Μηχανική Μάθηση

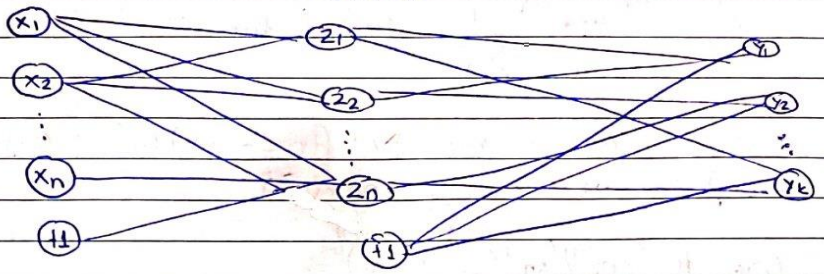
Εργασία 1η

Υπολογισμός μερικής παραγώγου για τον πίνακα παραμέτρων W(1):

Από τα δεδομένα προκύπτει το παρακάτω δίκτυο
1 κρυφού επιπέδου

x_n $z_1 = W^{(1)T} x_n$ 2 $z_2 = W^{(2)T} z_{nj}$ $y = \text{softmax}(z_2)$

$z_{nj} = h(z_1)$



Διαστάσεις
N×D N×M N×K

Χρησιμοποιώντας τον κανόνα της αλυσίδας για την μερική παράγωγο του πίνακα $W^{(2)}$ έχουμε:

$\frac{dE}{dW^{(2)}} = \frac{dE}{dy_{nk}} \cdot \frac{dy_{nk}}{dz_2} \cdot \frac{dz_2}{dW_2}$ όπου από (2) προκύπτει ότι η μερική παράγωγος σε διανυσματική μορφή είναι $(t - y)^T z \cdot \partial W^{(2)}$

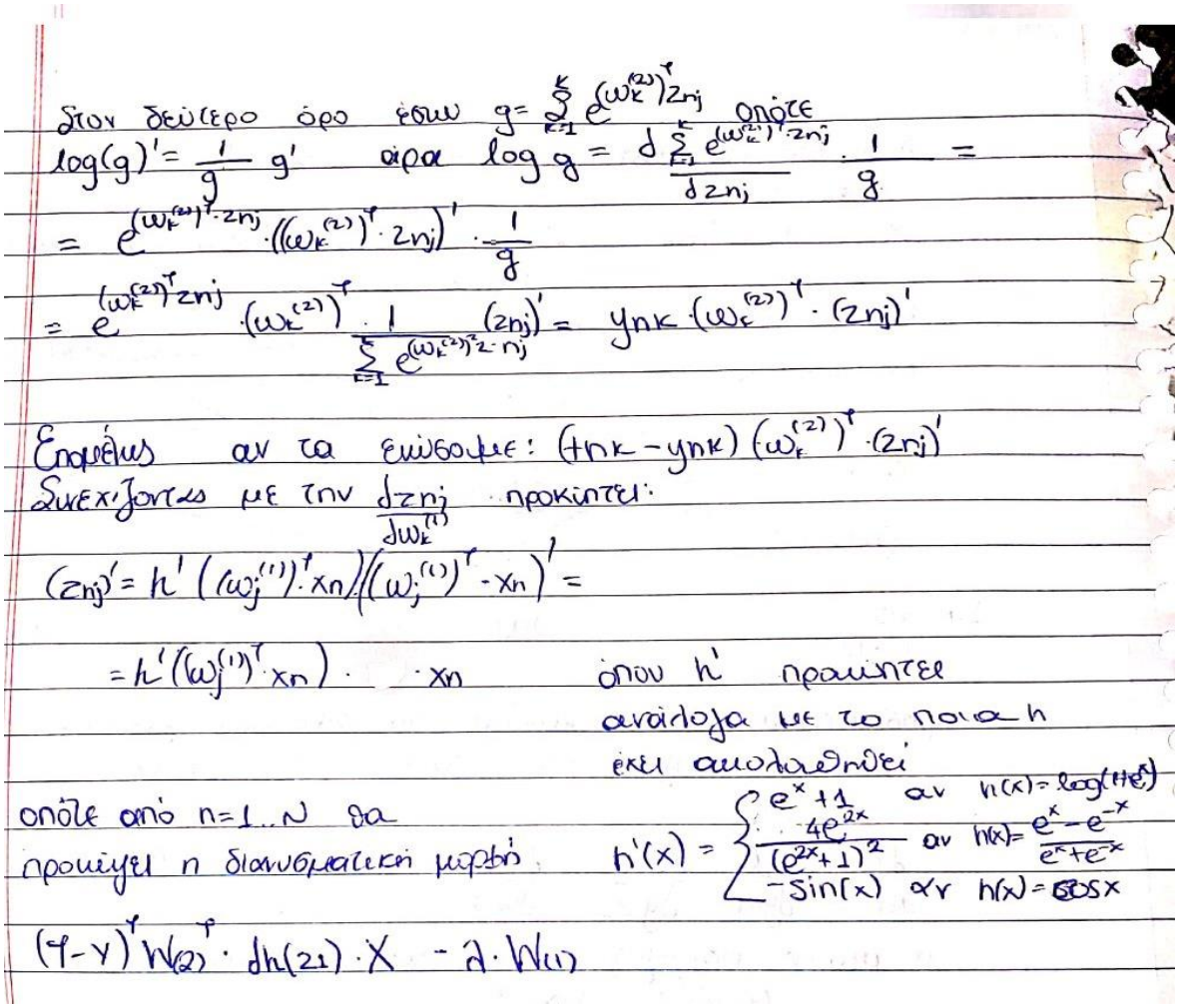
Για την μερική παραγωγο του πίνακα παραμέτρων $W^{(1)}$ προκύπτει από τον κανόνα της αλυσίδας:

$\frac{dE}{dW^{(1)}} = \frac{dE}{dy_{nk}} \cdot \frac{dy_{nk}}{dz_2} \frac{dz_2}{dz_{nj}} \frac{dz_{nj}}{dh} \frac{dh}{dz_1} \frac{dz_1}{dx_{in}}$

Επομένως απλοποιώ αρχικά το E_W σε:

$E_W = \sum_{n=1}^{N} \left[\left(\sum_{k=1}^{K} t_{nk} (w_k^{(2)})^T \cdot z_{nj} \right) - \log \left(\sum_{i=1}^{K} e^{(w_k^{(2)})^T \cdot z_{nj}} \right) \right] - \frac{\lambda}{2} \|w\|^2$

Επομένως, κρατώντας σταθερό το k, παραγωγίζω τα εσωτερικά αθροίσματα, όπου το πρώτο μέρος θα βγάλει $\frac{\partial t_{nk} (w_k^{(2)})^T z_{nj}}{\partial z_{nj}} = t_{nk} \cdot (w_k^{(2)})^T (z_{nj})'$



Στον δεύτερο όρο έστω $g = \sum_{k=1}^{K} e^{(w_k^{(2)})^T z_{nj}}$ οπότε

$\log(g)' = \frac{1}{g} g'$ άρα $\log g = \frac{d \sum_{k=1}^{K} e^{(w_k^{(2)})^T \cdot z_{nj}}}{d z_{nj}} \cdot \frac{1}{g} =$

$= e^{(w_k^{(2)})^T \cdot z_{nj}} \cdot ((w_k^{(2)})^T \cdot z_{nj})' \cdot \frac{1}{g}$

$= e^{(w_k^{(2)})^T z_{nj}} \cdot (w_k^{(2)})^T \cdot \frac{1}{\sum_{k=1}^{K} e^{(w_k^{(2)})^T z_{nj}}} (z_{nj})' = y_{nk} \cdot (w_k^{(2)})^T \cdot (z_{nj})'$

Επομένως αν τα ενώσουμε: $(t_{nk} - y_{nk})(w_k^{(2)})^T \cdot (z_{nj})'$

Συνεχίζοντας με την $\frac{d z_{nj}}{d w_k^{(1)}}$ προκύπτει:

$(z_{nj})' = h'((w_j^{(1)})^T \cdot x_n) \cdot ((w_j^{(1)})^T \cdot x_n)' =$

$= h'((w_j^{(1)})^T \cdot x_n) \cdot \cdot x_n$ όπου h' προκύπτει αναλογα με το ποια h έχει αποδοθεί

οπότε από $n=1..N$ θα προκύψει η διανυσματική μορφή

$h'(x) = \begin{cases} \frac{e^x + 1}{4e^{2x}} & \text{αν } h(x) = \log(1+e^x) \\ \frac{4e^{2x}}{(e^{2x}+1)^2} & \text{αν } h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \\ -\sin(x) & \text{αν } h(x) = \cos x \end{cases}$

$(Y - Y)^T W_{(2)}^T \cdot dh(Z_1) \cdot X - \lambda \cdot W_{(1)}$

Όπου Y είναι $N \times K$ $[Y]_{nk} = t_{nk}$, Y είναι $N \times K$ $[Y]_{nk} = y_{nk}$, $W_{(2)}$ είναι $K \times (M+1)$ $[W_{(2)}]_j = W_{(2)j}$, Z_1 είναι $N \times (D+1)$ $[Z_1] = (w_j^{(1)})^T \cdot x_n$, X είναι $N \times (D+1)$ $[X]_{n*} = x_n$

Έλεγχος ορθότητας παραγώγου(gradient check):

(Όπου branch υπονοείται ποιου είδους h έχει χρησιμοποιηθεί για 1-> έχει χρησιμοποιηθεί η (4),για 2->(5), για 3->(6))

Για **branch=1**:

```
The difference estimate for gradient of w1 is :  3.
7314403678045 95e-08
The difference estimate for gradient of w2 is :  1.
9524179606378 31e-08
```

Ο πίνακας **gradEw1:**

```
          0         1         2    ...       782
783       784
0 -1.286846 -0.082356 -0.061227  ... -0.056297 -0.
041131 -0.093540
1  1.184217 -0.012502 -0.007933  ... -0.014180 -0.
003952 -0.019458
2 -0.050533 -0.051718 -0.046464  ... -0.020543 -0.
084964 -0.016725
3  0.550271 -0.086635 -0.010744  ... -0.039295 -0.
068050 -0.034876
4  1.235503 -0.091354 -0.099080  ... -0.091643 -0.
066076 -0.078260
```

Ο πίνακας με τις υπολογισμένες μερικές παραγώγους του πίνακα W1(**numericalGrad1**):

```
          0         1         2    ...       782
783       784
0 -1.286846 -0.082356 -0.061227  ... -0.056297 -0.
041131 -0.093540
1  1.184217 -0.012502 -0.007933  ... -0.014180 -0.
003952 -0.019458
2 -0.050533 -0.051718 -0.046464  ... -0.020543 -0.
084964 -0.016725
3  0.550271 -0.086635 -0.010744  ... -0.039295 -0.
068050 -0.034876
4  1.235503 -0.091354 -0.099080  ... -0.091643 -0.
066076 -0.078260
```

Ο πίνακας **gradEw2:**

```
           0           1           2    ...          98
           99         100
0   50.680377   52.329452   47.082256   ...   48.
044936   46.476188  0.942522
1 -280.831464 -302.072888 -272.740471   ... -263.
423454 -271.633015 -5.051741
2   70.190386   77.743825   67.807868   ...   65.
080557   68.097963  0.915327
3   -0.097236   -0.087720   -0.034434   ...   -0.
068332   -0.004216 -0.040351
4   45.987220   50.583882   46.558675   ...   44.
524924   50.390404  0.996283
```

Ο πίνακας με τις υπολογισμένες μερικές παραγώγους του πίνακα W2(**numericalGrad2**):

```
           0           1           2    ...          98
           99         100
0   50.680377   52.329452   47.082256   ...   48.
044936   46.476188  0.942522
1 -280.831464 -302.072888 -272.740471   ... -263.
423454 -271.633015 -5.051741
2   70.190386   77.743825   67.807868   ...   65.
080557   68.097963  0.915327
3   -0.097236   -0.087720   -0.034434   ...   -0.
068332   -0.004216 -0.040351
4   45.987220   50.583882   46.558675   ...   44.
524924   50.390404  0.996283
```

Για **branch=2:**

```
The difference estimate for gradient of w1 is :  1.3610776186667728e-09
The difference estimate for gradient of w2 is :  1.3154008993154775e-09
```

Ο πίνακας **gradEw1:**

```
        0         1         2    ...       782       783       784
0 -0.018850 -0.044393 -0.008566   ... -0.005272 -0.049147 -0.044541
1 -0.041159 -0.086544 -0.024125   ... -0.002145 -0.020006 -0.049904
2 -0.024913 -0.056198 -0.071077   ... -0.026268 -0.083974 -0.067717
3 -0.083232 -0.063313 -0.061106   ... -0.083468 -0.045126 -0.045203
4 -0.056533 -0.048657 -0.014548   ... -0.030709 -0.098071 -0.007294
```

Ο πίνακας με τις υπολογισμένες μερικές παραγώγους του πίνακα W1(**numericalGrad1**):

	0	1	2	...	782	783	784
0	-0.018850	-0.044393	-0.008566	...	-0.005272	-0.049147	-0.044541
1	-0.041159	-0.086544	-0.024125	...	-0.002145	-0.020006	-0.049904
2	-0.024913	-0.056198	-0.071077	...	-0.026268	-0.083974	-0.067717
3	-0.083232	-0.063313	-0.061106	...	-0.083468	-0.045126	-0.045203
4	-0.056533	-0.048657	-0.014548	...	-0.030709	-0.098071	-0.007294

Ο πίνακας **gradEw2**:

	0	1	2	...	98	99	100
0	1.903387	1.880588	1.932792	...	1.944212	1.882090	1.880964
1	-0.231444	-0.229990	-0.193244	...	-0.256097	-0.235991	-0.257351
2	0.853732	0.800693	0.863084	...	0.872996	0.810201	0.785864
3	-0.768779	-0.775986	-0.794370	...	-0.845370	-0.770400	-0.814978
4	0.884098	0.882079	0.832438	...	0.844500	0.879104	0.808268

Ο πίνακας με τις υπολογισμένες μερικές παραγώγους του πίνακα W2(**numericalGrad2**):

	0	1	2	...	98	99	100
0	1.903387	1.880588	1.932792	...	1.944212	1.882090	1.880964
1	-0.231444	-0.229990	-0.193244	...	-0.256097	-0.235991	-0.257351
2	0.853732	0.800693	0.863084	...	0.872996	0.810201	0.785864
3	-0.768779	-0.775986	-0.794370	...	-0.845370	-0.770400	-0.814978
4	0.884098	0.882079	0.832438	...	0.844500	0.879104	0.808268

Για **branch=3**:

```
The difference estimate for gradient of w1 is :  3.
3297311574642663e-09
The difference estimate for gradient of w2 is :  1.
3267426046681408e-09
```

Ο πίνακας **gradEw1:**

```
        0         1         2    ...       782       783       784
0  0.397512 -0.085757 -0.007468  ... -0.033824 -0.046950 -0.001202
1  0.116594 -0.057599 -0.025138  ... -0.006893 -0.018204 -0.041813
2 -0.333871 -0.091438 -0.096014  ... -0.065823 -0.058399 -0.057247
3 -0.205122 -0.013641 -0.020694  ... -0.006126 -0.015016 -0.072315
4 -0.215117 -0.007863 -0.090348  ... -0.013509 -0.003418 -0.023780
```

Ο πίνακας με τις υπολογισμένες μερικές παραγώγους του πίνακα W1(**numericalGrad1**):

```
        0         1         2    ...       782       783       784
0  0.397512 -0.085757 -0.007468  ... -0.033824 -0.046950 -0.001202
1  0.116594 -0.057599 -0.025138  ... -0.006893 -0.018204 -0.041813
2 -0.333871 -0.091438 -0.096014  ... -0.065823 -0.058399 -0.057247
3 -0.205122 -0.013641 -0.020694  ... -0.006126 -0.015016 -0.072315
4 -0.215117 -0.007863 -0.090348  ... -0.013509 -0.003418 -0.023780
```

Ο πίνακας **gradEw2:**

```
        0         1         2    ...        98
99        100
0  0.413054 -0.381527 -0.756172  ...  0.448423 -0.
767041  0.475352
1 -0.070832 -0.116739 -0.158193  ... -0.044327  0.
059844 -0.250710
2  0.040044  0.615558  1.145443  ... -0.058972 -0.
603724 -0.184649
3 -0.144293 -0.016528  0.180443  ... -0.270002  0.
153597 -0.422026
4 -0.553497 -0.250625 -1.269200  ...  0.573506  1.
232543 -0.607512
```

Ο πίνακας με τις υπολογισμένες μερικές παραγώγους του πίνακα W2(**numerical Grad2**):

```
        0         1         2    ...        98
99        100
0  0.413054 -0.381527 -0.756172  ...  0.448423 -0.
767041  0.475352
1 -0.070832 -0.116739 -0.158193  ... -0.044327  0.
059844 -0.250710
2  0.040044  0.615558  1.145443  ... -0.058972 -0.
603724 -0.184649
3 -0.144293 -0.016528  0.180443  ... -0.270002  0.
153597 -0.422026
4 -0.553497 -0.250625 -1.269200  ...  0.573506  1.
232543 -0.607512
```

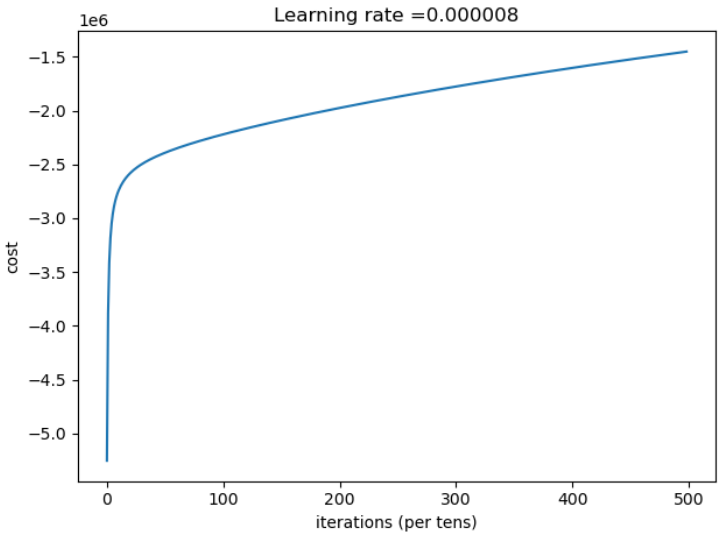
Αποτελέσματα πινάκων:

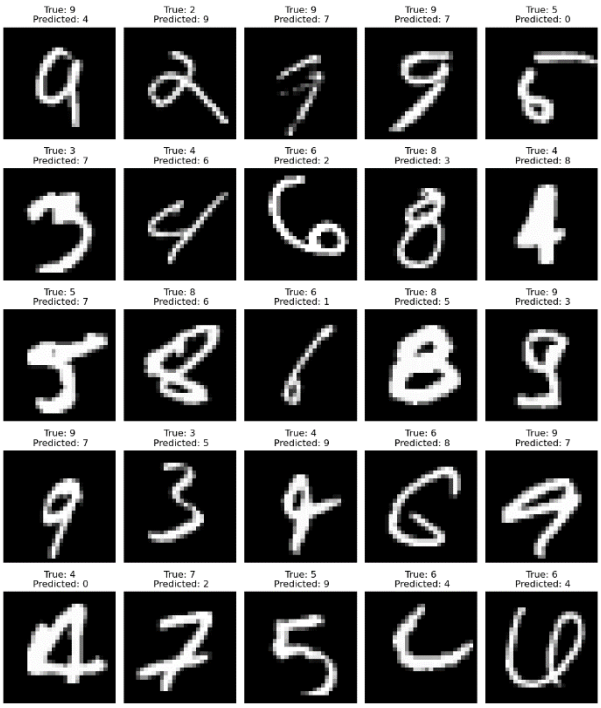
Για **MNIST Data**:

Ενδεικτικά αποτελέσματα για **M=100**

- Branch=1

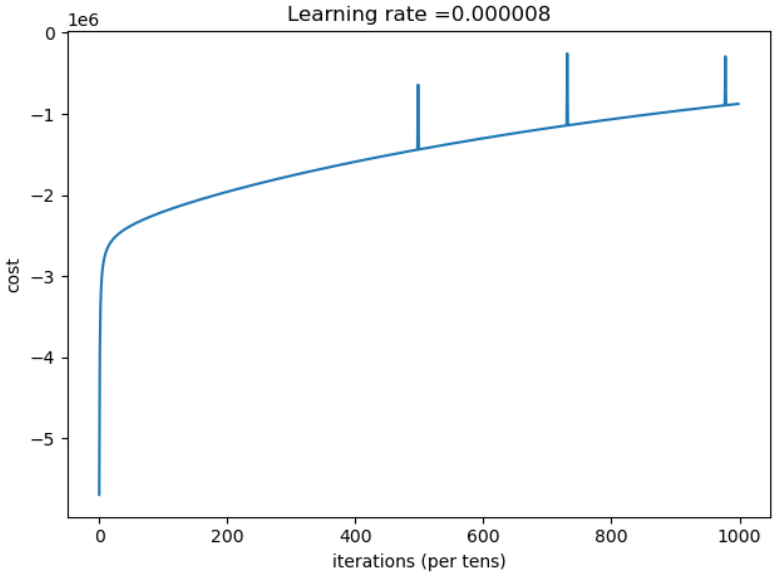
- Iteration=500,batch_size=100, lr= 8.333333333333334e-06

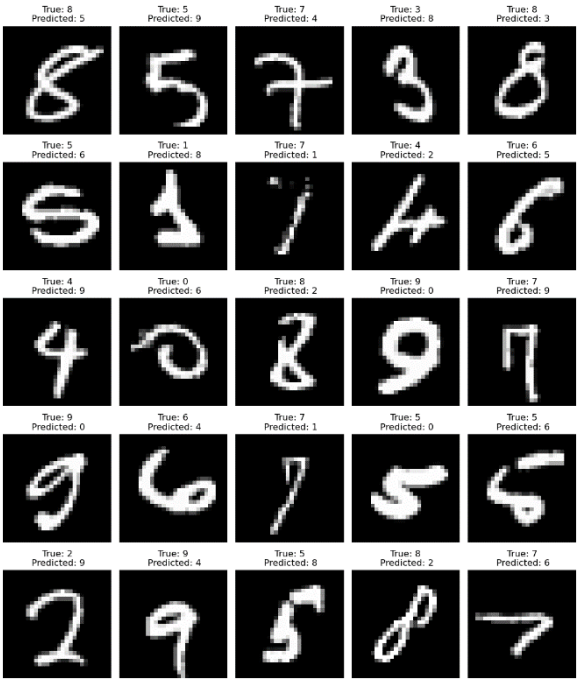




Accuracy: 0.9066

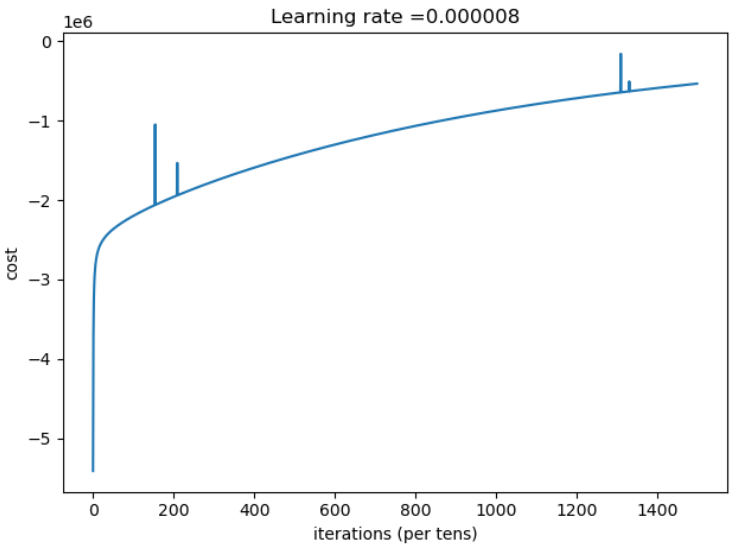
- Iteration=1000,batch_size=100, lr= 8.333333333333334e-06

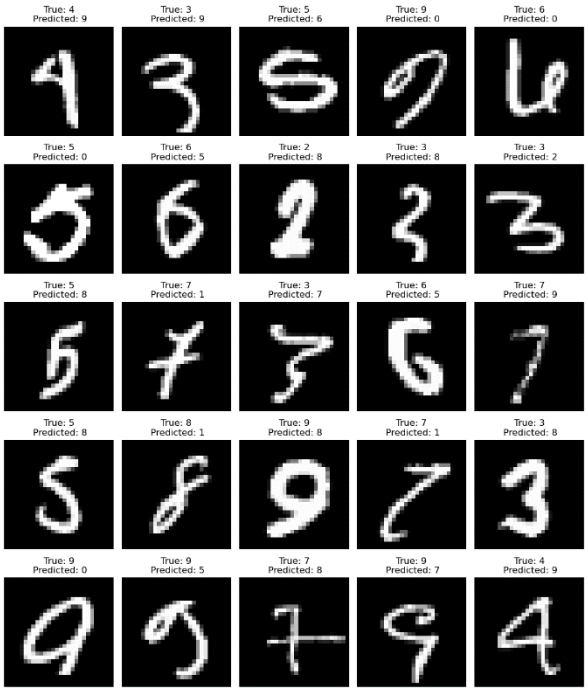




Accuracy: 0.9314

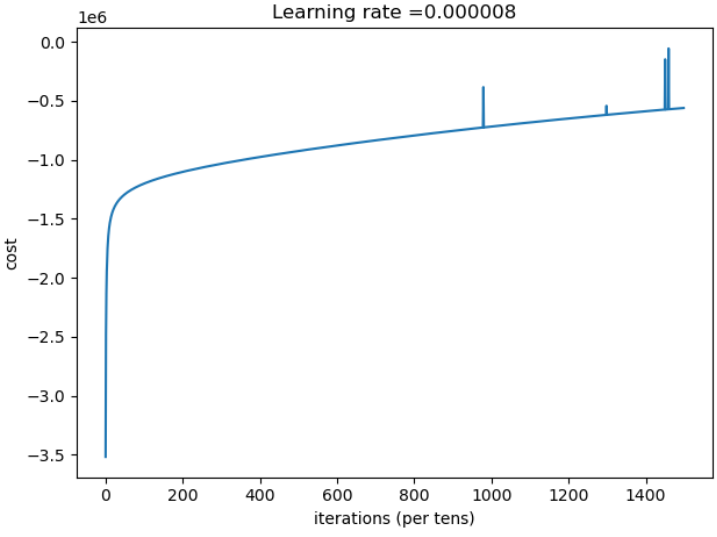
- Iteration=1500,batch_size=100, lr= 8.333333333333334e-06

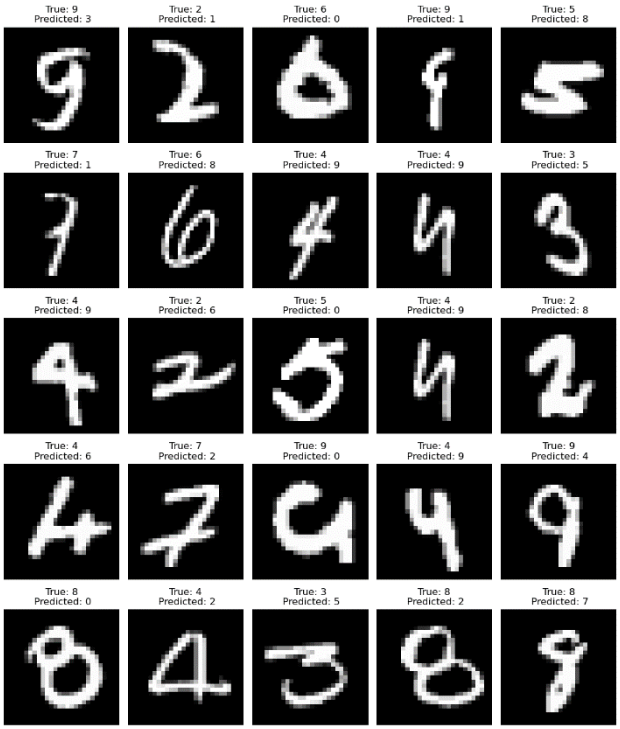




Accuracy: 0.9441

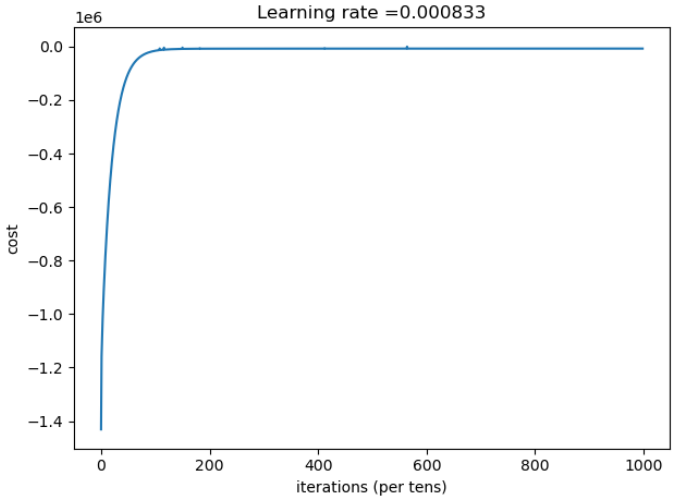
- Iteration=1500,batch_size=200, lr= 8.333333333333334e-06

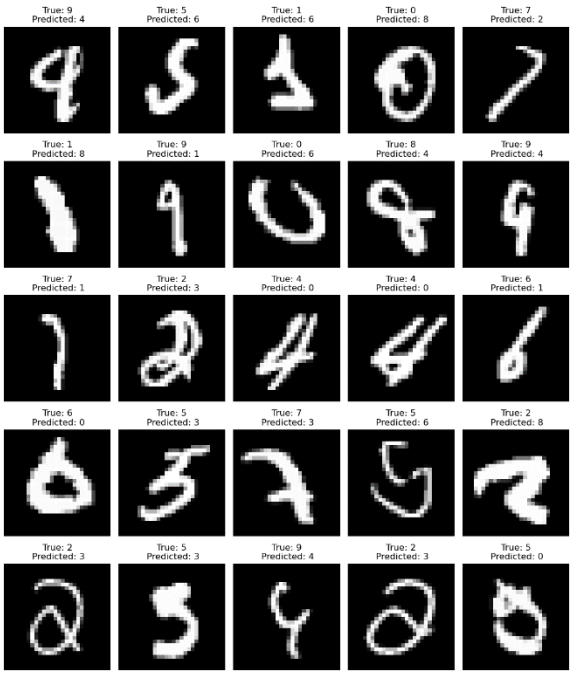




Accuracy: 0.9388

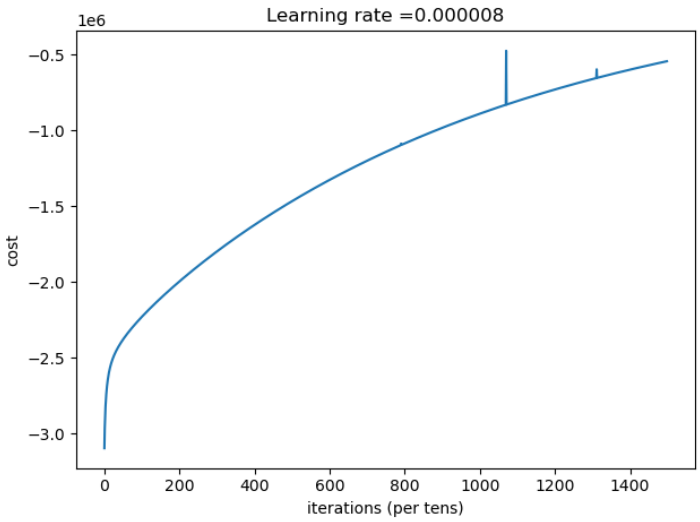
- Iteration=1000,batch_size=200, lr= 8.3333333333334e-06*100

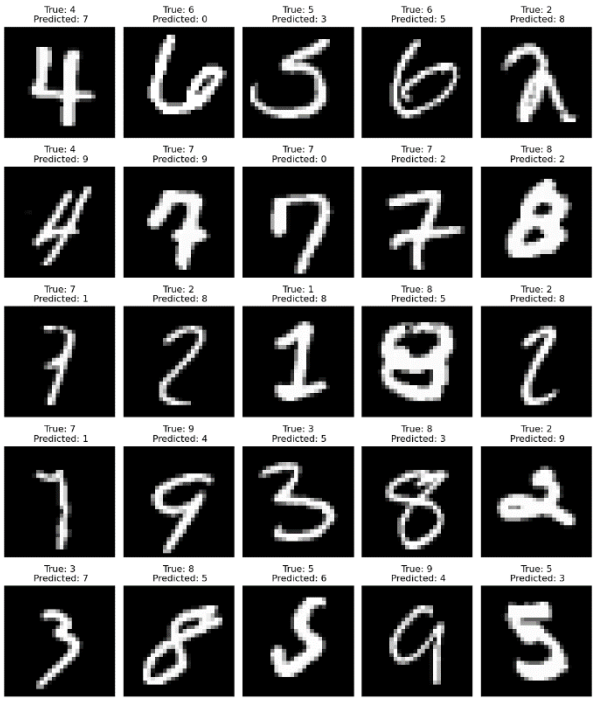




Accuracy: 0.9784

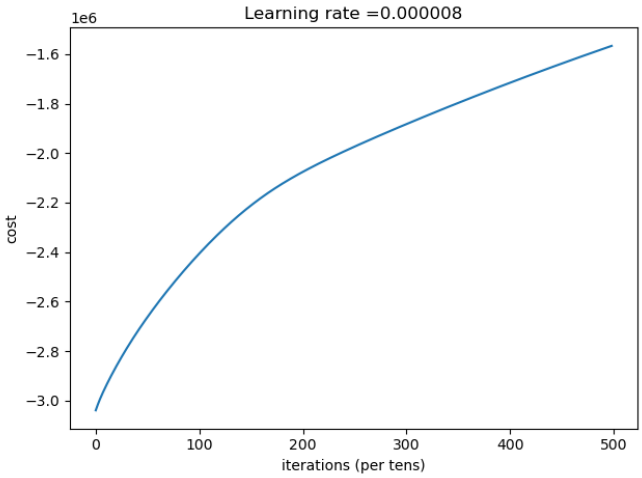
- Branch=2
  - Iteration=1500,batch_size=100, lr= 8.333333333333334e-06

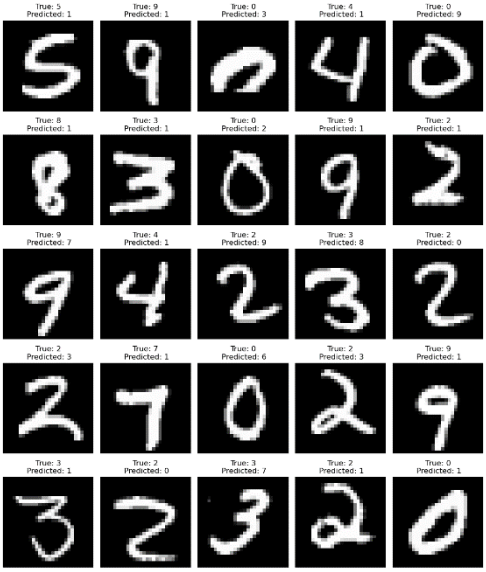




Accuracy: 0.9246

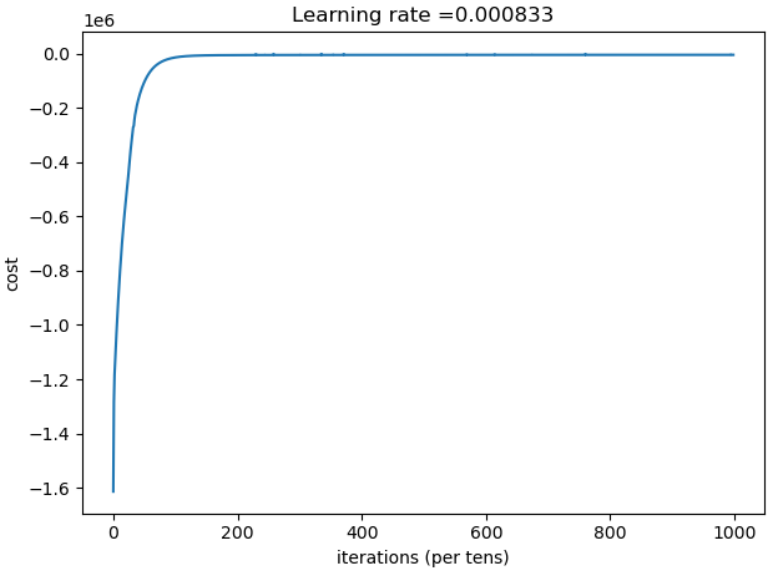
- Branch=3
  - Iteration=500,batch_size=100,lr=8. 33333333333334e-06

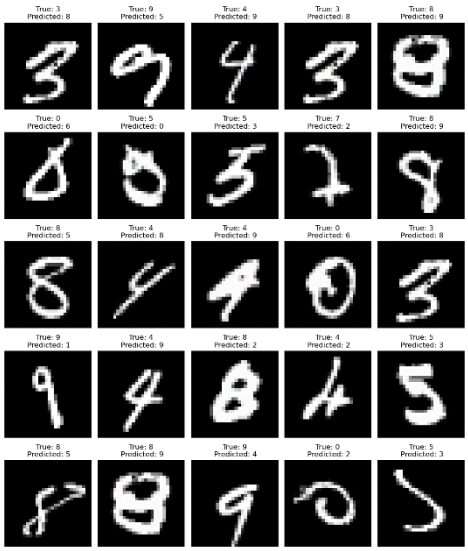




Accuracy: 0.1577

  - Iteration=1000,batch_size=200,lr=8.33333333333334e-06*100



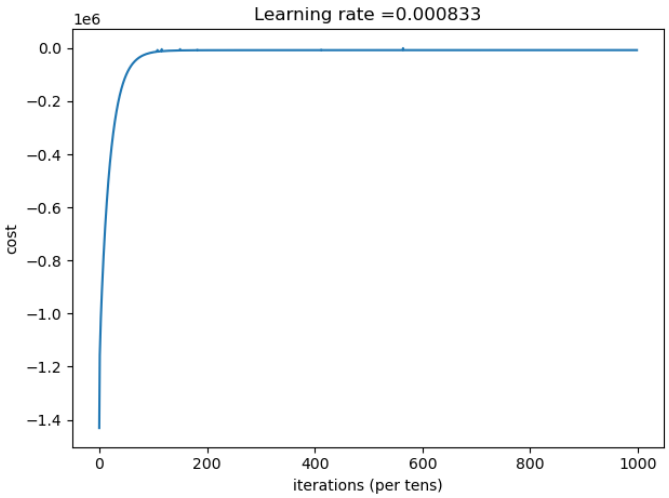


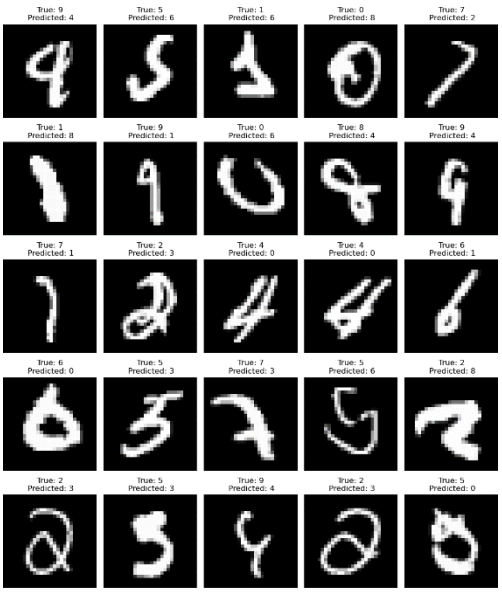
Accuracy: 0.9815

Ενδεικτικά αποτελέσματα για **M=200**

- Branch=1

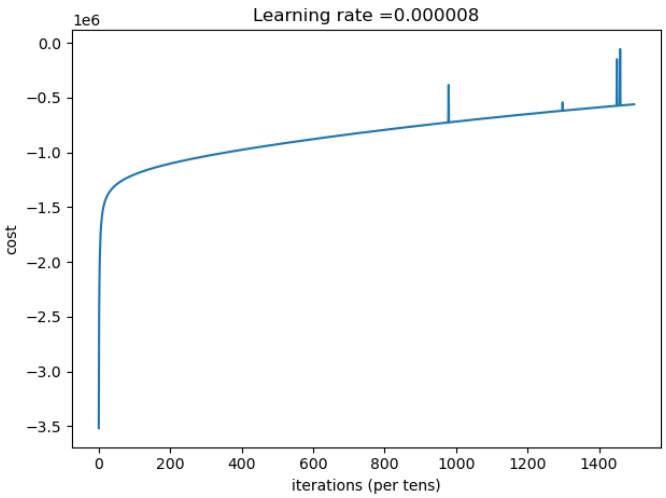
- Iteration=1000,batch_size=200, lr= 8.33333333333334e-06*100

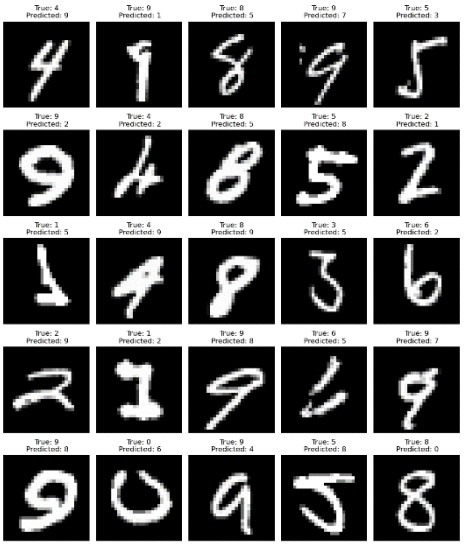




Accuracy: 0.9784

- Iteration=1500,batch_size=100, lr= 8.33333333333334e-06

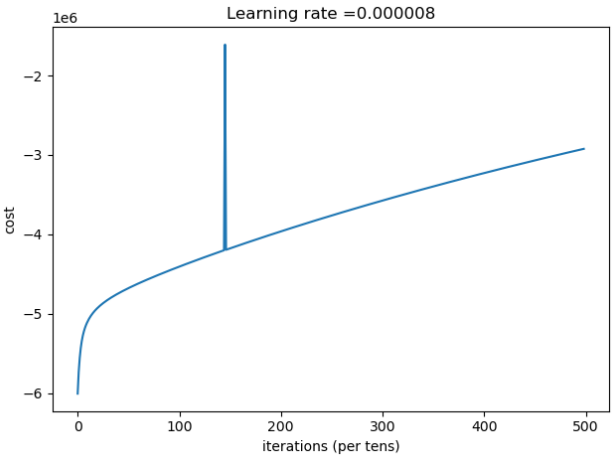


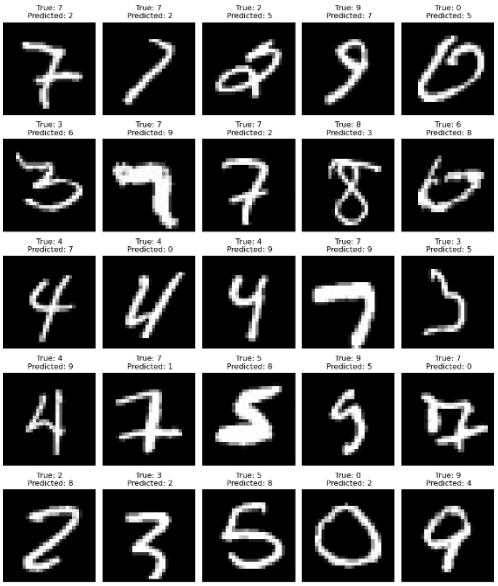


Accuracy:0.9467

- Branch=2

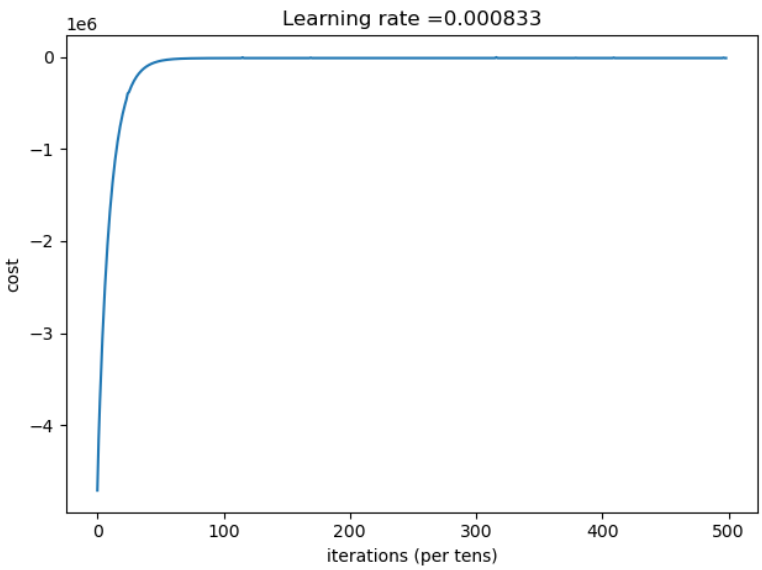
- Iteration=500,batch_size=100, lr= 8.333333333333334e-06

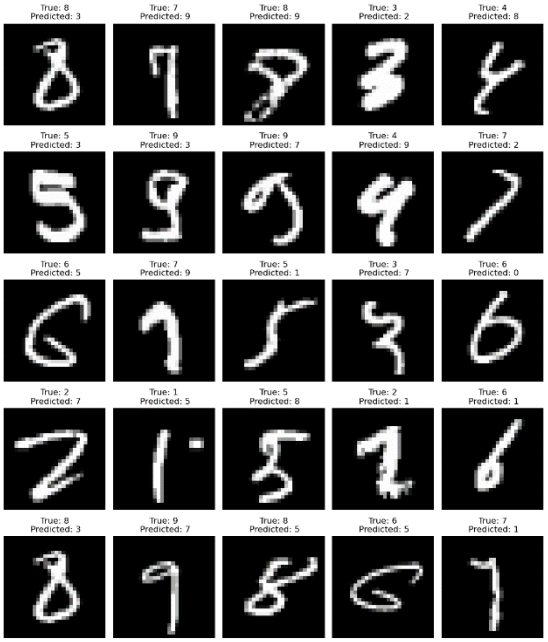




Accuracy: 0.8845

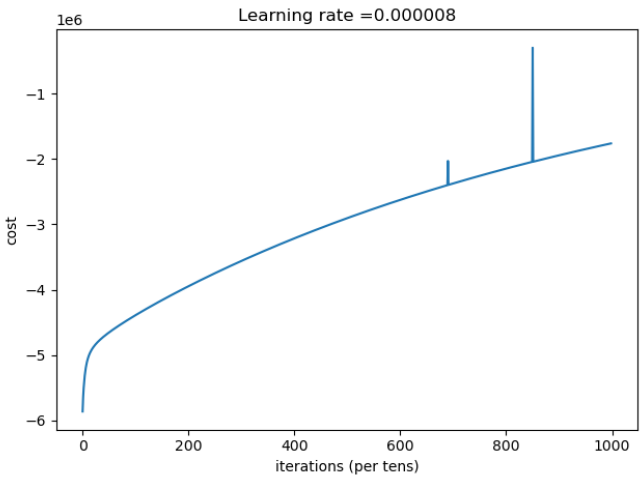
- Iteration=500,batch_size=100, lr= 8.33333333334e-06*100

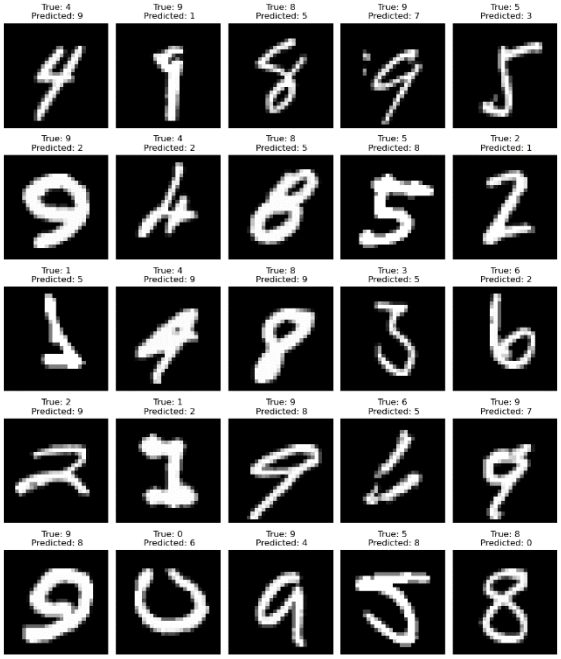




Accuracy:0.9784

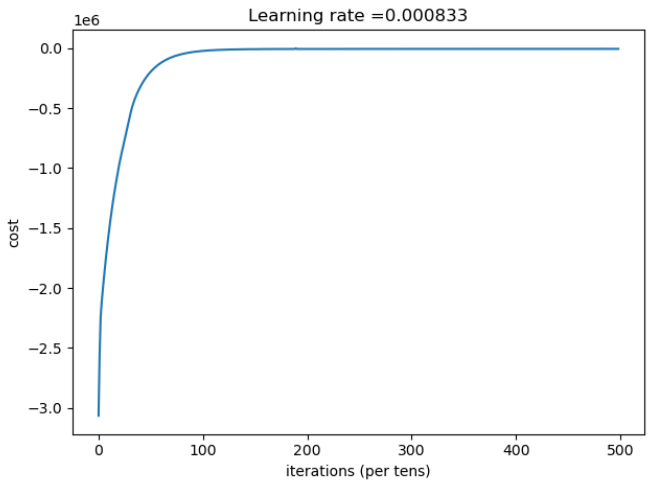
- Iteration=1000,batch_size=100, lr= 8.333333333333334e-06

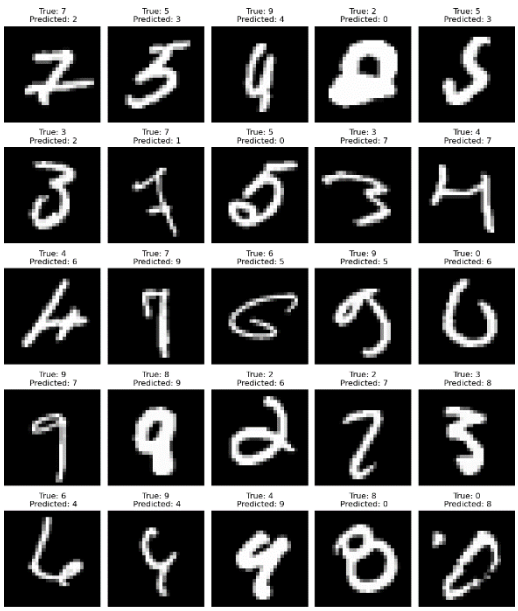




Accuracy: 0.9062

- Branch=3
  - Iteration=500,batch_size=200,lr=8.33333333333334 e-06*100



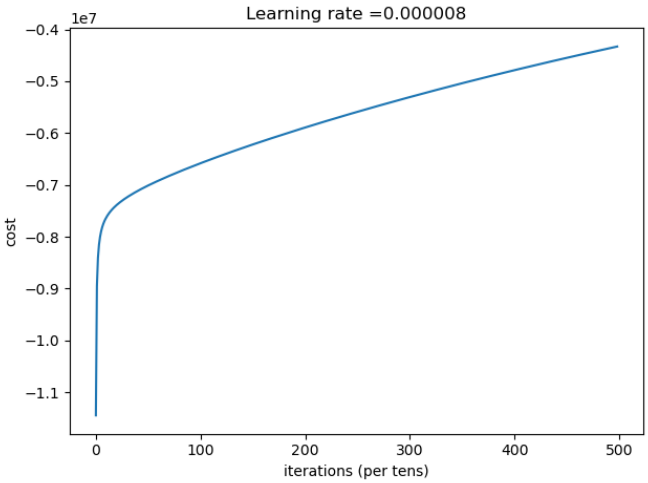


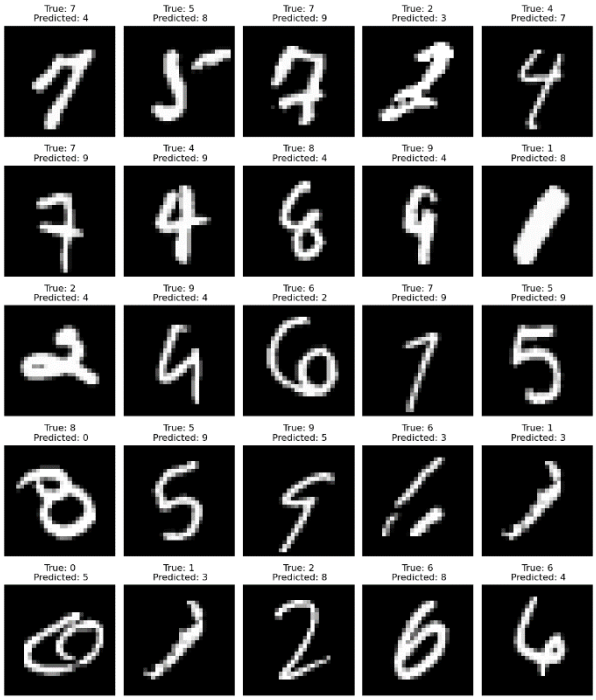
Accuracy: 0.9825

Ενδεικτικά αποτελέσματα για **M=300**

- Branch=1

- Iteration=500,batch_size=100,lr=8.3333333 3333334e-06

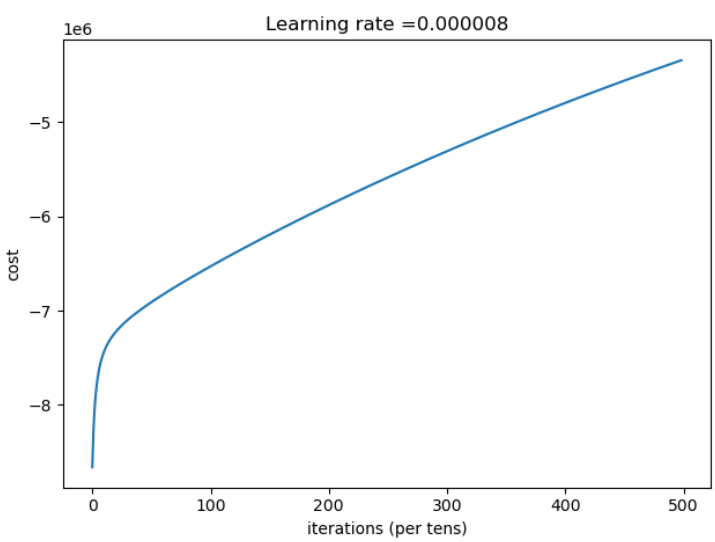


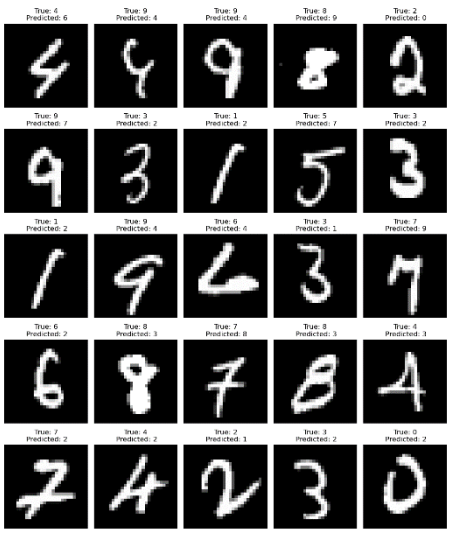


Accuracy: 0.9323

❖ Branch=2

- Iteration=500,batch_size=100,lr=8.33333333333334e-06





Accuracy: 0.8918

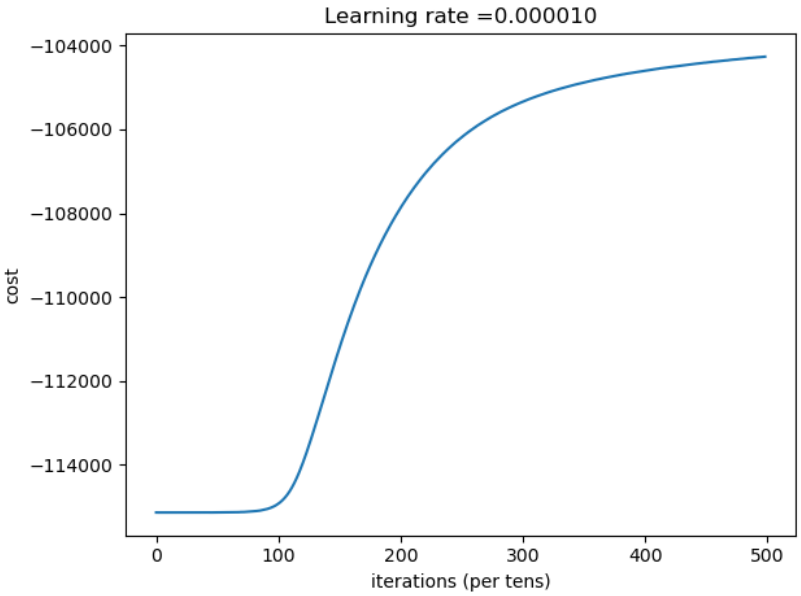
❖ Branch=3

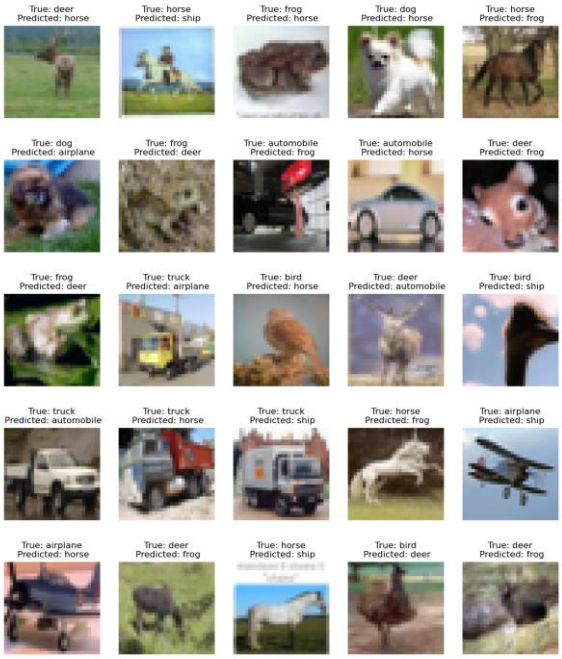
- Iteration=500,batch_size=100,lr=8.33333333333334e-06*10

Για **CIFAR-10 Data**:

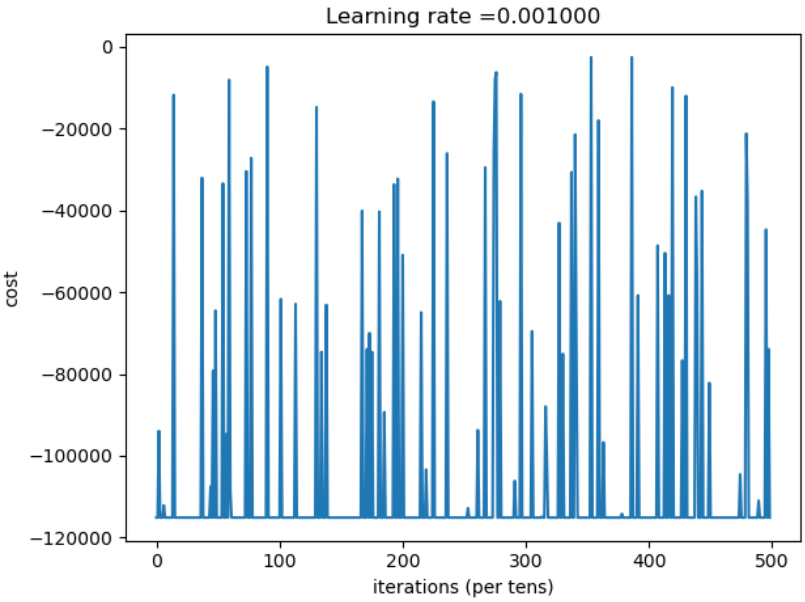
Ενδεικτικά αποτελέσματα για **M=100**

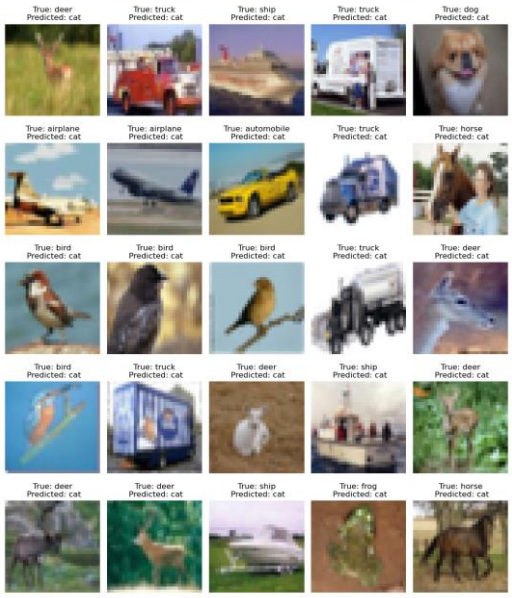
- Branch=1





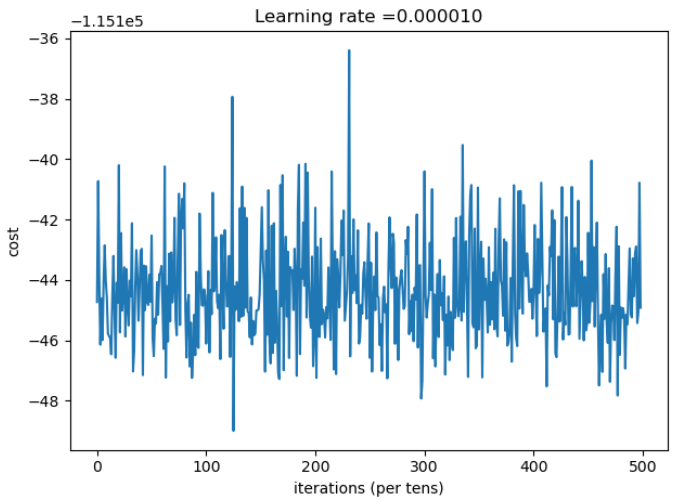
  - Iteration= 500 ,lr= 0.001 ,batch_size=100
    Accuracy: 0.1862
- Branch=2
  - Iteration= 500 ,lr= 0.001 ,batch_size=100

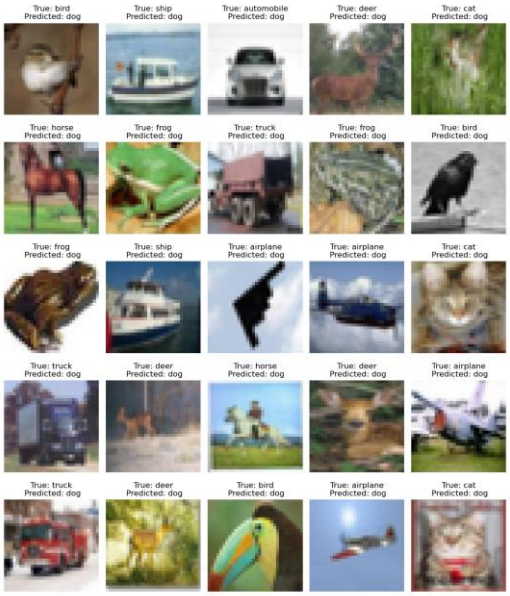




Accuracy: 0.1

- Branch=3
  - Iteration= 500 ,lr= 0.001 ,batch_size=100

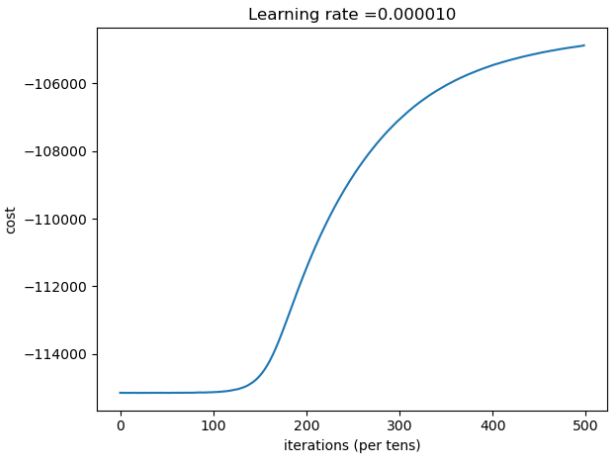


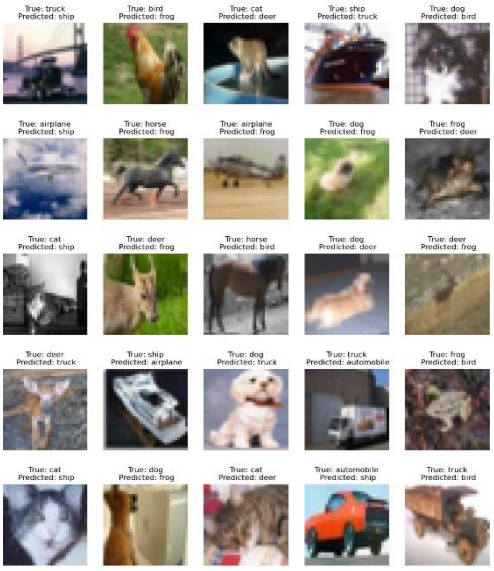


Accuracy: 0.1

Ενδεικτικά αποτελέσματα για **M=200**

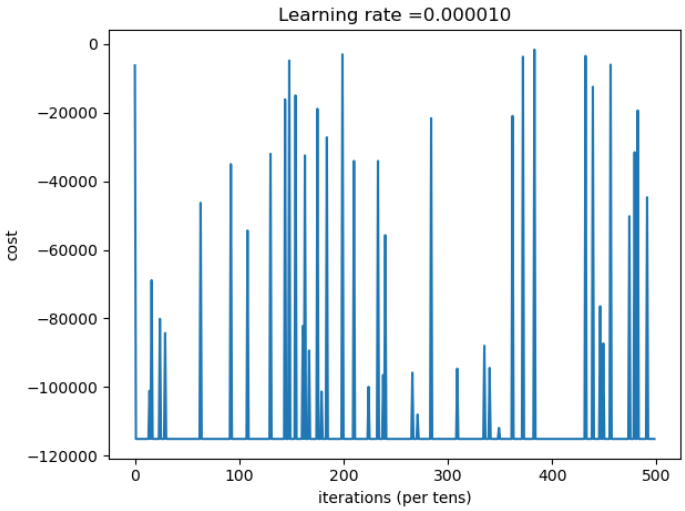
- <u>Branch=1</u>
  - Iteration= 500 ,lr= 1e-06,batch_size=100

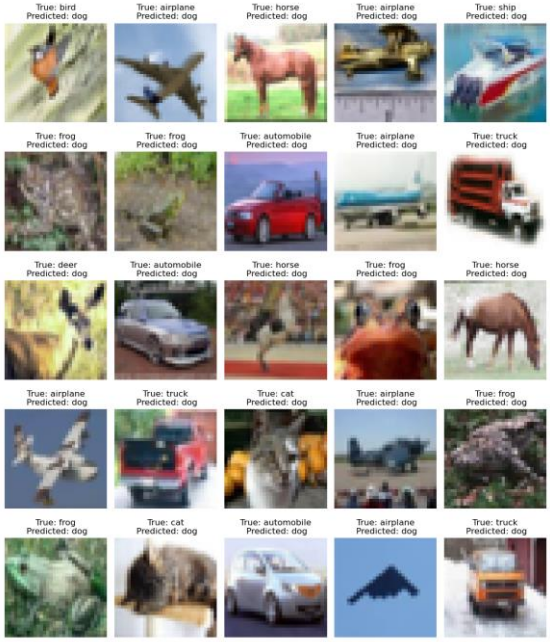




Accuracy: 0.1816

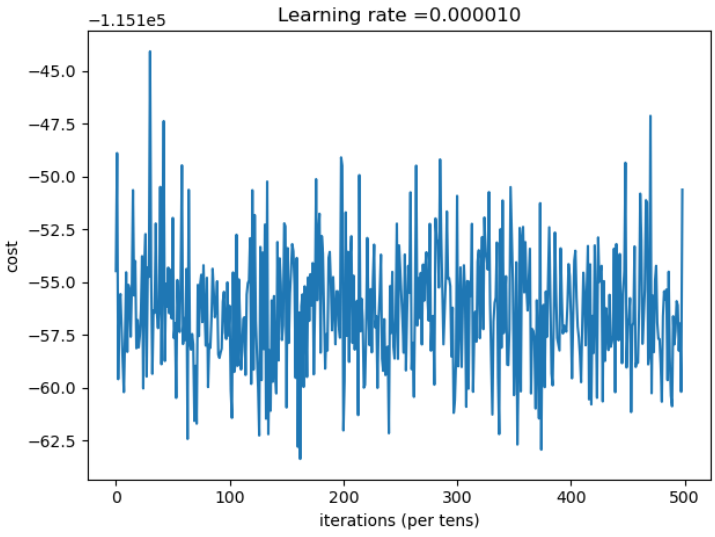
- Branch=2
  - Iteration= 500 ,lr= 1e-06,bat ch_size=100

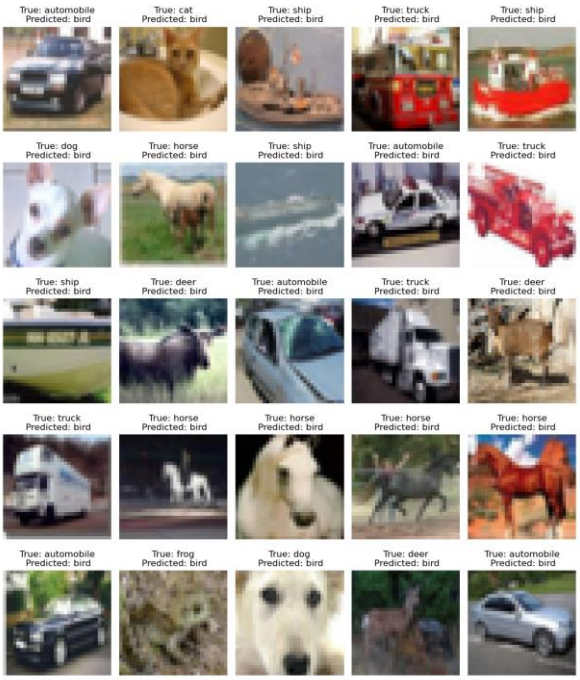




Accuracy=0.1

- Branch=3
  - Iteration= 500 ,lr= 1e-06,bat ch_size=100



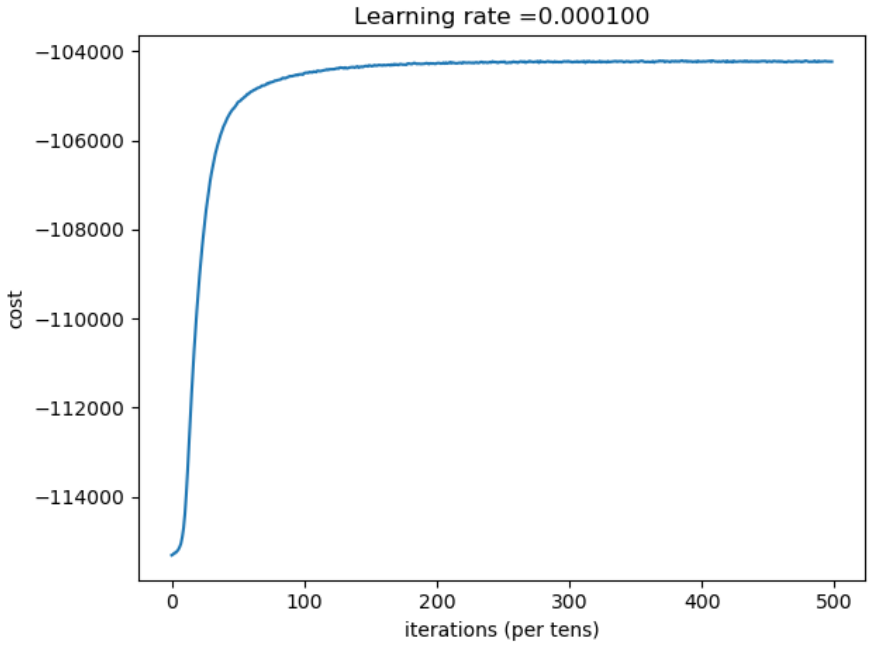


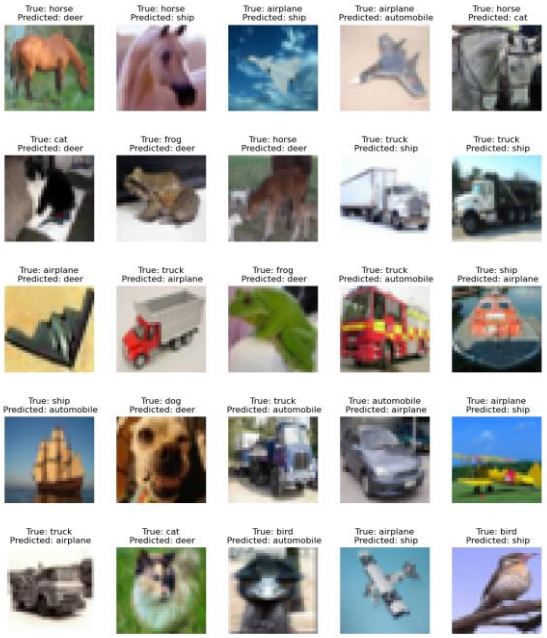
Accuracy=0.1

Ενδεικτικά αποτελέσματα για **M=300**

- Branch=1

- Iteration= 500 ,lr= 0.0001 ,batch_size=100

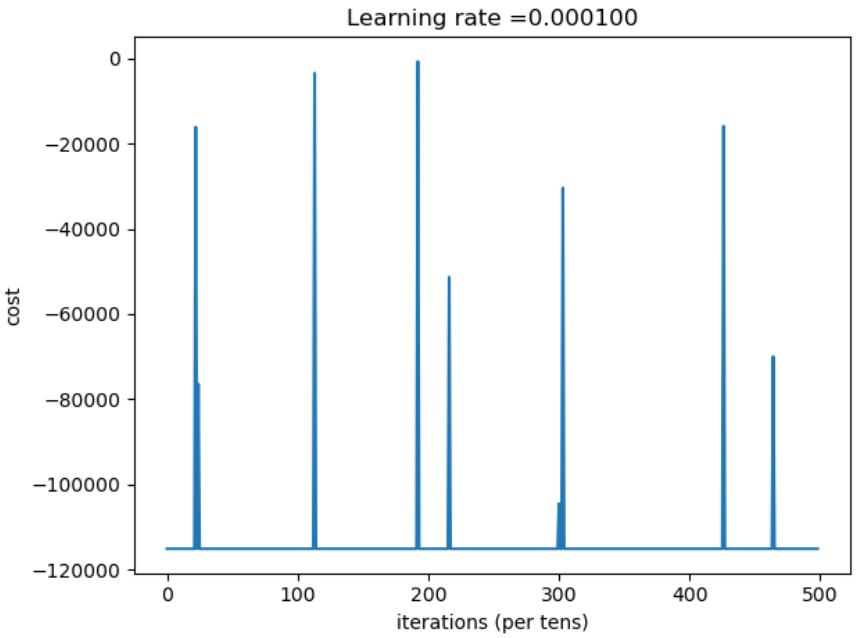


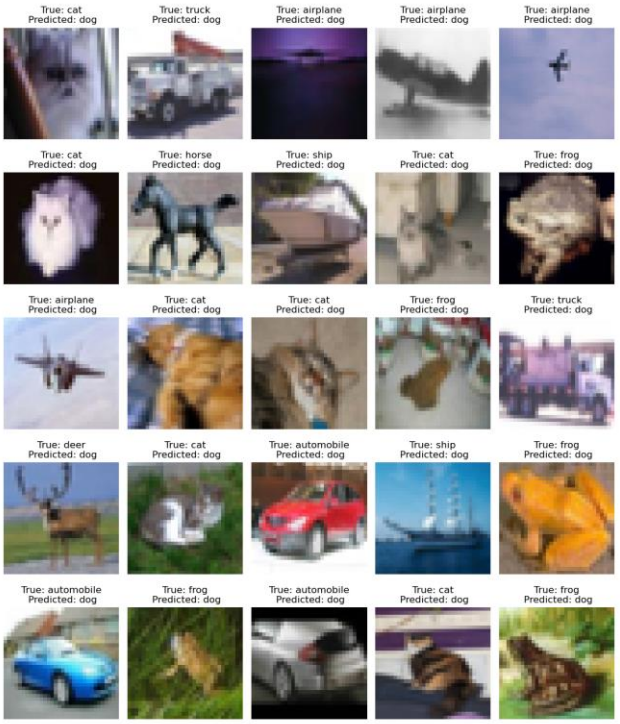


Accuracy: 0.192

❖ Branch=2

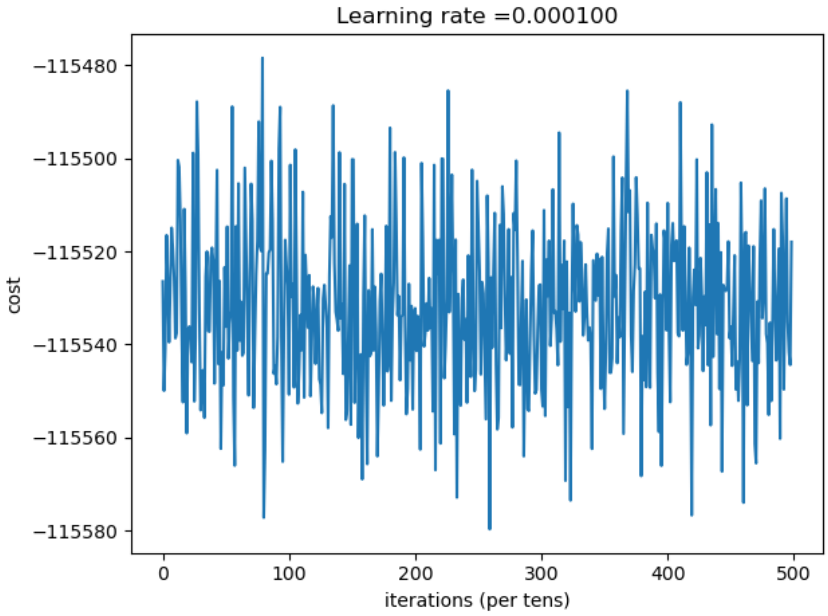
- Iteration= 500 ,lr= 0.0001 ,batch_size=100

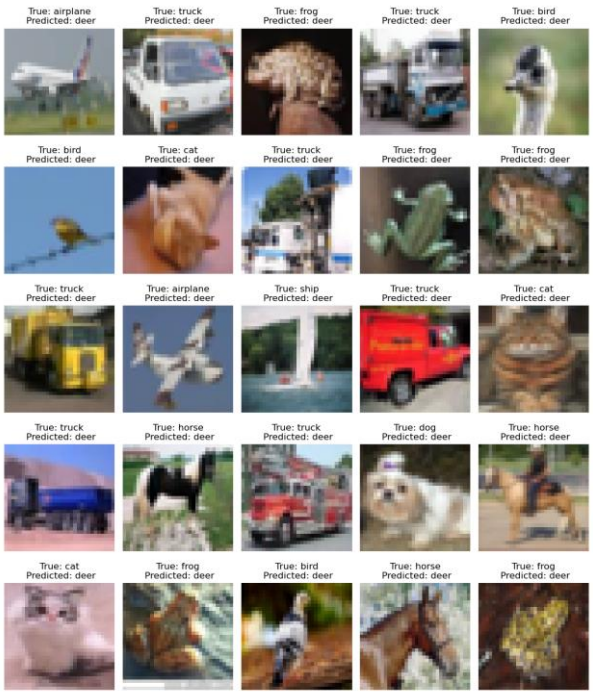




Accuracy=0.1

- Branch=3
  - Iteration= 500 ,lr= 0.0001 ,batch_size=100





Accuracy: 0.1