0.2768, Validation Loss: 0.2804
Epoch [13/20], Loss: 0.2639, Validation Loss: 0.2901
Epoch [14/20], Loss: 0.2584, Validation Loss: 0.2815
Epoch [15/20], Loss: 0.2505, Validation Loss: 0.2867
Epoch [16/20], Loss: 0.2420, Validation Loss: 0.2819
Epoch [17/20], Loss: 0.2386, Validation Loss: 0.2695
Epoch [18/20], Loss: 0.2319, Validation Loss: 0.2809
Epoch [19/20], Loss: 0.2247, Validation Loss: 0.2821
Epoch [20/20], Loss: 0.2226, Validation Loss: 0.2784
```

Test Accuracy: 90.53%

```
Epoch [1/20], Loss: 0.4646, Validation Loss: 0.3928
Epoch [2/20], Loss: 0.3639, Validation Loss: 0.3748
Epoch [3/20], Loss: 0.3344, Validation Loss: 0.3286
Epoch [4/20], Loss: 0.3076, Validation Loss: 0.3215
Epoch [5/20], Loss: 0.2913, Validation Loss: 0.3425
Epoch [6/20], Loss: 0.2762, Validation Loss: 0.3358
Epoch [7/20], Loss: 0.2660, Validation Loss: 0.3008
Epoch [8/20], Loss: 0.2513, Validation Loss: 0.3032
Epoch [9/20], Loss: 0.2428, Validation Loss: 0.3053
Epoch [10/20], Loss: 0.2336, Validation Loss: 0.2974
Epoch [11/20], Loss: 0.2222, Validation Loss: 0.3079
Epoch [12/20], Loss: 0.2149, Validation Loss: 0.3145
Epoch [13/20], Loss: 0.2068, Validation Loss: 0.2807
Epoch [14/20], Loss: 0.2002, Validation Loss: 0.2997
Epoch [15/20], Loss: 0.1880, Validation Loss: 0.2992
Epoch [16/20], Loss: 0.1885, Validation Loss: 0.2994
Epoch [17/20], Loss: 0.1786, Validation Loss: 0.2978
Epoch [18/20], Loss: 0.1754, Validation Loss: 0.3163
Epoch [19/20], Loss: 0.1729, Validation Loss: 0.3148
Epoch [20/20], Loss: 0.1652, Validation Loss: 0.3015
```

Test Accuracy: 90.28%

- batch_size = 256

```
Epoch [1/20], Loss: 0.6469, Validation Loss: 0.4248
Epoch [2/20], Loss: 0.4319, Validation Loss: 0.3661
Epoch [3/20], Loss: 0.4007, Validation Loss: 0.3313
Epoch [4/20], Loss: 0.3654, Validation Loss: 0.3182
Epoch [5/20], Loss: 0.3460, Validation Loss: 0.3276
Epoch [6/20], Loss: 0.3290, Validation Loss: 0.3108
Epoch [7/20], Loss: 0.3189, Validation Loss: 0.3015
Epoch [8/20], Loss: 0.3061, Validation Loss: 0.3024
Epoch [9/20], Loss: 0.3003, Validation Loss: 0.3156
Epoch [10/20], Loss: 0.2928, Validation Loss: 0.2951
Epoch [11/20], Loss: 0.2886, Validation Loss: 0.2976
Epoch [12/20], Loss: 0.2697, Validation Loss: 0.2853
Epoch [13/20], Loss: 0.2619, Validation Loss: 0.3002
Epoch [14/20], Loss: 0.2612, Validation Loss: 0.2798
Epoch [15/20], Loss: 0.2535, Validation Loss: 0.2798
Epoch [16/20], Loss: 0.2419, Validation Loss: 0.2894
Epoch [17/20], Loss: 0.2416, Validation Loss: 0.2934
Epoch [18/20], Loss: 0.2589, Validation Loss: 0.2868
Epoch [19/20], Loss: 0.2436, Validation Loss: 0.2895
Epoch [20/20], Loss: 0.2399, Validation Loss: 0.2854
```

Test Accuracy: 90.29%

```
Epoch [1/20], Loss: 0.4696, Validation Loss: 0.4301
Epoch [2/20], Loss: 0.3624, Validation Loss: 0.3591
Epoch [3/20], Loss: 0.3292, Validation Loss: 0.3502
Epoch [4/20], Loss: 0.3097, Validation Loss: 0.3314
Epoch [5/20], Loss: 0.2833, Validation Loss: 0.3357
Epoch [6/20], Loss: 0.2724, Validation Loss: 0.3160
Epoch [7/20], Loss: 0.2601, Validation Loss: 0.3046
Epoch [8/20], Loss: 0.2422, Validation Loss: 0.3132
Epoch [9/20], Loss: 0.2368, Validation Loss: 0.3139
Epoch [10/20], Loss: 0.2277, Validation Loss: 0.3117
Epoch [11/20], Loss: 0.2292, Validation Loss: 0.3022
Epoch [12/20], Loss: 0.2157, Validation Loss: 0.2972
Epoch [13/20], Loss: 0.2039, Validation Loss: 0.2972
Epoch [14/20], Loss: 0.1898, Validation Loss: 0.3077
Epoch [15/20], Loss: 0.1904, Validation Loss: 0.2889
Epoch [16/20], Loss: 0.1799, Validation Loss: 0.2899
Epoch [17/20], Loss: 0.1749, Validation Loss: 0.3225
Epoch [18/20], Loss: 0.1722, Validation Loss: 0.3440
Epoch [19/20], Loss: 0.1702, Validation Loss: 0.2953
Epoch [20/20], Loss: 0.1628, Validation Loss: 0.2983
```

Test Accuracy: 89.95%

3

در این پارت می‌خواهیم به مقایسه دو بهینه‌ساز Adam و SGD پردازیم و در مدل پیش‌فرض شبکه عمیق و عرضی که در سوال 1 پیاده‌سازی کردیم از بهینه‌ساز Adam استفاده کردیم و همان نتایج را نمایش می‌دهیم و SGD را هم مثل Adam در ادامه پیاده‌سازی می‌کنیم. برای افزایش سرعت همگرایی و کاهش نوسانات در فرایند آموزش از momentum استفاده کردیم و مقدار 0.9 را در نظر گرفتیم و در هر بروزرسانی وزن‌ها، 90 درصد از تغییرات قبلی (گام‌های قبلی) به عنوان نیرو به تغییرات جدید اضافه می‌شود و باعث می‌شود که مدل به طور مؤثرتری به سمت بهینه‌سازی بپردازد. اگر به تصاویر پایین دقت کنیم می‌بینیم که Adam دقت بهتری دارد اما تفاوت این دو بهینه‌ساز جلوگیری می‌شود. آنچنان بالا نیست و می‌توان گفت تقریباً عملکرد مشابهی دارند.

نتایج هر کدام از بهینه‌سازها روی مدل شبکه عمیق (سمت چی‌ها) و مدل شبکه عرضی (سمت راستی‌ها) را می‌توانیم در خروجی‌های زیر هم مشاهده کنیم:

- Adam

```
Epoch [1/20], Loss: 0.6040, Validation Loss: 0.3709
Epoch [2/20], Loss: 0.4475, Validation Loss: 0.3488
Epoch [3/20], Loss: 0.4073, Validation Loss: 0.3194
Epoch [4/20], Loss: 0.3813, Validation Loss: 0.3442
Epoch [5/20], Loss: 0.3602, Validation Loss: 0.3024
Epoch [6/20], Loss: 0.3481, Validation Loss: 0.3003
Epoch [7/20], Loss: 0.3333, Validation Loss: 0.3069
Epoch [8/20], Loss: 0.3208, Validation Loss: 0.2913
Epoch [9/20], Loss: 0.3116, Validation Loss: 0.2876
Epoch [10/20], Loss: 0.2998, Validation Loss: 0.2791
Epoch [11/20], Loss: 0.2900, Validation Loss: 0.2736
Epoch [12/20], Loss: 0.2859, Validation Loss: 0.2774
Epoch [13/20], Loss: 0.2769, Validation Loss: 0.2759
Epoch [14/20], Loss: 0.2678, Validation Loss: 0.2757
Epoch [15/20], Loss: 0.2605, Validation Loss: 0.2664
Epoch [16/20], Loss: 0.2513, Validation Loss: 0.2641
Epoch [17/20], Loss: 0.2465, Validation Loss: 0.2662
Epoch [18/20], Loss: 0.2433, Validation Loss: 0.2714
Epoch [19/20], Loss: 0.2389, Validation Loss: 0.2619
Epoch [20/20], Loss: 0.2313, Validation Loss: 0.2709
```

Test Accuracy: 90.04%

```
Epoch [1/20], Loss: 0.4695, Validation Loss: 0.3882
Epoch [2/20], Loss: 0.3813, Validation Loss: 0.3376
Epoch [3/20], Loss: 0.3478, Validation Loss: 0.3204
Epoch [4/20], Loss: 0.3233, Validation Loss: 0.3453
Epoch [5/20], Loss: 0.3061, Validation Loss: 0.3049
Epoch [6/20], Loss: 0.2937, Validation Loss: 0.3079
Epoch [7/20], Loss: 0.2767, Validation Loss: 0.3173
Epoch [8/20], Loss: 0.2650, Validation Loss: 0.3012
Epoch [9/20], Loss: 0.2542, Validation Loss: 0.2866
Epoch [10/20], Loss: 0.2434, Validation Loss: 0.2958
Epoch [11/20], Loss: 0.2295, Validation Loss: 0.2854
Epoch [12/20], Loss: 0.2276, Validation Loss: 0.2974
Epoch [13/20], Loss: 0.2181, Validation Loss: 0.2884
Epoch [14/20], Loss: 0.2072, Validation Loss: 0.2939
Epoch [15/20], Loss: 0.2050, Validation Loss: 0.2991
Epoch [16/20], Loss: 0.1954, Validation Loss: 0.2899
Epoch [17/20], Loss: 0.1847, Validation Loss: 0.2900
Epoch [18/20], Loss: 0.1829, Validation Loss: 0.2861
Epoch [19/20], Loss: 0.1744, Validation Loss: 0.2969
Epoch [20/20], Loss: 0.1730, Validation Loss: 0.2974
```

Test Accuracy: 90.19%

- SGD

```
Epoch [1/20], Loss: 0.9900, Validation Loss: 0.4806
Epoch [2/20], Loss: 0.5824, Validation Loss: 0.4030
Epoch [3/20], Loss: 0.5110, Validation Loss: 0.3695
Epoch [4/20], Loss: 0.4721, Validation Loss: 0.3602
Epoch [5/20], Loss: 0.4456, Validation Loss: 0.3442
Epoch [6/20], Loss: 0.4263, Validation Loss: 0.3335
Epoch [7/20], Loss: 0.4095, Validation Loss: 0.3283
Epoch [8/20], Loss: 0.3942, Validation Loss: 0.3221
Epoch [9/20], Loss: 0.3873, Validation Loss: 0.3118
Epoch [10/20], Loss: 0.3753, Validation Loss: 0.3175
Epoch [11/20], Loss: 0.3662, Validation Loss: 0.3086
Epoch [12/20], Loss: 0.3541, Validation Loss: 0.3046
Epoch [13/20], Loss: 0.3469, Validation Loss: 0.3058
Epoch [14/20], Loss: 0.3479, Validation Loss: 0.3005
Epoch [15/20], Loss: 0.3351, Validation Loss: 0.3018
Epoch [16/20], Loss: 0.3288, Validation Loss: 0.2990
Epoch [17/20], Loss: 0.3269, Validation Loss: 0.3004
Epoch [18/20], Loss: 0.3191, Validation Loss: 0.2934
Epoch [19/20], Loss: 0.3162, Validation Loss: 0.2938
Epoch [20/20], Loss: 0.3054, Validation Loss: 0.2910
```

Test Accuracy: 89.75%

```
Epoch [1/20], Loss: 0.5884, Validation Loss: 0.4191
Epoch [2/20], Loss: 0.4272, Validation Loss: 0.3767
Epoch [3/20], Loss: 0.3902, Validation Loss: 0.3576
Epoch [4/20], Loss: 0.3645, Validation Loss: 0.3489
Epoch [5/20], Loss: 0.3475, Validation Loss: 0.3334
Epoch [6/20], Loss: 0.3317, Validation Loss: 0.3259
Epoch [7/20], Loss: 0.3185, Validation Loss: 0.3255
Epoch [8/20], Loss: 0.3073, Validation Loss: 0.3121
Epoch [9/20], Loss: 0.2974, Validation Loss: 0.3091
Epoch [10/20], Loss: 0.2885, Validation Loss: 0.3020
Epoch [11/20], Loss: 0.2792, Validation Loss: 0.3024
Epoch [12/20], Loss: 0.2722, Validation Loss: 0.2969
Epoch [13/20], Loss: 0.2632, Validation Loss: 0.2981
Epoch [14/20], Loss: 0.2571, Validation Loss: 0.2932
Epoch [15/20], Loss: 0.2487, Validation Loss: 0.2898
Epoch [16/20], Loss: 0.2430, Validation Loss: 0.2914
Epoch [17/20], Loss: 0.2364, Validation Loss: 0.2831
Epoch [18/20], Loss: 0.2312, Validation Loss: 0.2916
Epoch [19/20], Loss: 0.2258, Validation Loss: 0.2883
Epoch [20/20], Loss: 0.2189, Validation Loss: 0.2805
```

Test Accuracy: 90.04%