Μηχανική Μάθηση: Εργασία 1

Τράκας Χριστόφορος - 318085

1 Μερικές παράγωγοι του πίνακα $W^{(1)}$

Θεωρούμε πως έχουμε το μοντέλο:

$$s^{(1)} = XW^{(1)}$$

$$Z = h(s^{(1)})$$

$$s^{(2)} = ZW^{(2)}$$

$$y = softmax(s^{(2)})$$

$$E = CEloss(y, t)$$

Η ποσότητα που ψάχνουμε είναι η :

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial s^{(2)}}\frac{\partial s^{(2)}}{\partial z}\frac{\partial z}{\partial s^{(1)}}\frac{\partial s^{(1)}}{\partial W^{(1)}} \tag{1}$$

Ενώ είναι:

$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial s^{(2)}}\frac{\partial s^{(2)}}{\partial W^{(2)}} \tag{2}$$

Ισχύει $\frac{\partial s^{(2)}}{\partial W^{(2)}} = Z$ και αφού μας δίνεται ότι $\frac{\partial E}{\partial W^{(2)}} = (T-Y)^T Z - \lambda W^{(2)}$ τότε απο την (1) :

$$\frac{\partial E}{\partial y}\frac{\partial y}{\partial s^{(2)}} = (T - Y) \tag{3}$$

Επίσης είναι:

$$\frac{\partial s^{(2)}}{\partial z} = W^{(2)} \tag{4}$$

$$\frac{\partial z}{\partial s^{(1)}} = h'(s^{(1)}) \tag{5}$$

$$\frac{\partial s^{(1)}}{\partial W^{(1)}} = X \tag{6}$$

Απο τις σχέσεις (3), (4), (5), (6) η (1) γίνεται

$$\frac{\partial E}{\partial W^{(1)}} = ((T - Y)W^{(2)}Z * h'(s^{(1)}))^T X$$

2 Αποτελέσματα εφαρμογής του μοντέλου. (Results.csv με Batch Size=100)

a. Για το Dataset MNIST:

1	DataSet	Hidden Layers (M)	Activation Function	epochs	Learning rate	Regularization	Error
2	mnist	100	softplus	10	0.01	0.1	0.0705
3	mnist	100	softplus	10	0.001	0.1	0.1002
4	mnist	100	softplus	20	0.01	0.1	0.0608
5	mnist	100	softplus	20	0.001	0.1	0.0902
6	mnist	100	tanh	10	0.01	0.1	0.0577
7	mnist	100	tanh	10	0.001	0.1	0.0883
8	mnist	100	tanh	20	0.01	0.1	0.0458
9	mnist	100	tanh	20	0.001	0.1	0.0815
10	mnist	100	cos	10	0.01	0.1	0.0454
11	mnist	100	cos	10	0.001	0.1	0.0921
12	mnist	100	cos	20	0.01	0.1	0.033
13	mnist	100	cos	20	0.001	0.1	0.0781
14	mnist	200	softplus	10	0.01	0.1	0.0687
15	mnist	200	softplus	10	0.001	0.1	0.0952
16	mnist	200	softplus	20	0.01	0.1	0.0591
17	mnist	200	softplus	20	0.001	0.1	0.0875
18	mnist	200	tanh	10	0.01	0.1	0.054
19	mnist	200	tanh	10	0.001	0.1	0.0843
20	mnist	200	tanh	20	0.01	0.1	0.0396
21	mnist	200	tanh	20	0.001	0.1	0.0789
22	mnist	200	cos	10	0.01	0.1	0.0368
23	mnist	200	cos	10	0.001	0.1	0.0868
24	mnist	200	cos	20	0.01	0.1	0.029
25	mnist	200	cos	20	0.001	0.1	0.0662
26	mnist	300	softplus	10	0.01	0.1	0.0692
27	mnist	300	softplus	10	0.001	0.1	0.0937
28	mnist	300	softplus	20	0.01	0.1	0.0566
29	mnist	300	softplus	20	0.001	0.1	0.0848
30	mnist	300	tanh	10	0.01	0.1	0.0516
31	mnist	300	tanh	10	0.001	0.1	0.0839
32	mnist	300	tanh	20	0.01	0.1	0.0369
33	mnist	300	tanh	20	0.001	0.1	0.0776
34	mnist	300	cos	10	0.01	0.1	0.0337
35	mnist	300	cos	10	0.001	0.1	0.074
36	mnist	300	cos	20	0.01	0.1	0.0261
37	mnist	300	cos	20	0.001	0.1	0.0596

b. Για το Dataset CIFAR-10:

1	DataSet	Hidden Layers (M)	Activation Function	epochs	Learning rate	Regularization	Error
2	cifar	100	softplus	10	0.01	0.1	0.5764
3	cifar	100	softplus	10	0.001	0.1	0.6024
4	cifar	100	softplus	20	0.01	0.1	0.5407
5	cifar	100	softplus	20	0.001	0.1	0.5877
6	cifar	100	tanh	10	0.01	0.1	0.564
7	cifar	100	tanh	10	0.001	0.1	0.5719
8	cifar	100	tanh	20	0.01	0.1	0.5708
9	cifar	100	tanh	20	0.001	0.1	0.5634
10	cifar	100	cos	10	0.01	0.1	0.5998
11	cifar	100	cos	10	0.001	0.1	0.566
12	cifar	100	cos	20	0.01	0.1	0.5591
13	cifar	100	cos	20	0.001	0.1	0.5545
14	cifar	200	softplus	10	0.01	0.1	0.5753
15	cifar	200	softplus	10	0.001	0.1	0.6059
16	cifar	200	softplus	20	0.01	0.1	0.5317
17	cifar	200	softplus	20	0.001	0.1	0.5812
18	cifar	200	tanh	10	0.01	0.1	0.5575
19	cifar	200	tanh	10	0.001	0.1	0.572
20	cifar	200	tanh	20	0.01	0.1	0.5317
21	cifar	200	tanh	20	0.001	0.1	0.5441
22	cifar	200	cos	10	0.01	0.1	0.624
23	cifar	200	cos	10	0.001	0.1	0.5701
24	cifar	200	cos	20	0.01	0.1	0.5448
25	cifar	200	cos	20	0.001	0.1	0.5318
26	cifar	300	softplus	10	0.01	0.1	0.5595
27	cifar	300	softplus	10	0.001	0.1	0.6201
28	cifar	300	softplus	20	0.01	0.1	0.5369
29	cifar	300	softplus	20	0.001	0.1	0.5744
30	cifar	300	tanh	10	0.01	0.1	0.5708
31	cifar	300	tanh	10	0.001	0.1	0.5608
32	cifar	300	tanh	20	0.01	0.1	0.5337
33	cifar	300	tanh	20	0.001	0.1	0.5583
34	cifar	300	cos	10	0.01	0.1	0.6322
35	cifar	300	cos	10	0.001	0.1	0.5537
36	cifar	300	cos	20	0.01	0.1	0.5538
37	cifar	300	cos	20	0.001	0.1	0.5193