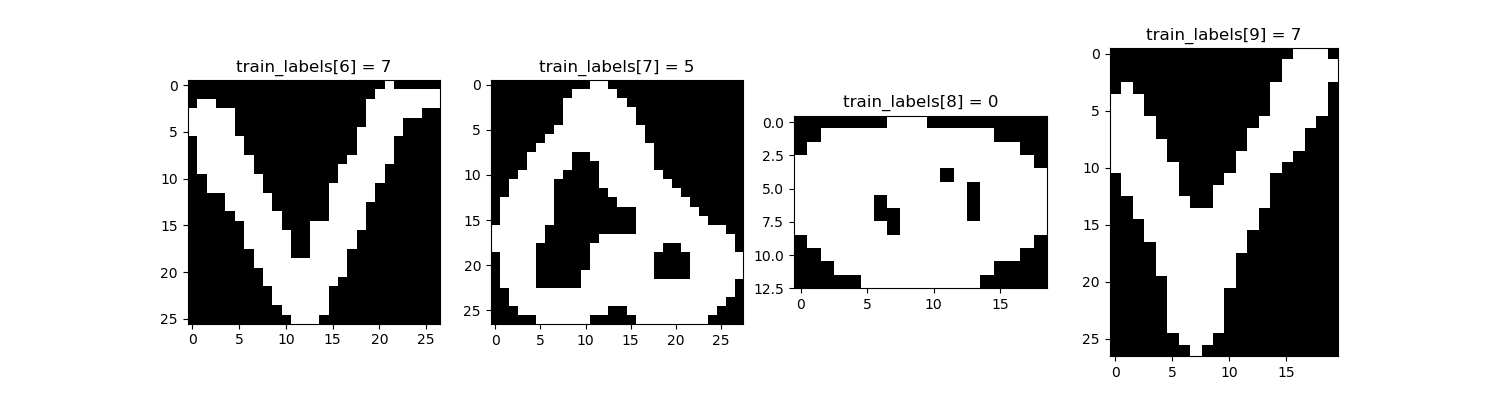
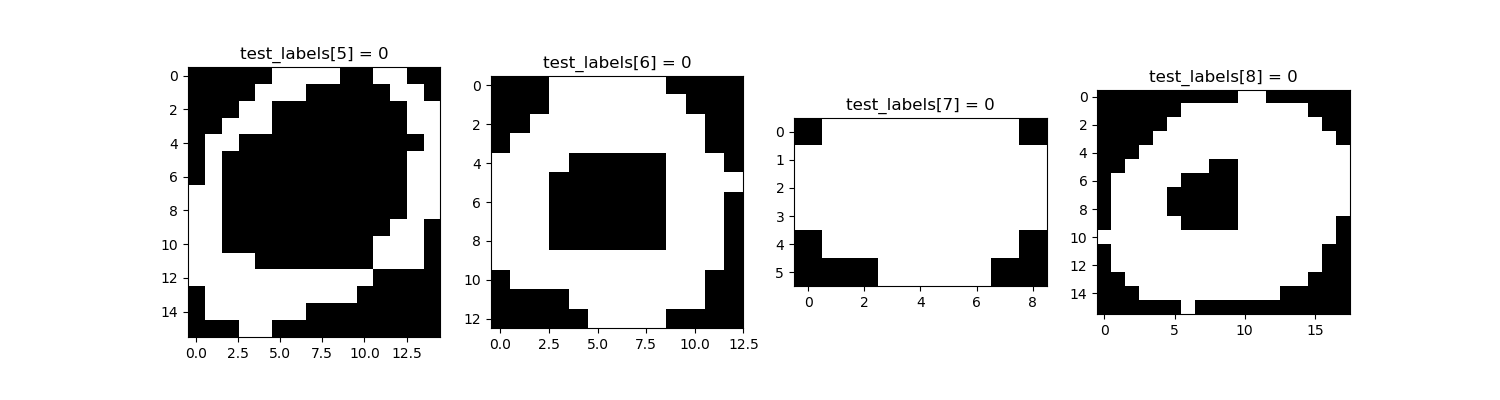
1:

a)



b)



precision recall f1-score support

0 0.99 0.98 0.99 2000

1 0.98 0.99 0.99 2000

2 0.96 0.95 0.96 2000

3 0.94 0.96 0.95 2000

4 0.97 0.97 0.97 2000

5 0.98 0.99 0.98 2000

6 0.95 0.98 0.97 2000

7 0.99 0.99 0.99 2000

8 0.99 0.98 0.99 2000

9 0.99 0.95 0.97 2000

accuracy 0.97 20000

macro avg 0.97 0.97 0.97 20000

weighted avg 0.97 0.97 0.97 20000

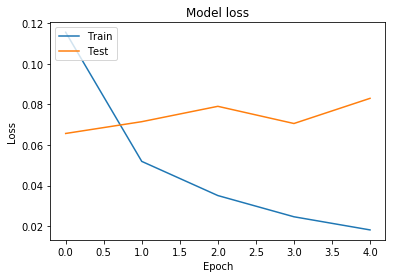
Test loss: 0.14845385479258824

Test accuracy: 0.97385

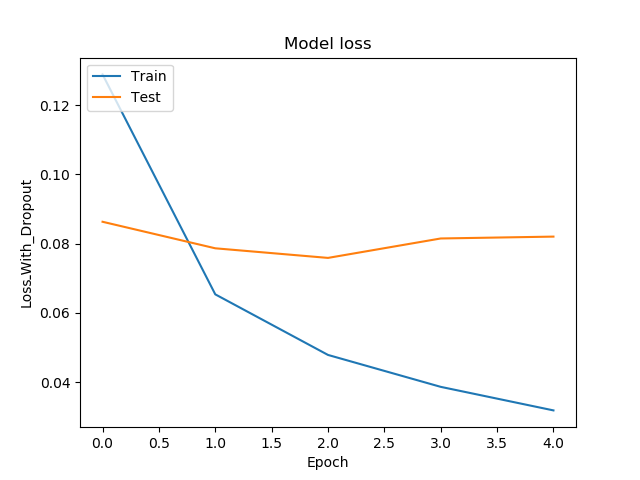
Test f1\_score: 0.973903092288971

Test precision: 0.9742692470550537

Test recall: 0.97355

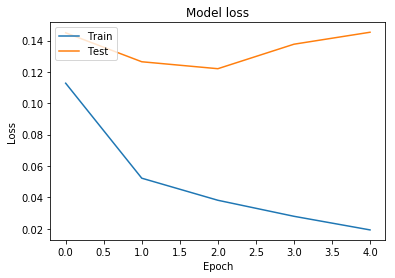


c)



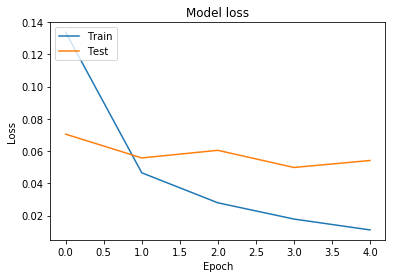
Dropout تعدادی از خروجی های یک لایه را صفر می کند که باعث می شود به مدل نویز اضافه شود که در نهایت منجر به بهتر جنرالایزشدن آن می شود و در نتیجه از overfitting جلوگیری می کند.

d)



در حالت عادي دسته بندها داده ها را به دو دسته داده هاي آموزش و داده هاي تست تقسیم مي كنند كه در آن داده هاي آموزش براي فرآيند يادگیري استفاده مي شوند و از داده هاي تست براي آزمايش دسته بند استفاده مي شود. اما در بعضي اوقات مخصوصا هنگام كار با شبكه هاي عصبي داده ها به سه دسته تقسیم مي شوند يعني در آن داده هاي اعتبارسنجي هم اضافه ميشوند. اگر فرايند آموزش خیلي طولاني باشد، پیش پردازش پیش مي آيد يعني شبكه خیلي به داده هاي آموزش حساس مي شود و اگر داده هاي جديد كمي متفاوت باشند، نتیجه ي دقیقي حاصل نمي شود. به همین دلیل در اين مقاله داده ها به سه دسته ي آموزش، اعتبارسنجي و آزمايش تقسیم مي شوند. اهمیت داده هاي اعتبارسنجي اين است كه از وقوع پیش پردازش جلوگیري ميكند.زماني كه فرايند آموزش توسط داده هاي بخش آموزش انجام مي گیرد، توسط داده ها يا اعتبارسنجي بررسي مي كنیم كه سیستم خیلي وابسته به داده هاي آموزش نباشد. هر چه تعداد اپوك ها بیشتر شود، میزان خطاي آموزش كاهش مي يابد اما به نقطه اي مي رسیم كه كم كم خطاي اعتبارسنجي افزايش مي يابد؛ اين نقطه همان جايي است كه ممكن است از آن به بعد پیش پردازش اتفاق بیافتد به همین دلیل فرايند آموزش در آن متوقف مي شود. می توان هم از validation\_data در متدfit استفاده کرد که در آن صورت این پارامتر یک دوتایی ازداده ها و برچسب هایشان می گیرد و همین طور می توان از validation\_split در متدfit استفاده کرد که یک عدد بین ۰ و ۱ می گیرد که این عدد نشان می دهد چه نسبتی از داده های برای آموزش برای این کار استفاده شوند

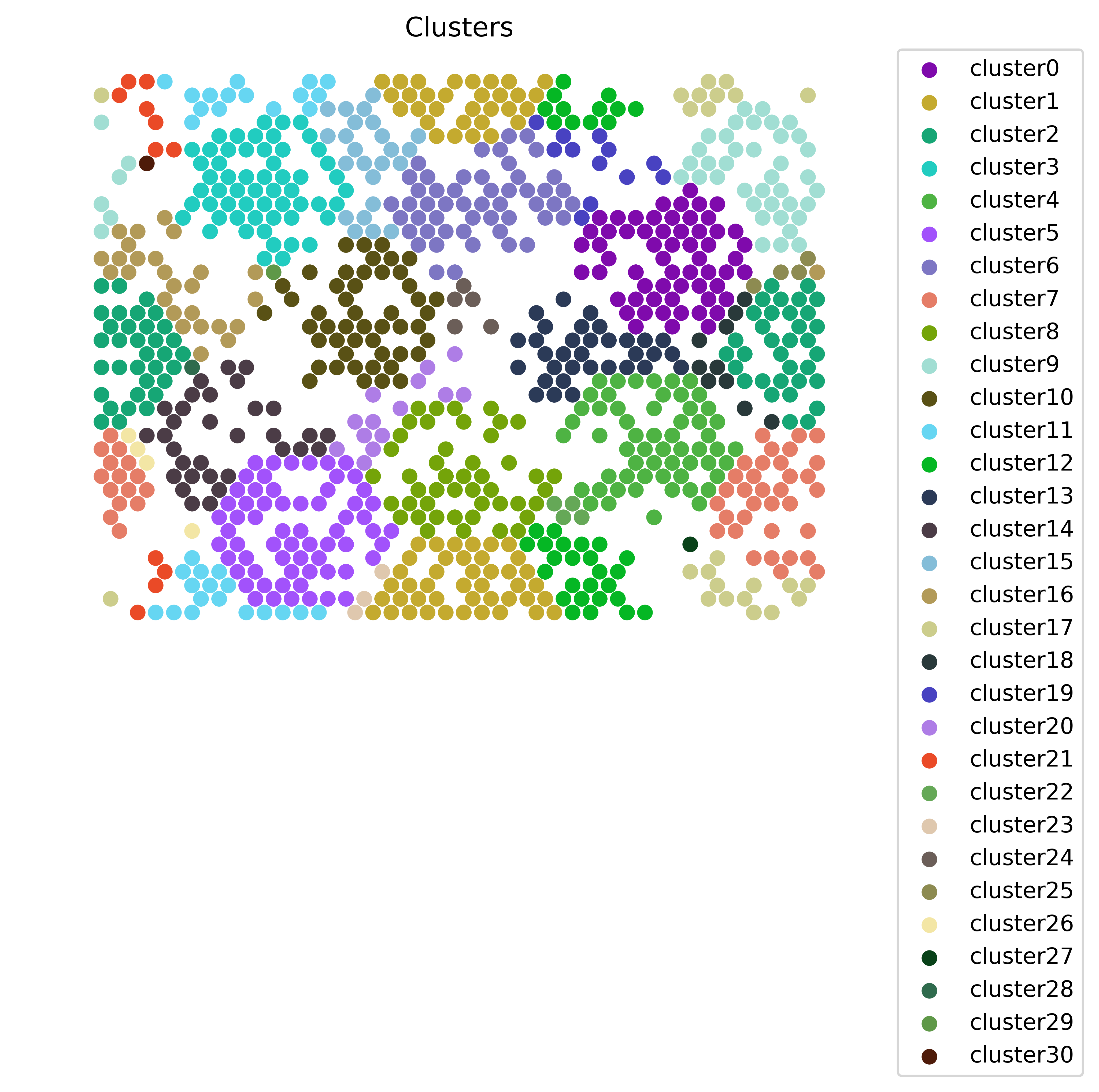
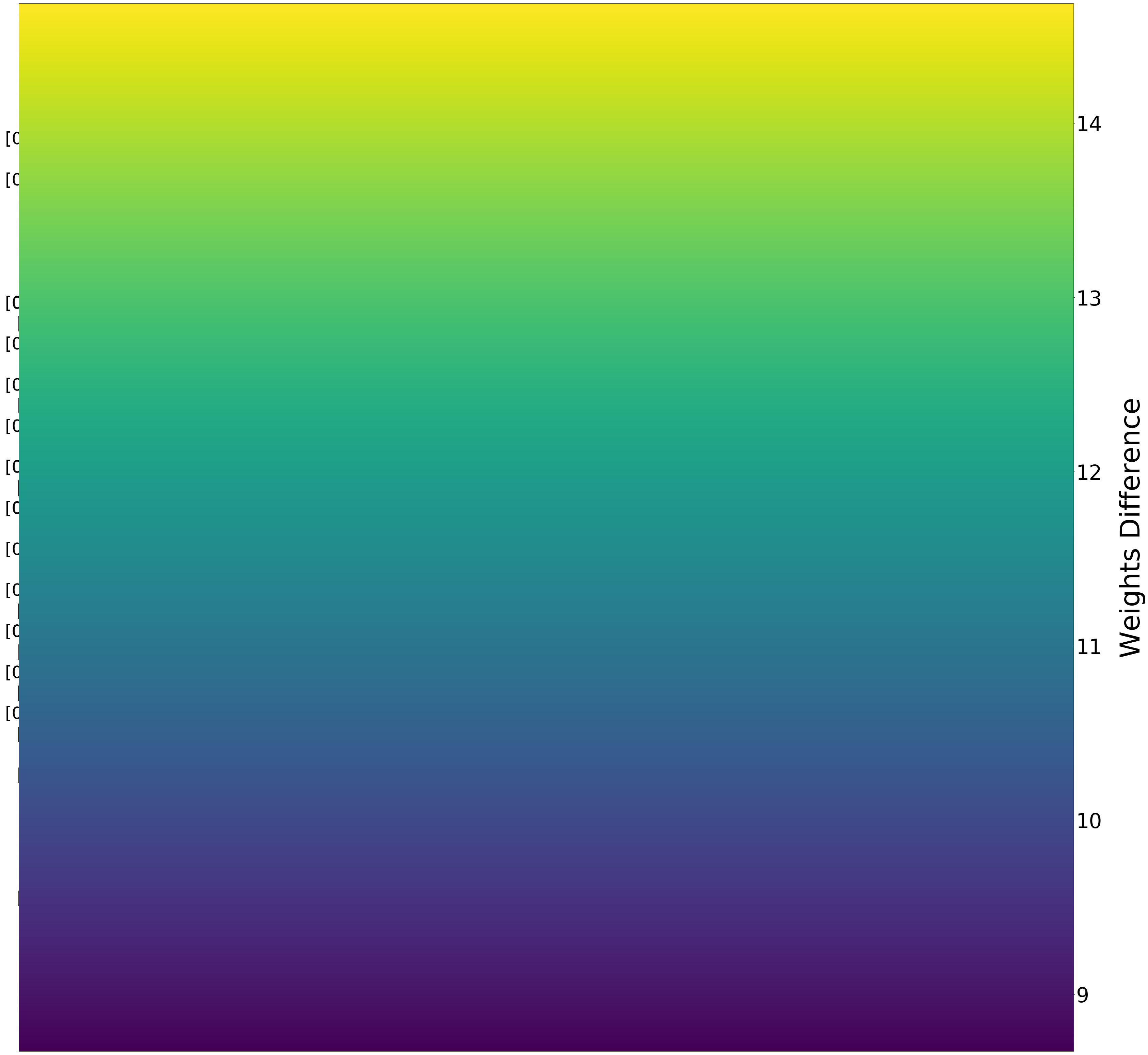
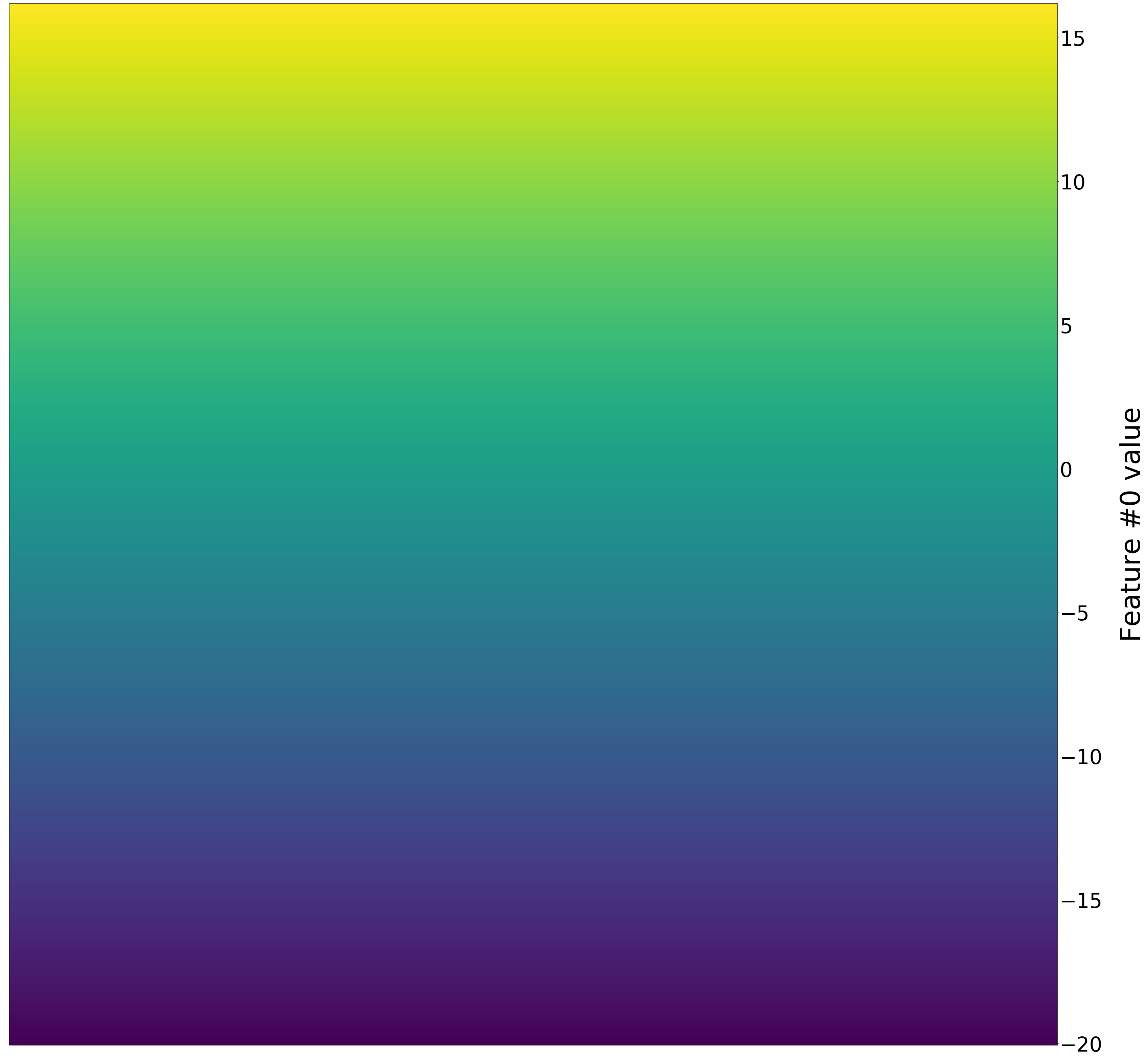
e)



در هر مرحله(step) از یادگیری تعدادی از نمونه ها را آموزش داده و پس از آن پارامترهای شبکه تنظیم می شود به این تعداد نمونه batchsize گفته می شود.البته انتخابbatchsize هم باید دقت کنید که نه زیاد بزرگ باشه و نه زیاد کوچک .از طرفی از batchsize برابر کل نمونه های آموزش باشد گرادیان پایدارتر می شود و شبکه به کندی همگرا می شود و از طرف دیگر اگر batchsize خیلی کوچک انتخاب شود گرادیان ناپایدار خواهید داشت که به طبع آن مجبورید نرخ یادگیری را کاهش دهید. Batch mode به زمانی گفته می شود که batch size به اندازه کل داده های آموزش باشد.Mini-batch mode به حالتی گفته می شود کهbatch size از اندازه کل داده های آموزش کمتر باشد.برای زمانی که تعداد داده های آموزش خیلی زیاد است. Stochastic mode به حالتی گفته می شود که batch size برابر ۱ است.

نتیجه گیری نهایی: برای غلبه بر مشکل overfitting که در شبکه های بزرگ با تعداد داده های زیاد رخ می دهد بهتر است در هر epoch داده های برای validation را عوض کنیم و نسبتی مناسب بعد از آموزش داده ها پارامترهای شبکه را عوض کنیم. همین طور استفاده از dropout هم راهگشا است.

2:



3:

