

Νευρωνικά Δίκτυα και Βαθιά Μάθηση
1. Σειρά Αναλυτικών Ασκήσεων

Ονοματεπώνυμο: Καπουγέρης Χάρης Ιάκωβος
Α.Μ.: 03120098
email: el20098@mail.ntua.gr

Άσκηση 1 (Multi Layer Perceptron - Regularization)

1.)

- Περίπτωση Α: Το μοντέλο υπερεκπαίδευεται (overfitting) στα δεδομένα εκπαίδευσης. Αυτό φαίνεται από το γεγονός ότι το σφάλμα στο σύνολο validation (L_v) αρχίζει να αυξάνεται μετά από ένα σημείο, ενώ το σφάλμα στο σύνολο training (L_T) συνεχίζει να μειώνεται.
- Περίπτωση Β: Το μοντέλο φαίνεται σαν να δεν υπερεκπαίδευεται. Το σφάλμα τόσο στο σύνολο εκπαίδευσης όσο και στο σύνολο επικύρωσης μειώνεται και σταθεροποιείται με την πάροδο των εποχών.

2.)

Περίπτωση Α: Η καμπύλη της επαναλήψεων είναι πριν το σημείο όπου το L_v αρχίζει να αυξάνεται. Αυτό το σημείο είναι όπου η καμπύλη του L_v είναι στο χαμηλότερο σημείο πριν αρχίσει να αυξάνεται. Αν υποθέσουμε ότι στο γραφικό φαίνεται η εποχή, πατάμε περίπου στο $\frac{1}{2}$.

Περίπτωση Β: Η καμπύλη της επαναλήψεων είναι εκεί που και οι 2 καμπύλες L_T και L_v έχουν σταθεροποιηθεί και δεν αλλάζουν σημαντικά με τις επόμενες επαναλήψεις. Αν στο γραφικό φαίνεται η εποχή, μπορεί να πατάμε είτε στο $\frac{1}{2}$ ή στο $\frac{2}{3}$ ή στο $\frac{3}{4}$.

3) • Dropout: Προσθέτοντας dropout layers μπορεί να μειωθεί η πιθανότητα υπερεκπαίδευσης, αναγκάζοντας το μοντέλο να γενικεύει καλύτερα.

• Early Stopping: Χρησιμοποιώντας early stopping, μπορούμε να σταματήσουμε την εκπαίδευση όταν το $Loss$ αρχίσει να αυξάνεται, αποφεύγοντας την υπερεκπαίδευση.

Οιόντως θα μπορούσατε να χρησιμοποιήσετε τεχνικές όπως Data-Augmentation, L^1 Regularization.

4) Το σύνολο ελέγχου (testing set) είναι απαραίτητο για να έχετε μια ανεξάρτητη εκτίμηση της επίδοσης του μοντέλου. Ενώ το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση του μοντέλου και το σύνολο επικύρωσης για τη ρύθμιση των υπερπαραμέτρων και την αποφυγή υπερεκπαίδευσης, το σύνολο ελέγχου παίζει για τελική εκτίμηση της γενικεύσεως του μοντέλου σε άγνωστα δεδομένα. Αλλά είναι κρίσιμο για να διαπιστώσετε πως θα αποδίδει το μοντέλο σε πραγματικές συνθήκες.

Άσκηση 2 (Representation Learning - Autoencoders)

- a) Έχουμε U_0, U_c (256×1) και 5 κρυφά επίπεδα $[500 \ 250 \ 50 \ 250 \ 500]$.
Κάθε layer αναπαριστάται από ένα διάνυσμα με 256 στοιχεία πίσω του skipgram.
• Άρα η είσοδος του auto-encoder θα είναι $\dim(x_i) = 256$
- β) Το μέγεθος του y_i θα είναι ίδιο με το μέγεθος της εισόδου: $\dim(y_i) = \dim(x_i) = 256$
- γ) Η συνδεδεμένη αναπαράσταση είναι ίση με το μέγεθος της στρώσης στο κέντρο του auto-encoder, δηλαδή 50

Άσκηση 3 (Recurrent Neural Networks)

a)
$$C = \frac{1}{4} \sum_{j=1}^4 i_j = \frac{1}{4} ([4, 0]^T, [0, 4]^T, [0, 0]^T, [0, 0]^T) = [1, 1]^T$$

- Υπολογισμός της κρυφής καταστάσης και της εξόδου σε κάθε χρονικό βήμα.

$$\boxed{\rightarrow t=1}$$

• Είσοδος: $x_1 = [0, 0, 1, 1]^T$

Hidden state:

$$h_1 = \text{ReLU}(W_{hx} \cdot x_1 + W_{hh} \cdot h_0 + \text{bias}_{in}) =$$

$$= \text{ReLU} \left(\begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot 0 + 0 \right) =$$

$$= \text{ReLU} \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

• Εξόδος :

$$y_1 = (W_{yh} \cdot h_1 + \text{bias}_{out}) = \begin{bmatrix} -5 & -5 \\ 0 & 3 \\ 1 & 2 \\ 2 & 2 \\ 3 & -1 \\ 2.9 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -10 \\ 3 \\ 3 \\ 4 \\ 2 \\ 2.9 \end{bmatrix}$$

Η τελευταία παράσταση είναι η 4, στην 4η γραμμή
η 2.9 στην οποία αντιστοιχεί ένα Cat

$\rightarrow t=2$

• Είσοδος: $x_2 = [1, -2, 0, 1, 1]^T$

Hidden State :

$$h_2 = \text{ReLU} (W_{hx} \cdot x_2 + W_{hh} \cdot h_1 + \text{bias}_{in}) =$$

$$= \text{ReLU} \left(\begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0 \right) =$$

$$= \text{ReLU} \left(\begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

• Εξόδο:

$$y_2 = (W_{yh} \cdot h_2 + \text{bias}_{out}) = \begin{bmatrix} -5 & -5 \\ 0 & 3 \\ 1 & 2 \\ 2 & 2 \\ 3 & -1 \\ 2.9 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -15 \\ 0 \\ 3 \\ 6 \\ 9 \\ 8.7 \end{bmatrix}$$

Η βεβαιότητα πιθανότητας είναι ~ 9 , οπότε
 $\hat{y} =$ pattern, οπότε αντιστοιχεί στο λέξη staring.

$$\rightarrow \epsilon = 3$$

• Είσοδος: $x_3 = [0, -1, 1, 1]^T$

• Hidden State:

$$\begin{aligned} h_3 &= \text{ReLU}(W_{hx} \cdot x_3 + W_{hh} \cdot h_2 + \text{bias}_{in}) = \\ &= \text{ReLU}\left(\begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \end{bmatrix} + 0\right) = \\ &= \text{ReLU}\left(\begin{bmatrix} 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \end{aligned}$$

• Εξόδο:

$$y_3 = (W_{yh} \cdot h_3 + \text{bias}_{out}) = \begin{bmatrix} -5 & -5 \\ 0 & 3 \\ 1 & 2 \\ 2 & 2 \\ 3 & -1 \\ 2.9 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -10 \\ 6 \\ 4 \\ 4 \\ -2 \\ 0 \end{bmatrix}$$

Η βεβαιότητα πιθανότητας αντιστοιχεί στο $\hat{y} =$ pattern (6)
 και αντιστοιχεί στο λέξη <stop>.

B) Attention Scores

$$\begin{aligned} \rightarrow \text{Score} &= Q \cdot K^T = h_0 \cdot [i_1, i_2, i_3, i_4]^T \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 4 & 0 \\ 0 & 4 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

$$\rightarrow \text{Scaled Score} = \frac{\text{Score}}{\sqrt{d_k}} = \frac{\begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}}{\sqrt{2}} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$$

\rightarrow Attention Weights:

$$\begin{aligned} P &= \text{softmax}(\text{Scaled Score}) = \text{softmax}(\begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}) \\ &= \begin{bmatrix} e^{0/4} & e^{0/4} & e^{0/4} & e^{0/4} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} \end{aligned}$$

Teđika:

$$\begin{aligned} Z &= P \cdot V = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 4 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

Kazadijate mo isto stavuoka. Enofirws y
Nijn napafiveri Cat

Άσκηση 4 (Convolutional Neural Networks)

a)

$$W_2 = \frac{(W_1 - f + 2P)}{S} + 1 \stackrel{P=0}{\Rightarrow} W_2 = \frac{W_1 - f + 1}{S}$$

$$H_2 = \frac{(H_1 - f + 2P)}{S} + 1 \Rightarrow H_2 = \frac{H_1 - f + 1}{S}$$

$$P_2 = k \cdot P_1$$

$$\left\{ \begin{array}{l} W_2 = \frac{227-11}{4} + 1 \\ H_2 = \frac{227-11}{4} + 1 \\ P_2 = 1 \text{ για 1 φίλτρο} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} W_2 = 55 \\ H_2 = 55 \\ P_2 = 1 \end{array} \right.$$

για $k=96$ φίλτρα έχουμε $P_2 = 96$
 οπότε $55 \times 55 \times 96$

β) Έχουμε $55 \cdot 55 = 3.025$ units/filter οπότε οπότε
 $3.025 \cdot 96 = 290.400$ units

γ) Για το επόμενο layer έχουμε:

$$(width \times height \times channels + 1) \times filters =$$

$$(11 \cdot 11 \cdot 3 + 1) \cdot 96 = 34.944 \text{ επιπλέον παραγόμενες} \text{ } \rightarrow$$

δ) Παίρνουμε flattened input $227 \cdot 227 \cdot 3 = 154.587$

παραδοσ εισόδου στο FF δίκτυο αφού έχουμε
256 units τε 256 biases θα έχουμε
συνολικά:

$$154.587 \cdot 256 + 256 = 39.574.528 \text{ εκπαιδευτικές παραμέτρους}$$

Άσκηση 5 (Generative models)

1) Variational Autoencoders (VAEs)

• Γενικεύσεις Αρχές

- Αρχιτεκτονική: Τα VAEs αποτελούνται από έναν encoder και έναν decoder. Ο encoder μετατρέπει την είσοδο σε μια μειωμένη έκδοση στο latent space και ο decoder παράγει την έξοδο από αυτή την κατάσταση.
- Ενός: Εκπαίδευση της κατανομής των δεδομένων μέσω της μεγιστοποίησης της κλίσης της προοπτικής της πιθανότητας των δεδομένων.

• Διαδικασία Εκπαίδευσης:

- Loss Function: Ο συνδυασμός reconstruction loss και Kullback-Leibler (KL) divergence. Το reconstruction loss μετράει τη διαφορά μεταξύ της πραγματικής και της παραχόμενης εικόνας ενώ ο KL divergence μετράει την απόσταση της κατανομής στο latent space από την προκαθορισμένη κανονική κατανομή.
- Εκπαίδευση: Backpropagation για τη μείωση της συνολικής απώλειας.

- Απαιτήσεις:

- Μνήμη/Χρόνος: Αναλόγως με το μέγεθος δαγών και την πολυπλοκότητα των δεδομένων. Αναμενών ονέως υπολογιστικής πόρας σε σχέση με GANs και Diffusion Models.
- Δεδομένα Εκπαίδευσης: Αναμενών μεγάλη ποσότητα δεδομένων για καλή απόδοση

- Αποεξελεγκτικότητα:

- Εφαρμογές: Image generation, denoising

- Προεξέλιξη: Ευθεία εκπαίδευση, ικανότητα για μεταβολή του latent space.

- Μειονεκτήματα: Θορύβη παραγόμενα δείγματα σε σχέση με τα άλλα μοντέλα.

2) [GANs]

- Σεβελιώδης Αρχή:

- Αρχιτεκτονική: Αποτελείται από 2 αναμνησικά δίκτυα, τον Generator και τον Discriminator. Ο Generator προσπαθεί να παράγει ψεύτικες αλλά ρεαλιστικές εικόνες ενώ ο Discriminator να διακρίνει ψεύδεις από πραγματικές εικόνες.

- Επίσκο: Ο Generator προσπαθεί να εκπαιδευτεί με δεδομένα που προσπαθούν να "ξεγεμίσουν" τον Discriminator, ο οποίος προσπαθεί να διακρίνει τις πραγματικές από τις ψεύτικες εικόνες.

Διαδικασία Εκπαίδευσης

- Loss Function: Binary cross-entropy loss για τον Discriminator και συνήθως για συνδυασμένη loss για τον Generator που περιλαμβάνει την binary cross-entropy loss αλλά μπορεί να περιλαμβάνει και άλλες παραλλαγές.

- Εκπαίδευση: Evaluating εκπαίδευση generator και Discrim.

Κάθε φορά που ο Discriminator γίνεται καλύτερος στη διακρίση των εισόδων ο Generator προσαρμόζεται για να εκπαιδευτεί καλύτερα δείχνοντας.

Αναζητήσεις

- Μνήμη/Χρόνος: Υψηλές απαιτήσεις σε ισχύ και χρόνο

λόγω της ανάγκης εκπαίδευσης 2 δικτύων

- Δεδομένα Εκπαίδευσης: Χρειάζονται μεγάλη ποσότητα δεδομένων υψηλής ποιότητας.

Ανοητείες

- Προβλήματα: Υψηλή μνήμη και πραγματικότητα των παραγόμενων δεδομένων

- Μειονεκτήματα: Δυσκολία εκπαίδευσης, πρόβλημα

mode collapse, όταν ο Generator παράγει μόνο περιορισμένο αριθμό διαφορετικών δειγμάτων.

3) Diffusion Models

• Θεμελιώδεις Αρχές

- Αρχιτεκτονική: Βασίζονται σε μια διαδικασία που ξεκλεί σταδιακά θόρυβο στα δεδομένα εισόδου και στη συνέχεια αναστρέφει αυτή τη διαδικασία για να δημιουργήσει νέα δεδομένα. Η διαδικασία αποτελείται από πολλές βήματα, όπου κάθε βήμα προσθέτει μικρό θόρυβο στην εικόνα.
- Στόχος: Μάθηση της αντιστροφής διαδικασίας της διάχυσης του θορύβου για τη δημιουργία καθαρών δεδομένων από θόρυβο.

• Διαδικασία Εκπαίδευσης

- Loss Function: Χρήση ενός τύπου loss που μετράει τη διαφορά μεταξύ της πραγματικής κατανομής και της κατανομής που παράγει το μοντέλο σε κάθε βήμα διάχυσης.
- Εκπαίδευση: Περιλαμβάνει την εκπαίδευση της αντιστροφής διαδικασίας διάχυσης μέσω πολλών παθών. Αυτό σημαίνει ότι το μοντέλο εκπαιδεύεται να αφαιρεί το θόρυβο από τα δεδομένα και να επιβεβαιώνει καθαρή εικόνα.

Παρατηρίες

- Μνήμη/Χρήσιμος: Υψηλές απαιτήσεις σε ισχύ λόγω της πολυάριθμης διαδικασίας εκπαίδευσης και των

πολλών βημάτων διαχύσιμα.

- Δεδομένα Εκπαίδευσης: Απαιτούν τεράστια ποσότητα δεδομένων και συνήθως πολλά βήματα εκπαίδευσης για καλή απόδοση.

Αποελευθεύσιμα

- Πλεονεκτήματα: Λαμβάνουν κομμάτια δεδομένων και υψηλής ποσότητας εικόνες.
- Μειονεκτήματα: Πολύπλοκη και χρονοβόρα εκπαίδευση, απαιτούν πολλούς πόρους.

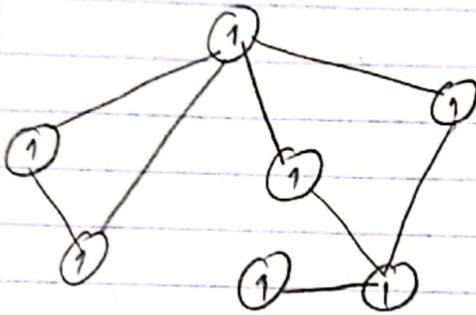
→ Τελικά για την ευχερή χρήση είναι εύκολο:

Χαρακτηριστικά	VAEs	GANs	Diffusion Models
Αρχιτεκτονική	Encoder-Decoder	Generator-Discrim.	Σειρά των βημάτων
Στόχος	Μείωση πιθανότητας \neq KL divergence	Αντικ. Εκπ.	Εκπαίδευση αντιστροφής
Loss Function	Reconst. loss + λ	Binary Cross entropy + adversarial loss	Διαφορά κατανοών αναπαραγωγής
Εκπαίδευση	Backpropagation	Excluz. Gen-Disc	Πολλοί βήματα
Απαιτούμ. Μνήμης/Χρόνου	Μέτριοι	Υψηλές	Πολλοί Υψηλές
// Δεδομένα	Μεγάλη ποσότητα	Υψηλή ποσότητα	Μεγάλη ποσότητα
Ποσότητα Αποελευθεύσιμων	Θαλάσσιες	Παράδειγμα + εικόνες	Υψηλή ποσότητα και πολλαπλά
Πλεονεκτήματα	Εύκολη εκπαίδευση	Υψηλή ποσότητα	Παράδειγμα βήματα
Μειονεκτήματα	Θαλάσσιες	mode collapse, ασταθής εκπ.	Πολύπλοκη και χρονοβόρα εκπαίδευση

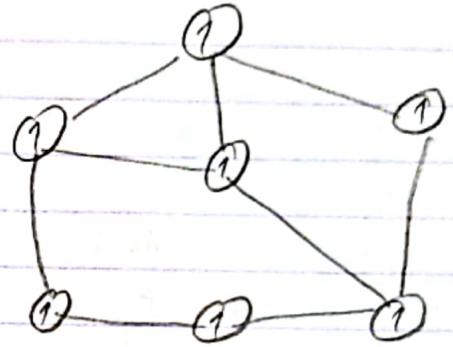
Άσκηση 6 (Graph Neural Networks)

Βήμα 1: Απεικόνιση

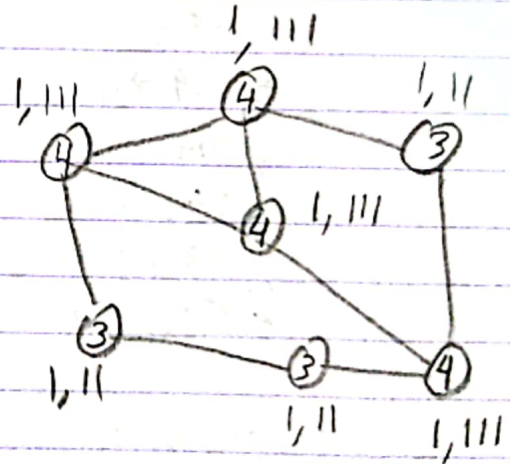
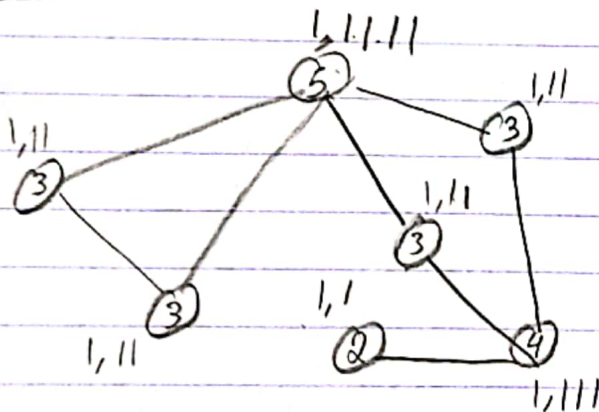
G_1



G_2



Βήμα 2ο: 1^η Εναγωγή



Hash

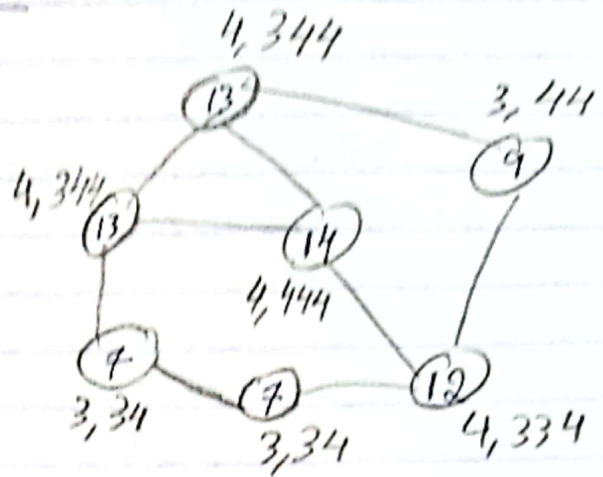
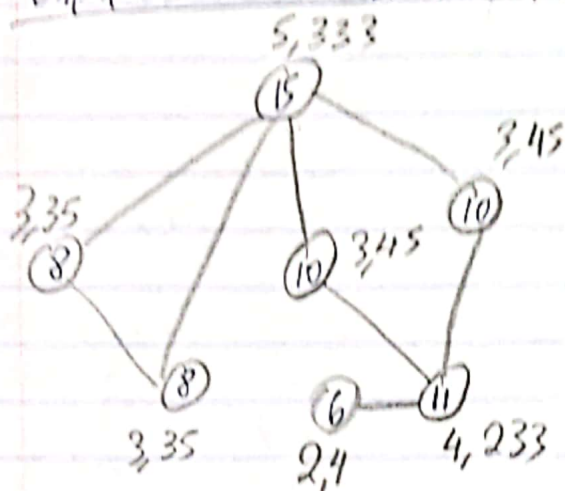
$1, 1 \rightarrow 2$

$1, 1, 1 \rightarrow 3$

$1, 1, 1, 1 \rightarrow 4$

$1, 1, 1, 1, 1 \rightarrow 5$

Βήμα 38: 2η Επανάληψη.



Hash

$2,4 \rightarrow 6$
 $3,34 \rightarrow 7$
 $3,35 \rightarrow 8$
 $3,44 \rightarrow 9$
 $3,45 \rightarrow 10$
 $4,233 \rightarrow 11$
 $4,334 \rightarrow 12$
 $4,344 \rightarrow 13$
 $4,444 \rightarrow 14$
 $5,333 \rightarrow 15$

→ Βλέπουμε πως οι 2 γραφοί δεν έχουν καμία κοινή ετικέτα μεταξύ τους, οπότε σταμάτατε το iterations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\phi(G_1)$	7	1	4	1	1	1	0	2	0	2	1	0	0	0	1
$\phi(G_2)$	7	0	3	4	0	0	2	0	1	0	0	1	2	1	0
Προϊκτο	49	0	12	4	0	0	0	0	0	0	0	0	0	0	0

$$K(\phi(G_1), \phi(G_2)) = \phi(G_1)^T \cdot \phi(G_2) = 49 + 12 + 4 = 65$$