



Mohammad Javad Ranjbar

810101173

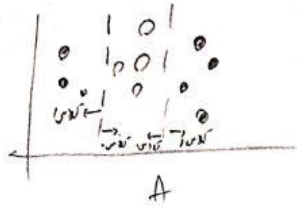
HW3

Machine Learning, Fall 2022

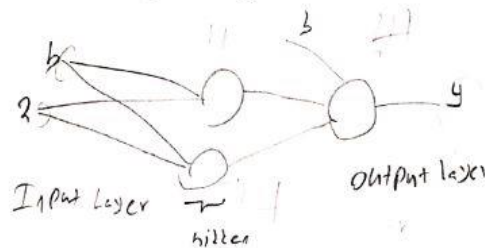
مهد جواد رشید نمرین سوم یادگیری ماشین ۸۱۰۱۱۷۳

۱-

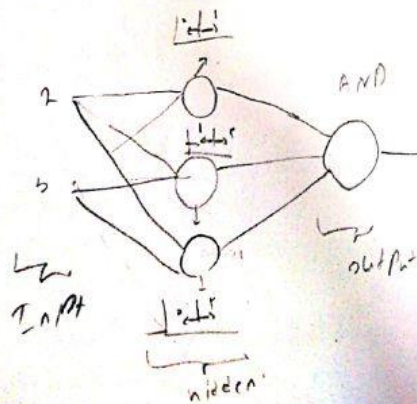
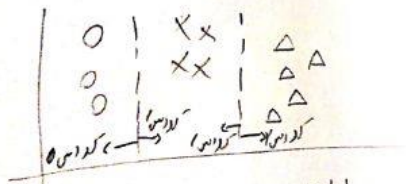
الف) با یک



با شبکه ای به صورت $2 \rightarrow 1 \rightarrow 2$ می توان دو نوع داده نهضی جدا ای شود را جدا کرد $\square \times \times$ حال
در این حای نیاز به دو خط جدا ساز داریم \triangle و این دو خط جدا ساز باید مشخص شود هر مفی
جهت کلاسی است و با هم کار کنند پس مثل نهایی به صورت زیر خلاصه بود



مکمل به مفی نیاز داریم به عبارتی ۳ نود در ۱ اگر دردی هم حساب کنیم ۵ نود در ۱



به یک شبکه ممی با یک لایه مفی و ۴ نود در ۱ (۴ نود در ۱)
نیاز داریم

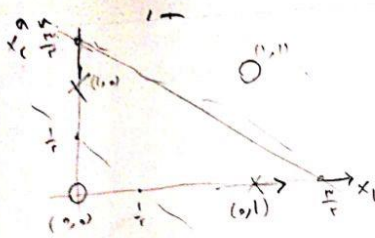
(ب)

(A) با استفاده از گیت NAND می‌باشد

x_1	x_2	y
0	0	$1 = f(0, 0)$
0	1	$1 = f(0, 1)$
1	0	$1 = f(1, 0, 1)$
1	1	$0 = f(1, 1, 2)$

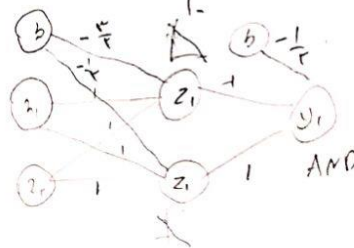
(B) گیت NOR می‌باشد

x_1	x_2	y
0	0	$f(0, 0) = 1$
0	1	$f(0, 1) = 0$
1	0	$f(1, 0) = 0$
1	1	$f(1, 1) = 0$



وزن لاگرانژ $\lambda_1 + \lambda_2 - \frac{1}{r} = 0$ خط اول
 $\lambda_1 + \lambda_2 - \frac{1}{r} = 0$ خط دوم

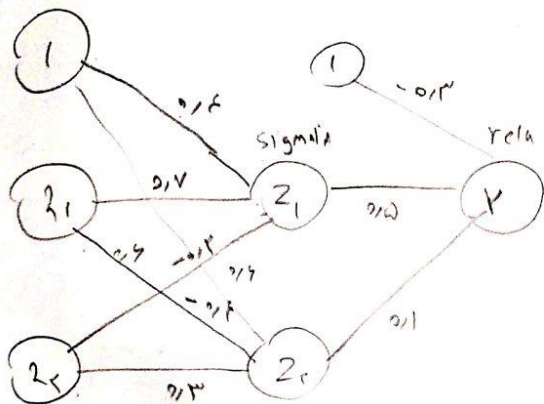
(ج) دو خط برای جداسازی نیاز داریم



همچنین از تابع Step استفاده می‌کنیم
 حال برای مقادیر مختلف λ_1 و λ_2 می‌کنیم

x_1	x_2	z_1	z_2	z_3	z_4	y
0	0	$0+0-\frac{1}{r} = -\frac{1}{r}$	$0+0-\frac{1}{r} = -\frac{1}{r}$	0	0	$0+0-\frac{1}{r} = -\frac{1}{r}$
0	1	$0+1-\frac{1}{r} = 1-\frac{1}{r}$	$0+1-\frac{1}{r} = 1-\frac{1}{r}$	0	1	$1+0-\frac{1}{r} = 1-\frac{1}{r}$
1	0	$1+0-\frac{1}{r} = 1-\frac{1}{r}$	$1+0-\frac{1}{r} = 1-\frac{1}{r}$	0	1	$1+0-\frac{1}{r} = 1-\frac{1}{r}$
1	1	$1+1-\frac{1}{r} = 2-\frac{1}{r}$	$1+1-\frac{1}{r} = 2-\frac{1}{r}$	1	1	$1+1-\frac{1}{r} = 2-\frac{1}{r}$

(5)



$$x = 0.1$$

$$Z_1 = (0.1 \cdot x_1 + 0.1 \cdot x_2 + 0.1) \Rightarrow Z_1 = \text{sigmoid}(Z_1) \quad Z_1(0.1) = 0.1 \cdot 1 + 0.1 \cdot 1 = 0.2 \Rightarrow Z_1 = 0.2$$

$$Z_2 = (0.1 \cdot x_1 + 0.1 \cdot x_2 + 0.1) \quad Z_2 = \text{sigmoid}(Z_2) \quad Z_2(0.1) = -0.1 + 0.1 = 0 \Rightarrow Z_2 = 0.2$$

$$Y = 0.2 \cdot Z_1 + 0.1 \cdot Z_2 + 0.1 = 0.1$$

$$Y(0.1) = 0.1 \quad \Rightarrow \quad J = \frac{1}{2} (y - \text{target})^2 \Rightarrow J(0.1) = \frac{1}{2} (0.1 - 0)^2 = \frac{0.01}{2}$$

$$\frac{\partial J}{\partial V} = \frac{\partial}{\partial V} \frac{1}{2} (V^T Z - \text{target})^2 = Z^T (V^T Z - \text{target})$$

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial Z} \times \frac{\partial Z}{\partial W} = V (Z - \text{target}) \times \text{Sigmoid}'(Z) \times X$$

$$V = V - \alpha \frac{\partial J}{\partial V} = \begin{bmatrix} -0.7 \\ 0.2 \\ 0.1 \end{bmatrix} - 0.1 \times \begin{bmatrix} 1 \\ 0.42 \\ 0.00 \end{bmatrix} \times 0.17 = \begin{bmatrix} -0.324 \\ 0.18 \\ 0.107 \end{bmatrix}$$

$$W_{01}^* = W_{01} - V_1 \times \text{Error} \times \text{Sigmoid}'(Z_1) \times X_{10} = 0.1 - 0.1 \times 0.17 \times 0.17 \times 0.17 = 0.11$$

$$W_{02}^* = W_{02} - V_2 \times \text{Error} \times \text{Sigmoid}'(Z_2) \times X_{20} = 0.4 - 0.1 \times 0.17 \times 0.17 \times 0.17 = 0.40$$

$$W_{11}^* = W_{11} - V_1 \times \text{Error} \times \text{Sigmoid}'(Z_1) \times X_{11} = 0.49$$

$$W_{12}^* = W_{12} - V_2 \times \text{Error} \times \text{Sigmoid}'(Z_2) \times X_{21} = -0.17$$

$$W_{21}^* = W_{21} - V_1 \times \text{Error} \times \text{Sigmoid}'(Z_1) \times X_{12} = 0.17$$

$$W_{22}^* = W_{22} - V_2 \times \text{Error} \times \text{Sigmoid}'(Z_2) \times X_{22} = -0.21$$

۳ - تحلیل به ندرت: Y کا ضد ۴ تابلو

ارزان

$$H(Y) = \frac{4}{10} \log \frac{10}{4} - \frac{6}{10} \log \frac{10}{6} = 0.97$$

$$I(Y; x_i) = H(Y) - H(Y|x_i)$$

$$H(Y|x_i) = + \frac{2}{10} H(Y|x_i = \text{سر}) + \frac{4}{10} H(Y|x_i = \text{نبرد}) + \frac{4}{10} H(Y|x_i = \text{آب})$$

$$H(Y|x_i = \text{سر}) = - \frac{2}{10} \log \frac{10}{2} - \frac{8}{10} \log \frac{10}{8} = 0$$

$$H(Y|x_i = \text{نبرد}) = - \frac{1}{4} \log \frac{4}{1} - \frac{3}{4} \log \frac{4}{3} = 0.81$$

$$H(Y|x_i = \text{آب}) = - \frac{1}{4} \log \frac{4}{1} - \frac{3}{4} \log \frac{4}{3} = 0.81$$

$$H(Y|x_i) = 0.81 \times \frac{4}{10} \times 2 + 0 = 0.648 \Rightarrow I(Y, x_i) = 0.97 - 0.648 = 0.322$$

$$H(Y) = \frac{4}{10} H(Y|x_i = \text{نقی}) + \frac{6}{10} H(Y|x_i = \text{جس})$$

$$H(Y|x_i = \text{نقی}) = - \frac{4}{10} \log \frac{10}{4} - \frac{6}{10} \log \frac{10}{6} = 0.81$$

$$H(Y|x_i = \text{جس}) = - \frac{5}{6} \log \frac{6}{5} - \frac{1}{6} \log \frac{6}{1} = 0.65$$

$$H(Y) = \frac{4}{10} \times 0.81 + \frac{6}{10} \times 0.65 = 0.71 \Rightarrow I(Y, x_i) = 0.97 - 0.71 = 0.26$$

$$H(Y) = \frac{4}{10} H(Y|x_i = \text{ارزان قیمت}) + \frac{6}{10} H(Y|x_i = \text{گران قیمت})$$

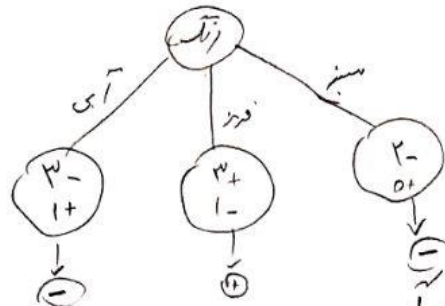
$$H(Y|x_i = \text{ارزان قیمت}) = - \frac{1}{4} \log \frac{4}{1} - \frac{3}{4} \log \frac{4}{3} = 0.81$$

$$H(Y|x_i = \text{گران قیمت}) = - \frac{2}{5} \log \frac{5}{2} - \frac{3}{5} \log \frac{5}{3} = 0.91$$

$$H(Y) = \frac{4}{10} \times 0.81 + \frac{6}{10} \times 0.91 = 0.87 \Rightarrow I(Y, \text{قیمت}) = 0.97 - 0.87 = 0.1$$

$$\arg \max (I(x, z_i)) = 0.22 \leftarrow \text{پس رنگ بهترین و بزرگ است}$$

درخت تصمیم به صورت زیر خواهد بود



اگر به اساس بهترین ای در هر برگ تصمیم بگیریم

$$\text{Accuracy} = \frac{3+3+2}{4+4+2} \cdot \frac{1}{10}$$

$$\text{Sensitivity} = \frac{3}{3+2} = \frac{3}{5}$$

$$\text{Precision} = \frac{3}{3+1} = \frac{3}{4}$$

$$\text{Specificity} = \frac{5}{5+1} = \frac{5}{6}$$

اگر اطراف نمودار بهترین لیل کارافرد باشد بصورت زیر انتخاب می شود

$$\sum_{j=0}^M \underbrace{P(w|x)}_{\frac{1}{r^i}} \underbrace{(1 - P(w+x))^{M-i}}_{\frac{1}{r^{n-i}}} \binom{n}{i}$$

اگر کار فرد باشد و ما توجه به برابری خاص بین خواص راس

$$\sum_{j=0}^{\frac{K(K-1)}{r}} \binom{n}{i} \times \frac{1}{r^i} \times \frac{1}{r^{n-i}} = \frac{1}{r^n} \sum_{j=0}^{\frac{K(K-1)}{r}} \binom{n}{j}$$

(ب) کافی است اما را قائله ای کنیم

$$P_{K=1}(e) = \frac{1}{r^n} \sum_{j=0}^{\frac{K(K-1)}{r}} \binom{n}{j} = \frac{1}{r^n} \quad P_{K=0} = \frac{1}{r^n} \sum_{j=0}^{\frac{K(K-1)}{r}} \binom{n}{j} \Rightarrow \sum \left(\frac{1}{r} \right)^j > 1$$

↑ مشخص است که هر چه بزرگتر K باشد، بیشتر است

$$\lim_{n \rightarrow \infty} P_n(x) = \lim_{n \rightarrow \infty} \frac{1}{r^n} \sum_{j=1}^{K-1} \frac{n!}{(n-j)!} = \lim_{n \rightarrow \infty} \frac{1}{r^n} \left(\frac{K-1}{r} \right) = \lim_{n \rightarrow \infty} \frac{K-1}{r^{n+1}} = 0 \quad (2)$$

$$v_n = \int e^{\frac{(x-x_i)}{h_n}} = h_n \quad (3)$$

قلت اول : $a, 1, 2, \dots, n$ و $x - x_i$

$$P_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} \phi\left(\frac{x - x_i}{h_n}\right) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} e^{-\frac{(x-x_i)}{h_n}}$$

$$\bar{P}_n(x) = E(P_n(x)) = \frac{1}{n} \sum_{i=1}^n E\left(\frac{1}{h_n} e^{-\frac{(x-x_i)}{h_n}}\right) = \int_a^a \frac{1}{ah_n} e^{-\frac{x-x_i}{h_n}} dx$$

$$\bar{P}_n(x) = \frac{1}{a} e^{\frac{x}{h_n}} \left(e^{\frac{a}{h_n}} - 1 \right)$$

قلت : $x < 0$ و $x > 0$ و $x = 0$ في $P_n(x)$ و $\bar{P}_n(x)$
 قلت : $a < x < a$

$$\bar{P}_n(x) = \int_0^a \frac{1}{ah_n} e^{-\frac{(x-y)}{h_n}} dy = \frac{1}{a} e^{-\frac{x}{h_n}} \left(e^{\frac{a}{h_n}} - 1 \right) = \frac{1}{a} (1 - e^{-\frac{a}{h_n}})$$

بس معادلتين را نقسم على بعضهما

$$\bar{P}_n(x) = \begin{cases} \frac{1}{a} (1 - e^{-\frac{a}{h_n}}) & x < a \\ \frac{1}{a} (e^{\frac{a}{h_n}} - 1) e^{-\frac{x}{h_n}} & a \leq x \end{cases}$$

$$\frac{P_n(x) - \bar{P}_n(x)}{P_n(x)} < 0.1 \Leftrightarrow \frac{\frac{1}{a} e^{-\frac{x}{h_n}}}{\frac{1}{a}} < 0.1 \Rightarrow h_n \leq \frac{0.1 a}{\ln(1.1)}$$

سوال ۶:

الف) دیتاست را به دو بخش آموزش و تست تقسیم می‌کنیم داده‌ی آموزش دارای ۵۰۰۰۰ داده و داده‌ی تست دارای ۱۰۰۰۰ داده می‌باشد.

```
(x_train, y_train), (x_test, y_test) = cifar10.load_data()
```

همینطور در مرحله‌ی آموزش دو دهم از داده‌ی آموزش را برای ارزیابی انتخاب می‌کنیم.

```
history = model.fit(x_train, y_train, batch_size=batch_size, epochs=epochs, callbacks=[CB], validation_split=0.2)
```

همینطور هرکدام از لیبیل‌ها را تبدیل به one hot vector می‌کنیم یعنی به ازای ایندکس هرکلاسی که لیبیل به آن متعلق باشد مقدار آن کلاس برابر با یک و سایر کلاس‌ها برابر با صفر خواهد بود.

معماری شبکه به صورت زیر می‌باشد:

علت انتخاب این پارامترهای شبکه معمولاً با سعی و خطا صورت می‌گیرد من با استفاده از شبکه پیشنهادی در خود سایت تئوسورفلو برای این وظیفه این شبکه را انتخاب کردم.

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 32, 32, 32)	896
conv2d_5 (Conv2D)	(None, 30, 30, 32)	9248
max_pooling2d_2 (MaxPooling2D)	(None, 15, 15, 32)	0
dropout_3 (Dropout)	(None, 15, 15, 32)	0
conv2d_6 (Conv2D)	(None, 15, 15, 64)	18496
conv2d_7 (Conv2D)	(None, 13, 13, 64)	36928
max_pooling2d_3 (MaxPooling2D)	(None, 6, 6, 64)	0
dropout_4 (Dropout)	(None, 6, 6, 64)	0
flatten_1 (Flatten)	(None, 2304)	0
dense_2 (Dense)	(None, 128)	295040
dropout_5 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 10)	1290
Total params: 361,898		
Trainable params: 361,898		
Non-trainable params: 0		

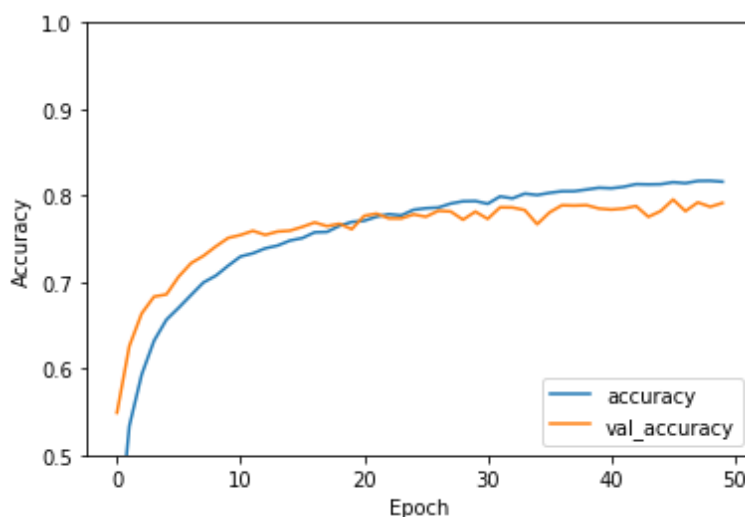
در لایه آخر ۱۰ نورون (به ازای هر کلاس یک نورون) قرار داده‌ایم به همراه تابع فعال ساز softmax تا بتوانیم به سادگی با argmax گرفتن کلاس را انتخاب کنیم. از آنجا که داده‌ی ما کتگوریکال می‌باشد و همچنین تعداد کلاس‌ها بیشتر از دوتا می‌باشد از categorical_crossentropy استفاده می‌کنیم.

نرخ یادگیری: می‌توان نرخ یادگیری را به صورت دیفالت مقداری که خود مدل انتخاب می‌کند را داشته باشد.

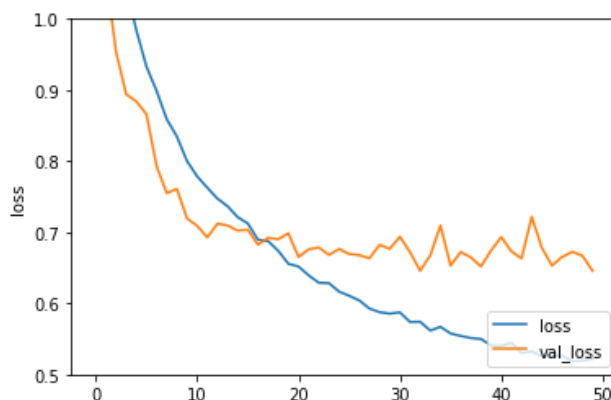
در مورد نرخ یادگیری باید دقت کرد که در صورتی که خیلی بزرگ انتخاب شود ممکن است شبکه همگرا نشود یا در صورتی که خیلی کوچک انتخاب شود ممکن است سرعت همگرایی به شدت کم بوده باشد و یا شبکه همگرا نشود. در اینجا برای تست مقدار ۰/۰۰۱ را برای آموزش انتخاب کرده‌ایم با نرخ کم شدن ۱۰ به توان منهای شش که در طول آموزش مقدار این نرخ یادگیری کاهش یابد.

بهینه‌ساز: چندین بهینه‌ساز مانند SGD یا Adams یا RMSprop می‌توان استفاده کرد، در بیشتر مواقع بهینه‌ساز Adams بهترین و سریع‌ترین جواب را می‌دهد. در اینجا از RMSprop و Adams استفاده کردیم. که Adams سریعتر همگرا شد و در نهایت نتیجه‌ی بهتری می‌دهد.

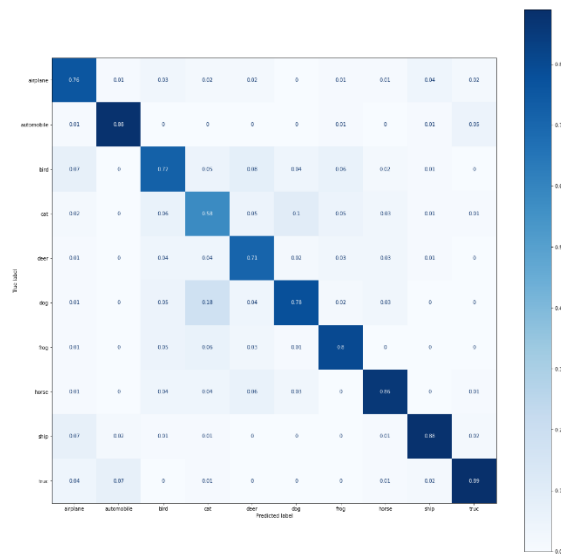
برای جلوگیری از استفاده از مدل overfit شده در هر مرحله با استفاده از call back مدل با کمتری مقدار loss را ذخیره می‌کنیم. نمودار دقت برای Adams:



نمودار loss برای Adams:

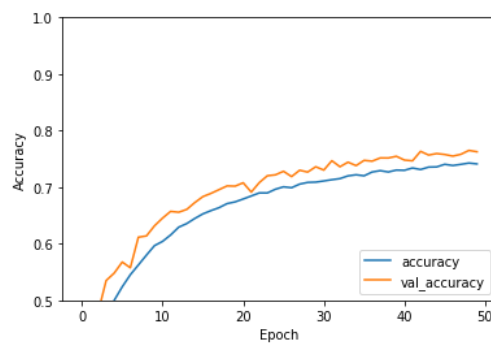


ماتریس درهم‌ریختگی برای Adams:

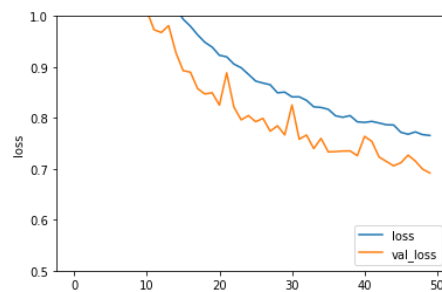


نمودار دقت برای RMSprop:

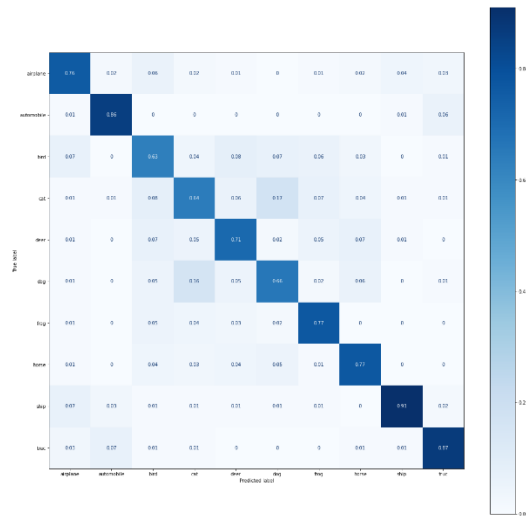
البته از آنجا که در نمودار زیر نیز مشخص است، شبکه به صورت کامل آموزش ندیده است، و هنوز به overfit شدن نزدیک نشده تا آموزش را متوقف کنیم (کولب مدت زمان بیشتری اجازه استفاده از GPU را نداد)



نمودار لاس برای RMSprop:



ماتریس در هم ریختگی برای RMSprop:



سوال ۷:

Slope= 0.6142214557837571 Intercept= [-0.25197169]
 <Figure size 432x288 with 0 Axes>

