

Μηχανική Μάθηση

Project A

Τσόγκας Ευάγγελος 3150185

Περιεχόμενα

1. Μερικές Παράγωγοι Πίνακα $W^{(1)}$	3
2. Παραδείγματα Εφαρμογής για το MNIST	4
2.1 Αποτελέσματα για $M=100$ hidden units	4
2.2 Αποτελέσματα για $M=200$ hidden units	6
2.3 Αποτελέσματα για $M=300$ hidden units	7
3. Παραδείγματα Εφαρμογής για το CIFAR-10.....	9
3.1 Αποτελέσματα για $M=100$ hidden units	9
3.2 Αποτελέσματα για $M=200$ hidden units	11
3.3 Αποτελέσματα για $M=300$ hidden units	12
4. Καλύτερα Αποτελέσματα	14

1. Μερικές Παράγωγοι Πίνακα $W^{(1)}$

Συνάρτηση κόστους:

$$E(w) = \sum_{n=1}^N \sum_{k=1}^K t_{nk} \log y_{nk} - \frac{\lambda}{2} ||w||^2$$

Έχοντας τα εξής δεδομένα:

- $\alpha = (w_j^{(1)})^T x_n$
- $h(\alpha)$ η συνάρτηση ενεργοποίησης του hidden layer
- $(T - Y)^T Z - \lambda W^{(2)}$ οι μερικές παράγωγοι για τις παραμέτρους $W^{(2)}$

Οι μερικές παράγωγοι για τις παραμέτρους $W^{(1)}$ δίνονται από την σχέση:

$$h'(\alpha)(T - Y)W^{(2)}X - \lambda W^{(1)}$$

όπου $h'(\alpha)$ είναι η παράγωγος της συνάρτησης ενεργοποίησης $h(\alpha)$, ο T περιέχει τα δεδομένα εξόδου για τα δεδομένα εκπαίδευσης, ο Y τις τιμές των softmax πιθανοτήτων και ο X τα δεδομένα εκπαίδευσης.

Παράδειγμα αποτελέσματος της συνάρτησης gradcheck για το MNIST:

```
The difference estimate for gradient of w1 is : 3.0309401453931795e-07
The difference estimate for gradient of w2 is : 2.686775384397322e-07
```

2. Παραδείγματα Εφαρμογής για το MNIST

Σε όλα τα παραδείγματα χρησιμοποιούνται οι εξής παράμετροι:

- Μέγεθος minibatch = 200
- Αριθμός επαναλήψεων = 300
- Learning rate $\eta = 0.0025$ (0.5/μέγεθος minibatch)
- $\lambda = 0.03$

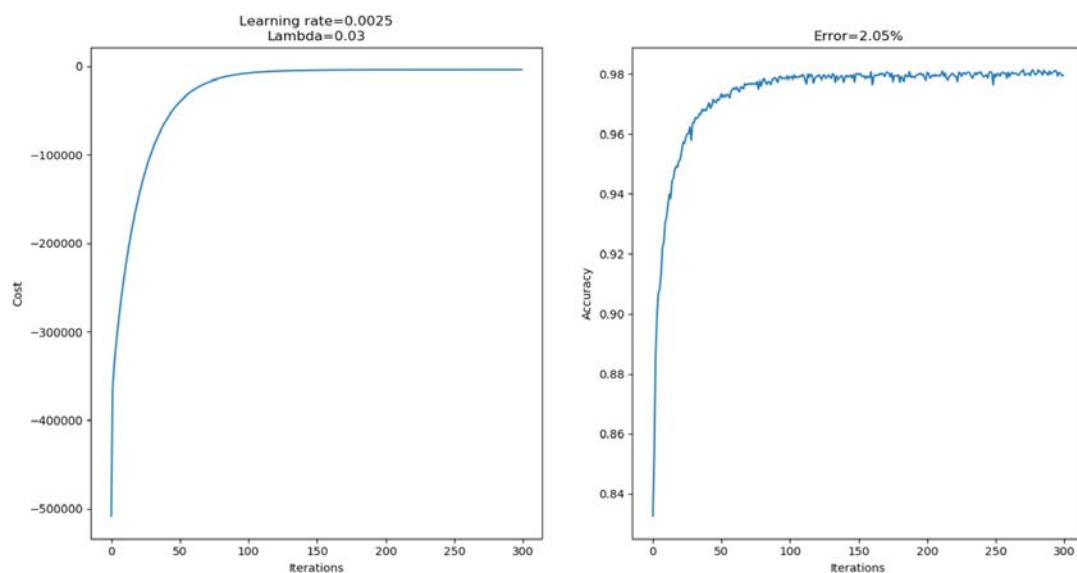
Τα κόστη στα διαγράμματα των παραδειγμάτων αποτελούν το άθροισμα των κοστών όλων των minibatches για κάθε επανάληψη.

2.1 Αποτελέσματα για M=100 hidden units

α) Συνάρτηση Ενεργοποίησης $h(a) = \log(1 + e^a)$

Επανάληψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

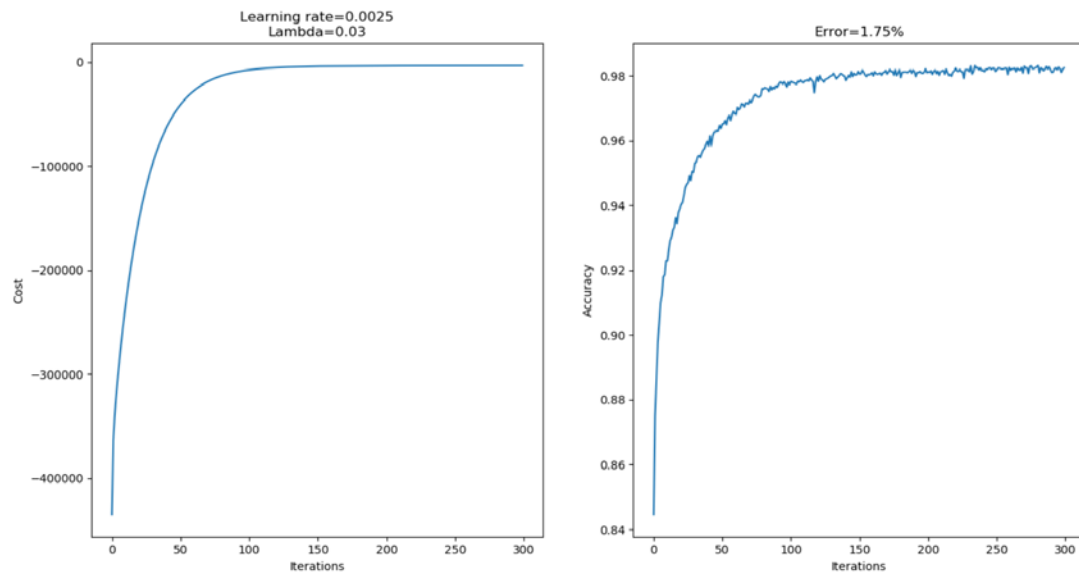
Error = 2.05%



β) Συνάρτηση Ενεργοποίησης $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

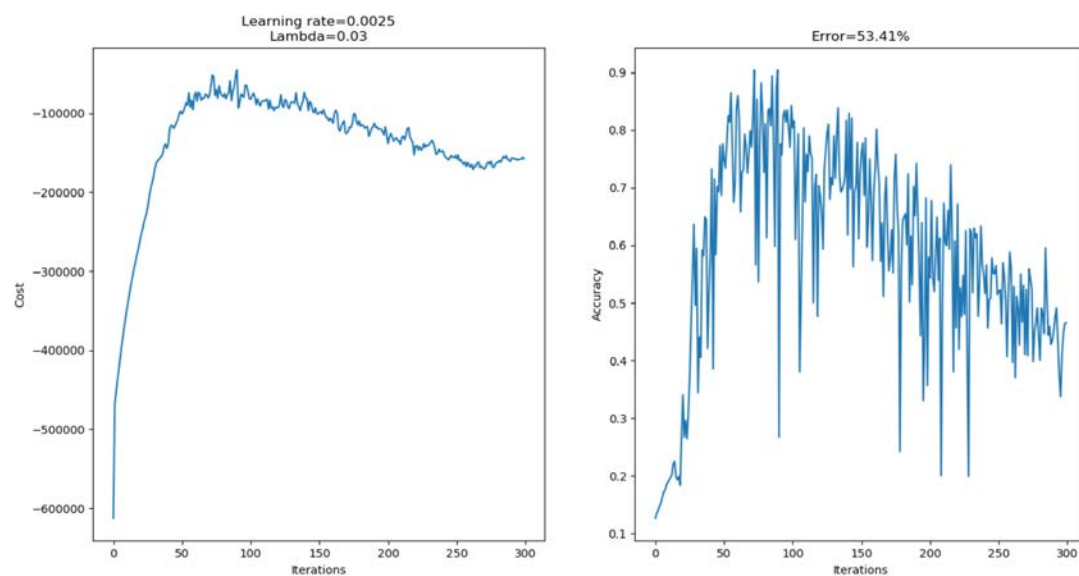
Error = 1.75%



γ) Συνάρτηση Ενεργοποίησης $h(a) = \cos(a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

Error = 53.41%

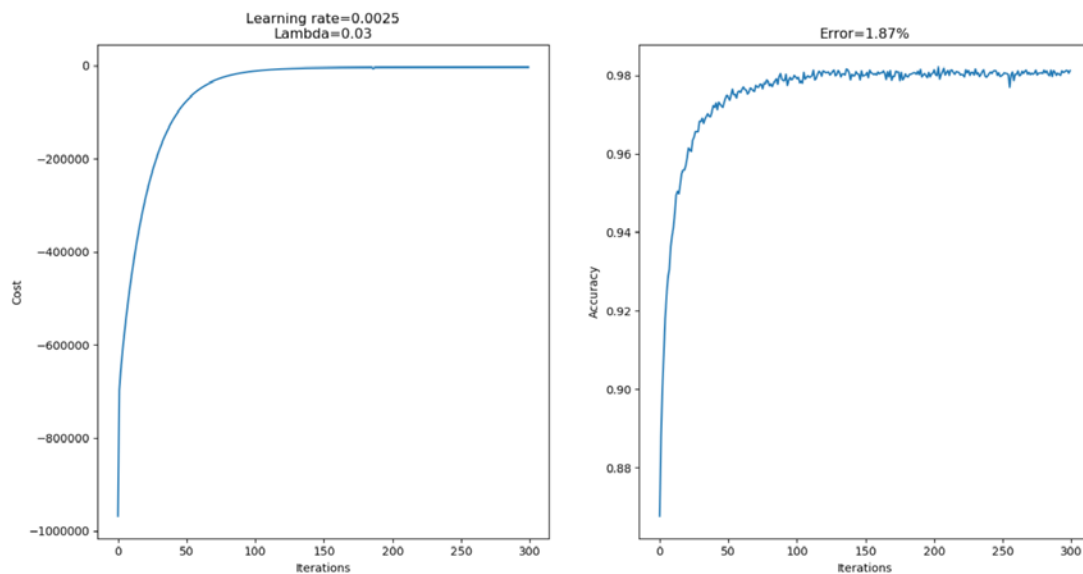


2.2 Αποτελέσματα για M=200 hidden units

α) Συνάρτηση Ενεργοποίησης $h(a) = \log(1 + e^a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

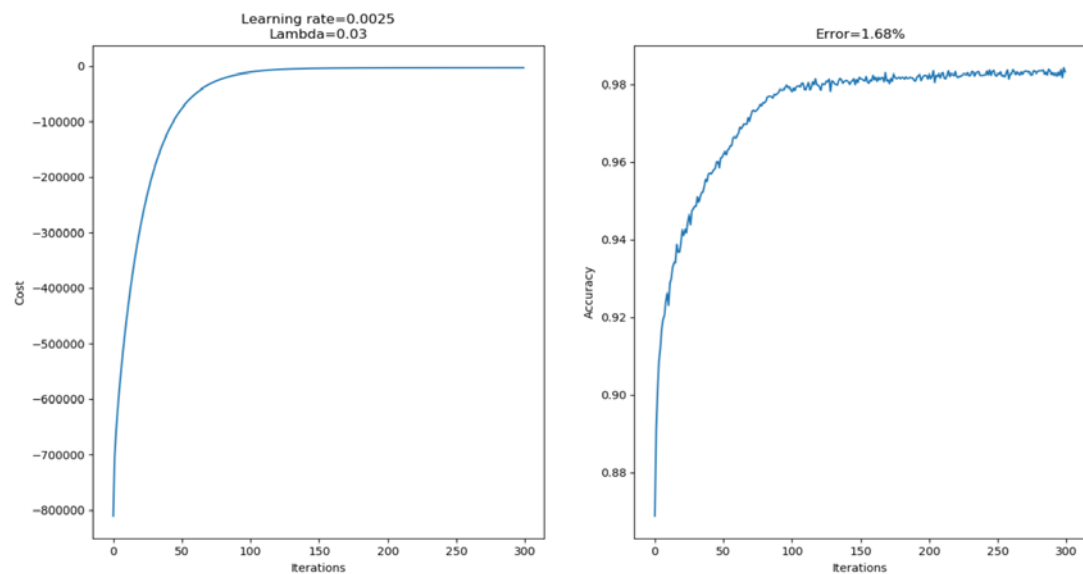
Error = 1.87%



β) Συνάρτηση Ενεργοποίησης $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

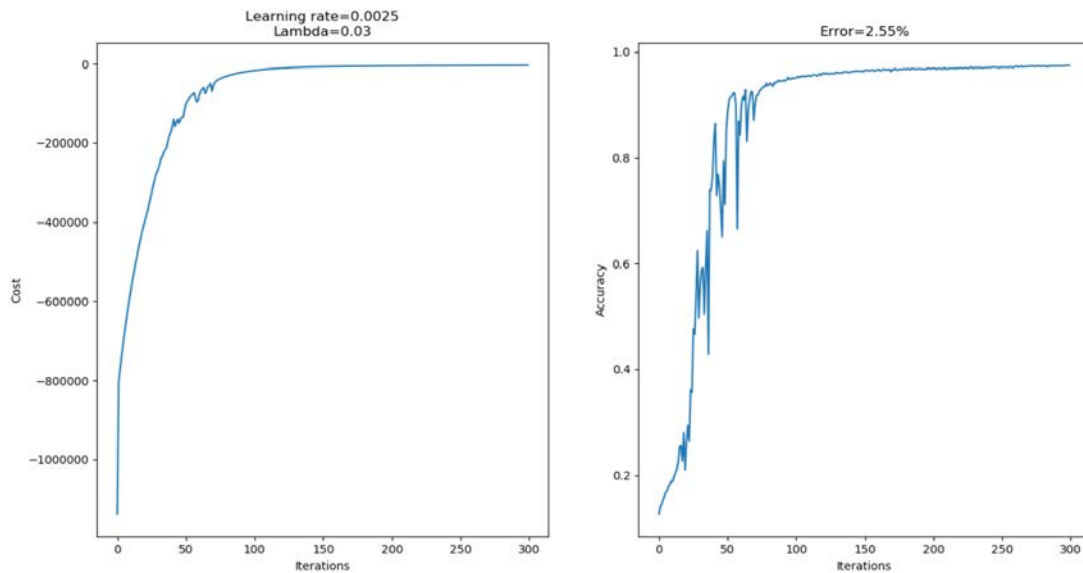
Error = 1.68%



γ) Συνάρτηση Ενεργοποίησης $h(a) = \cos(a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

Error = 2.55%

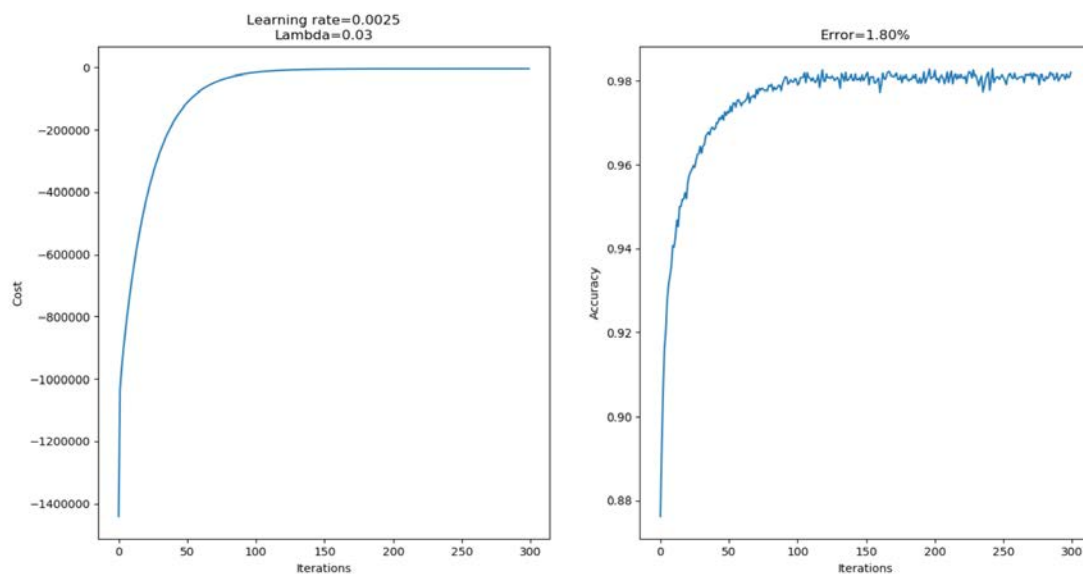


2.3 Αποτελέσματα για M=300 hidden units

α) Συνάρτηση Ενεργοποίησης $h(a) = \log(1 + e^a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

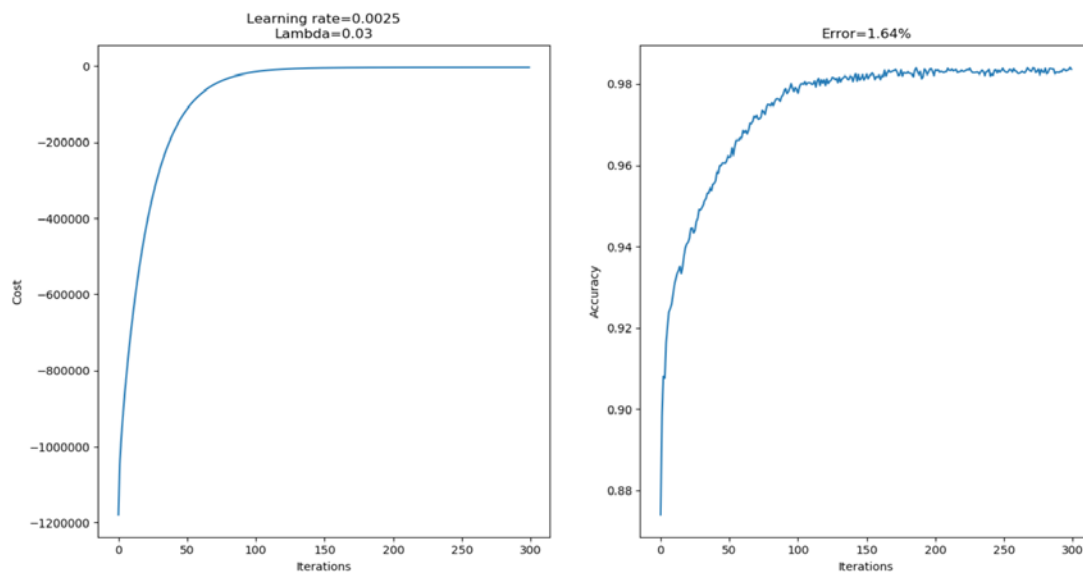
Error = 1.80%



β) Συνάρτηση Ενεργοποίησης $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

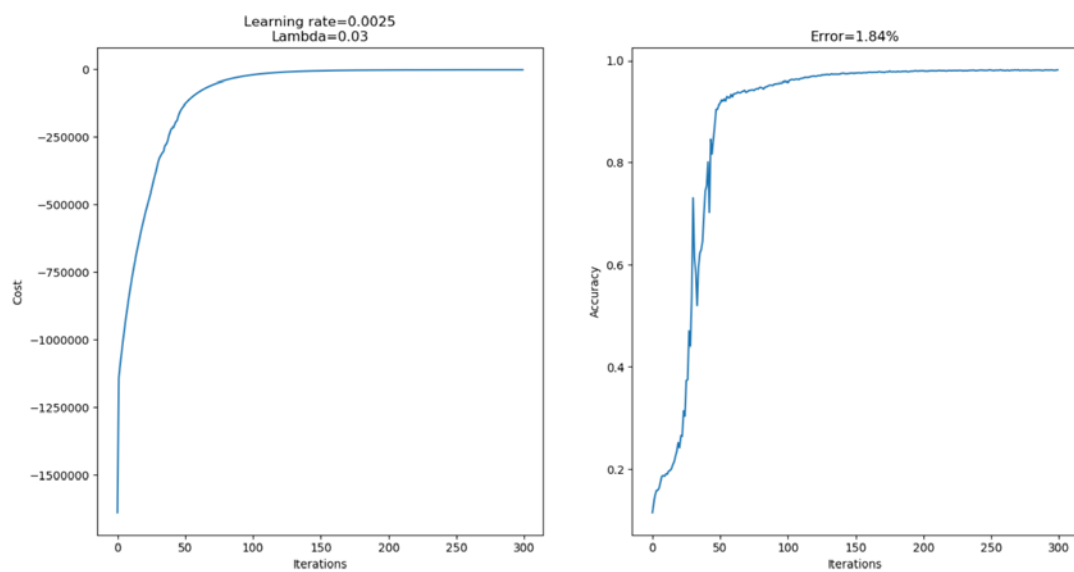
Error = 1.64%



γ) Συνάρτηση Ενεργοποίησης $h(a) = \cos(a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

Error = 1.84%



3. Παραδείγματα Εφαρμογής για το CIFAR-10

Σε όλα τα παραδείγματα χρησιμοποιείται οι εξής παράμετροι:

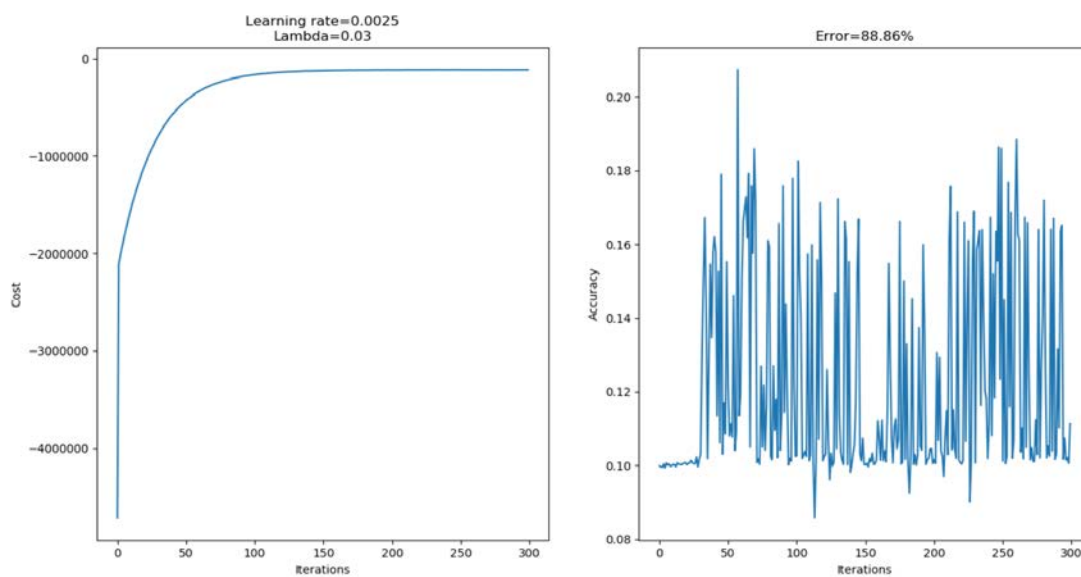
- Μέγεθος minibatch = 200
- Αριθμός επαναλήψεων = 300
- Learning rate $\eta = 0.0025$ (0.5/μέγεθος minibatch)
- $\lambda = 0.03$

3.1 Αποτελέσματα για M=100 hidden units

α) Συνάρτηση Ενεργοποίησης $h(a) = \log(1 + e^a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

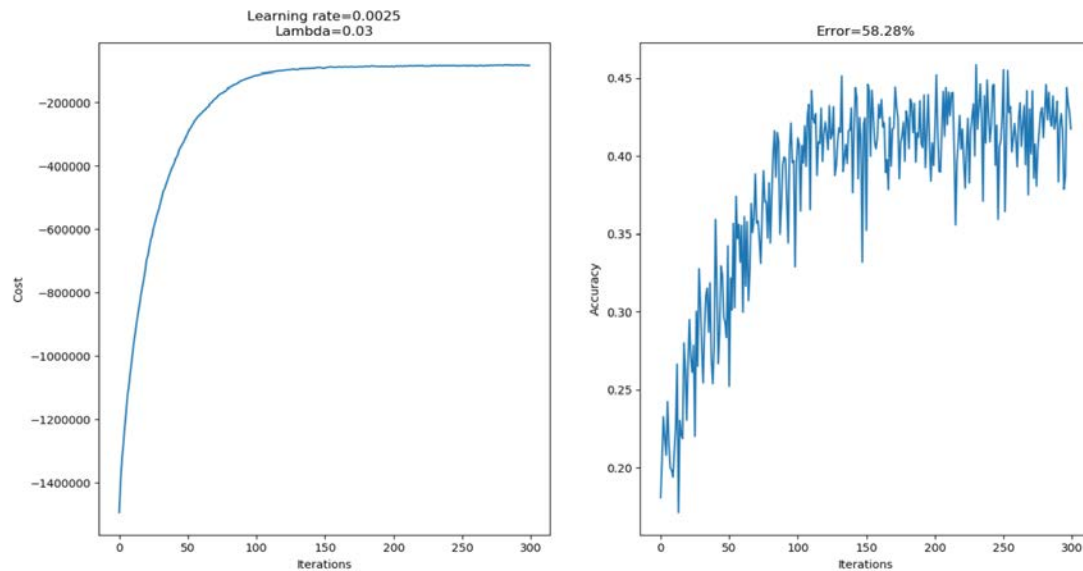
Error = 88.86%



β) Συνάρτηση Ενεργοποίησης $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

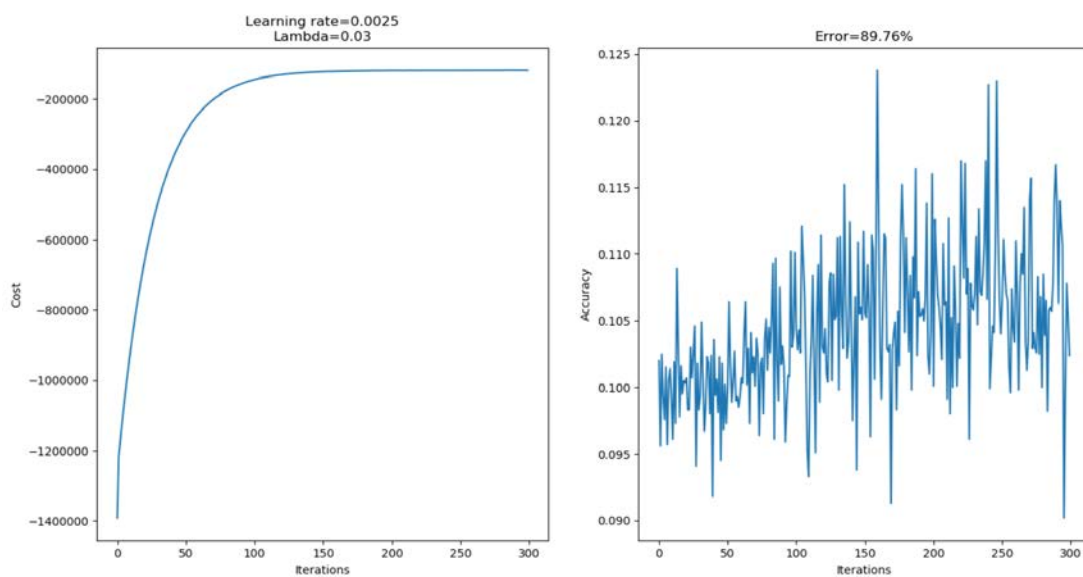
Error = 58.28%



γ) Συνάρτηση Ενεργοποίησης $h(a) = \cos(a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

Error = 89.76%

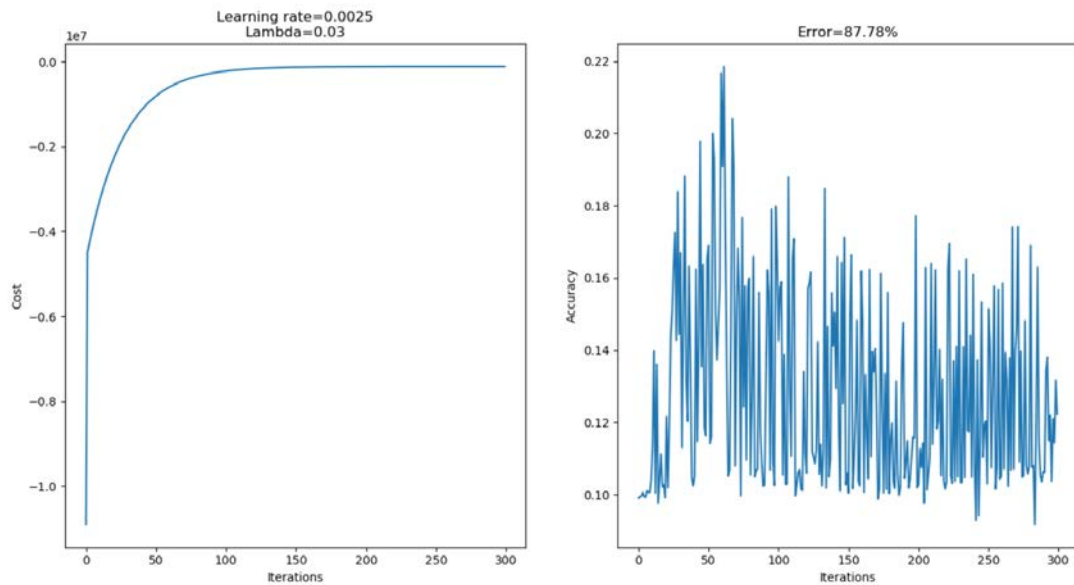


3.2 Αποτελέσματα για M=200 hidden units

α) Συνάρτηση Ενεργοποίησης $h(a) = \log(1 + e^a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

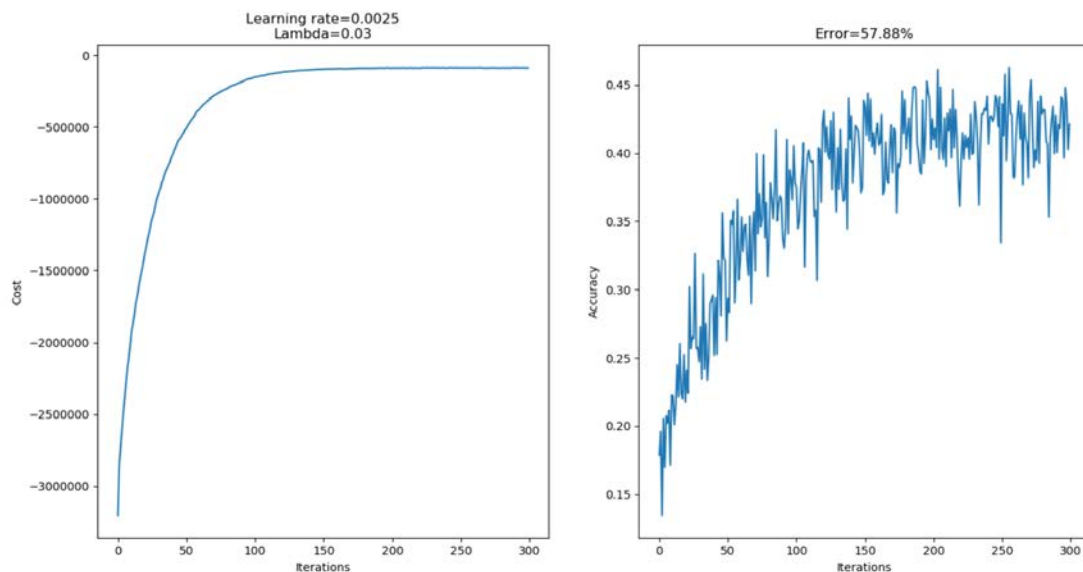
Error = 87.78%



β) Συνάρτηση Ενεργοποίησης $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

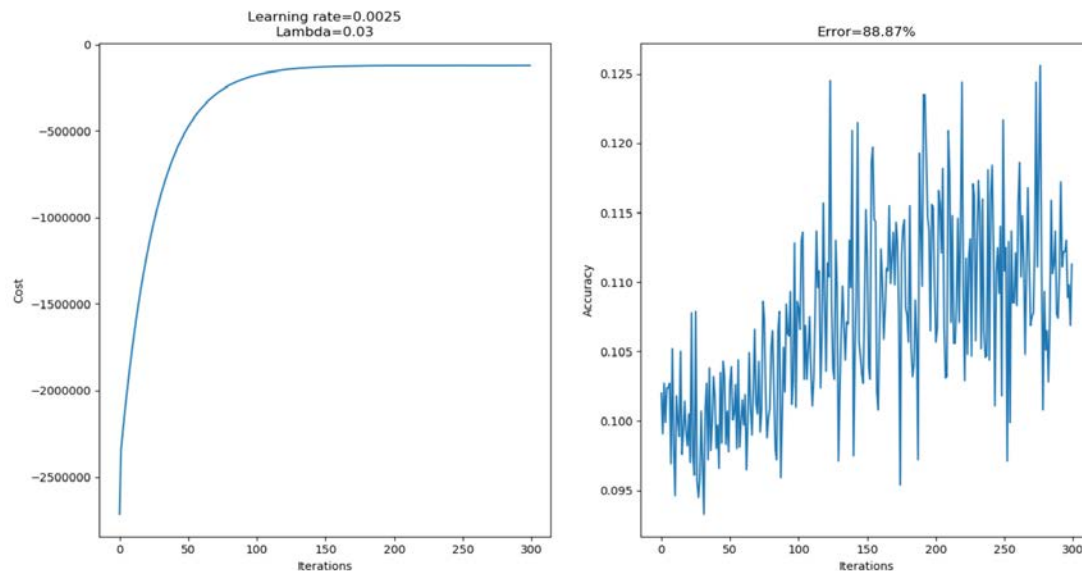
Error = 57.88%



γ) Συνάρτηση Ενεργοποίησης $h(a) = \cos(a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

Error = 88.87%

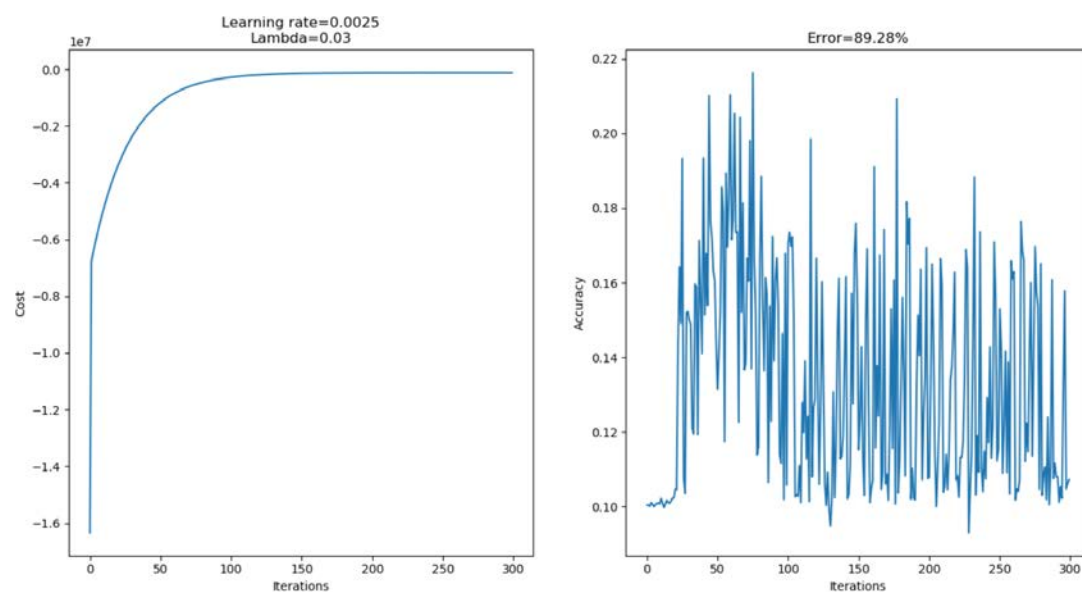


3.3 Αποτελέσματα για M=300 hidden units

α) Συνάρτηση Ενεργοποίησης $h(a) = \log(1 + e^a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

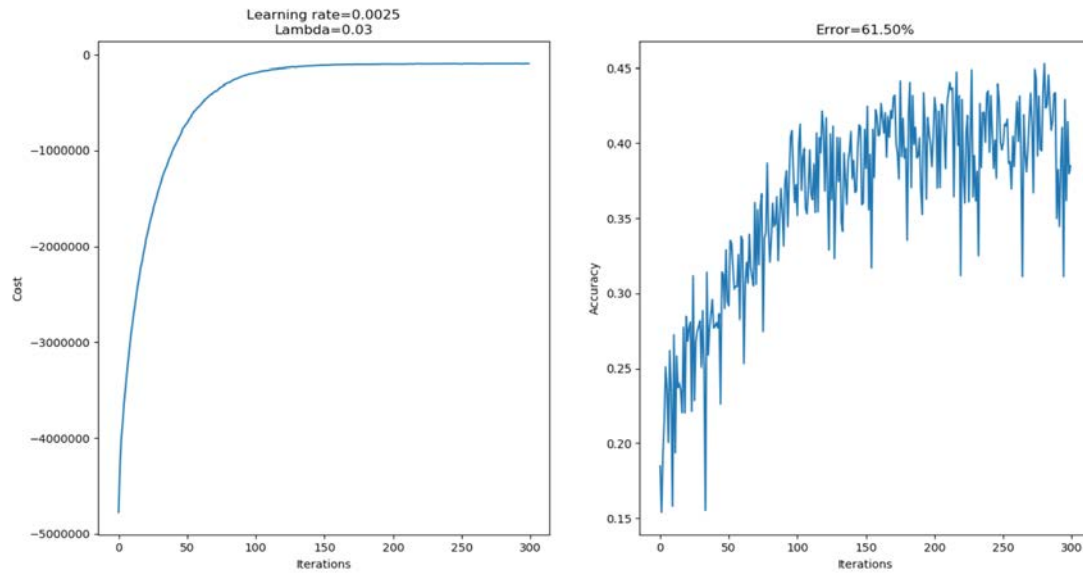
Error = 89.28%



β) Συνάρτηση Ενεργοποίησης $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

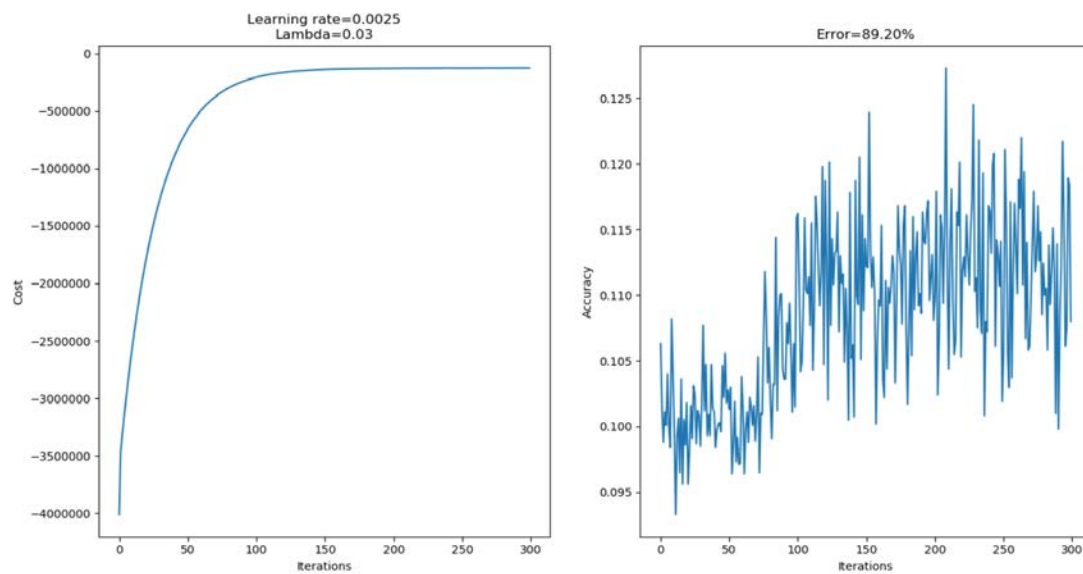
Error = 61.50%



γ) Συνάρτηση Ενεργοποίησης $h(a) = \cos(a)$

Επαναλήψεις = 300, $\eta = 0.0025$, $\lambda = 0.03$

Error = 89.20%



4. Καλύτερα Αποτελέσματα

MNIST

Τα καλύτερα αποτελέσματα παρουσιάστηκαν για **M = 300 hidden layers** και συνάρτηση ενεργοποίησης την $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Error = 1.64%

CIFAR

Τα καλύτερα αποτελέσματα παρουσιάστηκαν για **M = 200 hidden layers** και συνάρτηση ενεργοποίησης την $h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

Error = 57.88%