

ΔΕΥΤΕΡΑ 21/06/22 - «ΒΑΘΙΑ ΜΑΘΗΣΗ» - ΓΡΑΠΤΗ ΕΞΕΤΑΣΗ

ΕΠΩΝΥΜΟ:

ΟΝΟΜΑ:

ΙΔΙΟΤΗΤΑ:

ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ:

Σε παρένθεση τα μόρια κάθε θέματος (σύνολο 102 μόρια)

**Θέμα 01 (6)**

Περιγράψτε, σύντομα, τη λειτουργία του self-attention και εξηγήστε ποια είναι η διαφορά του από το multi-head attention.

**Θέμα 02 (3)**

Ποια είναι η λειτουργία του residual connection; (Πολλαπλές απαντήσεις)

A. Βοηθά στη προσαρμογή των βαρών των νευρώνων, μειώνοντας το υπολογιστικό κόστος.

☒ B. Βελτιώνει τη συνολική απόδοση σε ένα ευρύ φάσμα εφαρμογών, αυξάνοντας το υπολογιστικό κόστος.

☒ Γ. Παρέχει μια άλλη διαδρομή στα δεδομένα με σκοπό να φτάσουν στα τελευταία επίπεδα του νευρωνικού δικτύου παρακάμπτοντας ορισμένα επίπεδα.

Δ. Επιτυγχάνει την μεταφορά γνώσης από προεκπαιδευμένα μοντέλα

**Θέμα 03 (3)**

Σημείωση: στις ερωτήσεις με αστερίσκο \* μπορείτε να επιλέξετε περισσότερες από μία απαντήσεις

I. Το PCA είναι τεχνική για

(0.5 μονάδα)

A. Εύρεση ορθογώνιων διανυσμάτων του συνόλου δεδομένων

B. Εκτίμηση του αριθμού των διαστάσεων του συνόλου δεδομένων

☒ Γ. Εύρεση των πιο σημαντικών βάσεων για αναπαράσταση του συνόλου δεδομένων

Δ. Εύρεση στοιχείων του συνόλου δεδομένων

(0.5 μονάδα) \*

- II. Τα PCA είναι τεχνική για
- ☒ Α. Μείωση διαστάσεων
  - ☐ Β. Εκπαίδευση δεδομένων
  - ☐ Γ. Ελέγχους χαρακτηριστικών
  - ☐ Δ. Κατανομοποίηση διακύμανσης

III. Έστω  $X: X = U\Sigma V$  ένας SVD πίνακας.

(1 μονάδα) \*

Ποια από τα παρακάτω ισχύει για όλα τα  $X$ ;

- ☒ Α.  $\text{rank}(X) = \text{rank}(\Sigma)$ .
- ☐ Β. Εάν όλες οι singular values είναι μοναδικές, τότε ο SVD είναι μοναδικός.
- ☐ Γ. Η πρώτη στήλη του πίνακα  $V$  είναι ιδιοδιάνυσμα του  $X$ .
- ☐ Δ. Οι singular values και οι ιδιοτιμές του  $X^T X$  είναι ίδιες.

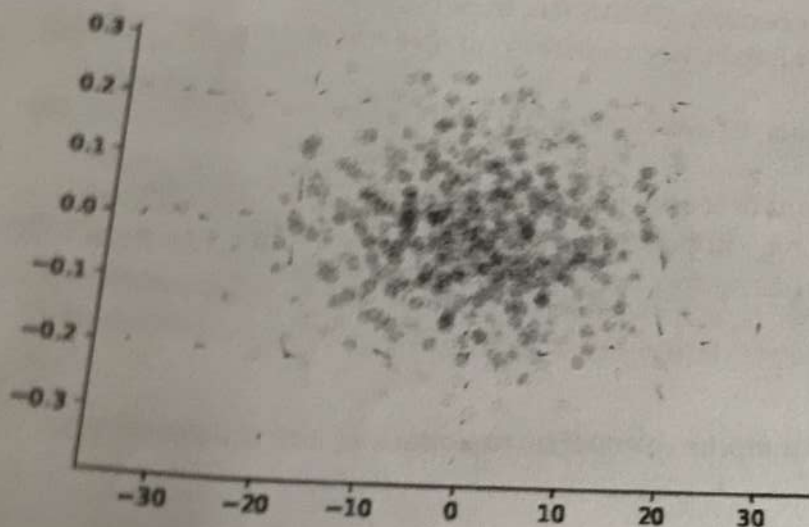
IV. Αν τα  $p_1$  και  $p_2$  είναι και τα δύο διανύσματα κύριων συνιστωσών, ποιες από τις ακόλουθες προτάσεις είναι σωστές

(1 μονάδα) \*

- ☒ Α. Το  $p_1$  είναι ορθογώνιο στο  $p_2$
- ☐ Β. Το  $p_1$  είναι παράλληλο στο  $p_2$
- ☐ Γ. Η διακύμανση κατά μήκος  $p_1$  είναι μεγαλύτερη από τη διακύμανση κατά μήκος  $p_2$
- ☐ Δ. Η διακύμανση κατά μήκος  $p_2$  είναι μεγαλύτερη από τη διακύμανση κατά μήκος  $p_1$

#### Θέμα 04 (6)

Στην ακόλουθη εικόνα απεικονίζονται 1.000 δείγματα που προέρχονται από μια δισδιάστατη πολυμεταβλητή κανονική κατανομή. Ποιος από τους παρακάτω πίνακες θα μπορούσε να είναι ο πίνακας συνδιακύμανσης της κατανομής; (Δώστε προσοχή στις τιμές στους άξονες!)



☒ Α.  $\begin{bmatrix} 100 & 0 \\ 0 & 0.01 \end{bmatrix}$

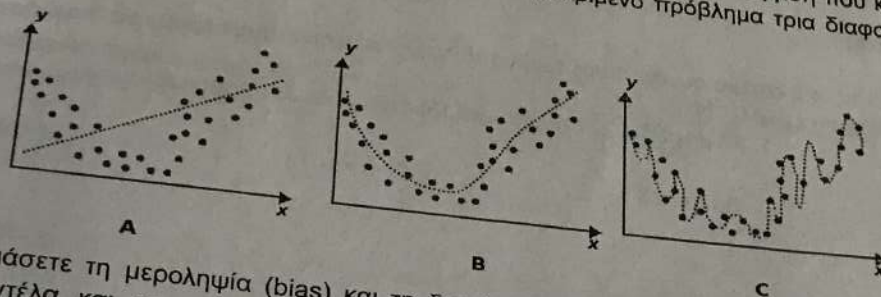
☐ Β.  $\begin{bmatrix} 10 & 0 \\ 0 & -0.1 \end{bmatrix}$

☐ Γ.  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$



Θέμα 05 (5)

Στο παρακάτω σχήμα, με μπλε κουκίδες απεικονίζονται τα δεδομένα ενός προβλήματος παλινδρόμησης μιας διάστασης, ενώ με διακεκομμένες γραμμές η προσέγγιση που κάνουν για τη σχέση μεταξύ εισόδου και εξόδου για το συγκεκριμένο πρόβλημα τρία διαφορετικά μοντέλα, τα A, B και C.



Να σχολιάσετε τη μεροληψία (bias) και τη διακύμανση (variance) που επιτυγχάνουν τα 3 αυτά μοντέλα και σε περίπτωση που σε κάποιο/α από αυτά δεν βρίσκετε τη σχέση μεροληψίας-διακύμανσης ιδανική, να προτείνετε συντομές, τεκμηριωμένες λύσεις αντιμετώπισης.

Θέμα 06 (4)

Επιλέγουμε να επιλύσουμε ένα πρόβλημα επιβλεπόμενης μάθησης μέσω bagging, χρησιμοποιώντας  $n$  διαφορετικά μοντέλα και παρατηρούμε ότι η συνδιασπορά των σφαλμάτων τους είναι η μισή από τη διασπορά τους. Ποια είναι η αναμενόμενη τιμή τετραγώνου του σφάλματος του συνολικού μοντέλου και ποια η σχέση της με την απόδοση των επιμέρους μοντέλων;

Θέμα 07 (6)

Ποιο είναι το αποτέλεσμα της προσθήκης όρου ποινής σε έναν υπερπλήρη αυτοκωδικοποιητή;

Θέμα 08 (3)

Ποια από τις παρακάτω προτάσεις είναι λανθασμένη

- A. Οι συστατικοί αυτοκωδικοποιητές μαθαίνουν καλύτερες αναπαραστάσεις των δεδομένων σε σύγκριση με την ανάλυση κυρίων συνιστωσών γιατί εκμεταλλεύονται την ιδιότητα διαμοιρασμού των παραμέτρων.
- B. Ένας υποπλήρης αυτοκωδικοποιητής με γραμμική συνάρτηση κωδικοποίησης καλύπτει τον ίδιο υποχώρο με την ανάλυση κυρίων συνιστωσών
- C. Η μη-επιβλεπόμενη προπαίδευση βασίζεται σε μοντέλα που μαθαίνουν λανθάνουσες αναπαραστάσεις της εισόδου τους
- ☒ D. Για να μπορέσουμε να χρησιμοποιήσουμε τεχνικές μάθησης μέσω μεταφοράς από το πεδίο A στο πεδίο B, θα πρέπει οι παράγοντες που εξηγούν τις διακυμάνσεις στο A να είναι σχετικές με τις διακυμάνσεις που αναμένουμε να εντοπίσουμε στο B



Θέμα 09 (6)

Περιγράψτε ένα πρόβλημα που μπορεί να εμφανιστεί κατά την εκπαίδευση ενός Wasserstein GAN (WGAN) καθώς και πιθανούς τρόπους αντιμετώπισής του

Θέμα 10 (6)

Ποιες από τις παρακάτω συναρτήσεις ενεργοποίησης διευκρινιστή μπορούν να "κολλήσουν" την εκπαίδευση ενός GAN

1. Σιγμοειδής, 2. Υπερβολική Εφαπτομένη 3. Γραμμική

A. 1 και 2

B. 2 και 3

C. 3 και 1

D. Όλες

Θέμα 11 (6)

Δίνονται οι παρακάτω πίνακες Q και R ενός συστήματος αυτοενισχυόμενης μάθησης βασισμένο στο Q-learning (στον πίνακα ανταμοιβών R, όσες εγγραφές έχουν τιμή -1 υποδηλώνουν μη-επιτρεπτά ζεύγη κατάστασης-δράσης).

State	Action					
	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

State	Action					
	0	1	2	3	4	5
0	0	0	0	0	80	0
1	0	0	0	64	0	100
2	0	0	0	64	0	0
3	0	80	51	0	80	0
4	64	0	0	64	0	100
5	0	80	0	0	80	100

Έστω ότι ο πράκτορας βρίσκεται στην κατάσταση 4. Ποια δράση θα επιλέξει ο πράκτορας με βάση τον κανόνα μετάβασης για την τεχνική επανάληψης τιμών και πως θα διαμορφωθεί ο πίνακας Q (θεωρήστε  $\gamma=0.85$ );

Θέμα 12 (3)

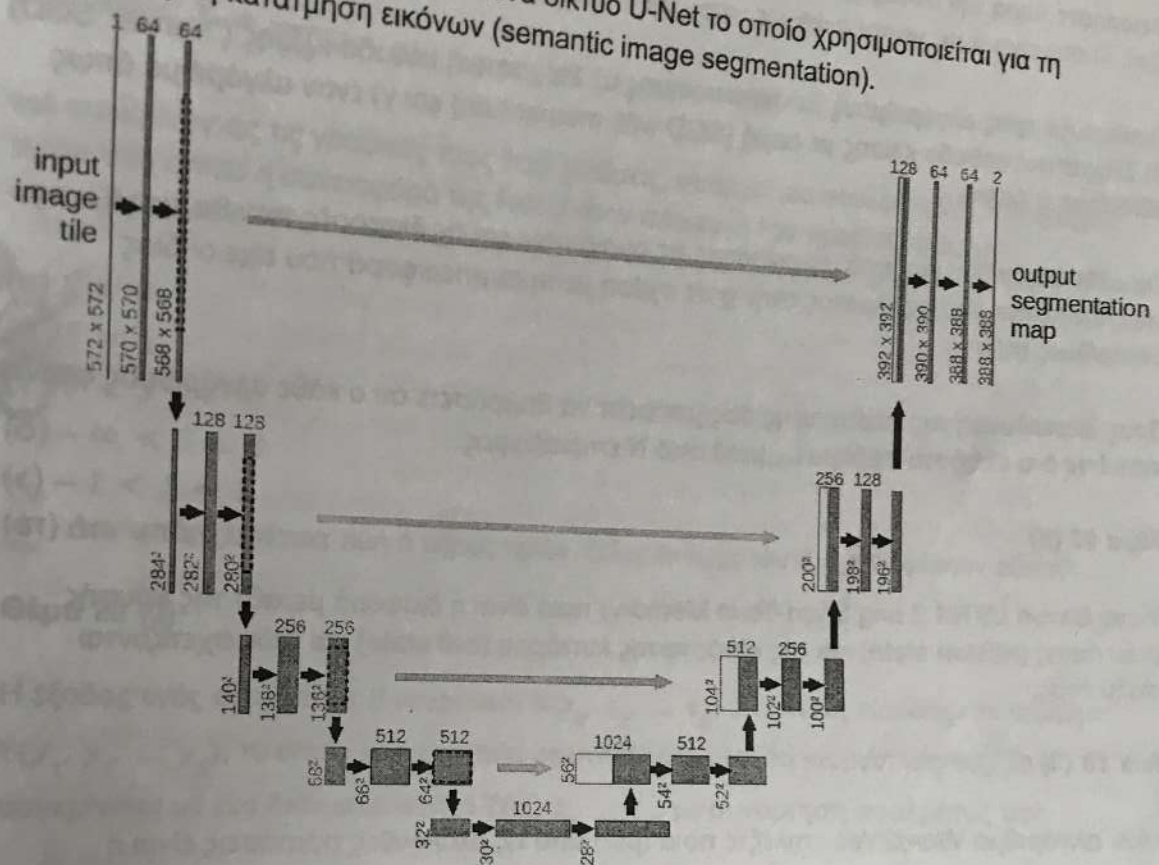
Ποιά από τις παρακάτω προτάσεις για τη (βαθιά) αυτοενισχυόμενη μάθηση είναι ορθή;

- A. Η χρήση συναρτήσεων αποτίμησης κατάστασης απαντάται τόσο στη model-based αυτοενισχυόμενη μάθηση, όσο και στη model-free
- B. Ο αλγόριθμος Q-learning προϋποθέτει ότι τα δεδομένα του είναι ακολουθιακά
- C. Ο αλγόριθμος Deep Q-learning (DQN) αντιμετωπίζει τα ζητήματα ευστάθειας του απλού Q-learning αλγορίθμου είτε μέσω επανάληψης εμπειρίας ή μέσω "παγώματος" του Q-δικτύου στόχου.
- D. Ο αλγόριθμος Deep Q-learning (DQN) προσεγγίζει τον στόχο μάθησης μέσω της τεχνικής επανάληψης τιμών

Ποια είναι η βασική σχεδιαστική διαφορά των μοντέλων Fast R-CNN και Faster R-CNN και ποιον αντίκτυπο έχει αυτή στην επίδοσή τους;

#### Θέμα 14 (6)

Στο ακόλουθο σχήμα απεικονίζεται ένα δίκτυο U-Net το οποίο χρησιμοποιείται για τη σημασιολογική κατάτμηση εικόνων (semantic image segmentation).



- Τί είδους πράξεις αναπαριστούν τα βέλη με κατεύθυνση προς τα πάνω;
- Τι σημαίνουν τα βέλη προς τα δεξιά, μικρά και μεγάλα;
- Ποιοι από όλους τους μετασχηματισμούς (βέλη προς οποιαδήποτε κατεύθυνση) έχουν βάρη που εκπαιδεύονται;

#### Θέμα 15 (3)

Ένα βαθύ νευρωνικό δίκτυο περιλαμβάνει ένα επίπεδο νέου τύπου, το "ScaleBias". Το "ScaleBias" επιτελεί την πράξη  $y = sx + b$ , όπου  $y$  είναι το διάνυσμα εξόδου,  $x$  το διάνυσμα εισόδου, και  $s$  και  $b$  είναι διανύσματα κλίμακας (scale) και κατωφλίου (bias). Οι δείκτες συμβολίζουν το  $i$ -οστό στοιχείο των διανυσμάτων, τα οποία έχουν όλα τις ίδιες διαστάσεις.

Αν σας δίνονται οι παράγωγοι του σφάλματος (loss gradient) προς την έξοδο  $y$ ,  $dL/dy$ , υπολογίστε συμβολικά τις παραγώγους του σφάλματος ως προς τα  $s$  και  $b$ . Θα είχε νόημα να βάλουμε ένα επίπεδο "ScaleBias" πριν ή/και μετά από ένα πλήρως διασυνδεδεμένο (fully connected) επίπεδο; Αιτιολογήστε.



### Θέμα 16 (9)

Ένα βαθύ νευρωνικό δίκτυο υλοποιεί την συνάρτηση  $f(X, W)$  όπου  $X$  είναι η είσοδος και  $W$  ένα διάνυσμα βαρών. Θεωρήστε ότι η συνάρτηση σφάλματος που βελτιστοποιείται για την  $f$  προκύπτει ότι είναι πλήρως συμμετρική (σφαιρική) ως προς κάποια βέλτιστα βάρη  $W_{opt}$ .

Θεωρήστε τώρα την συνάρτηση  $g(X, V) = f(X, DV)$  όπου  $D$  είναι ένας διαγώνιος πίνακας, με  $D = I$ . Τι σημαίνει αυτό γεωμετρικά ως προς την συνάρτηση σφάλματος της  $g$ ;

Διαθέτουμε τρεις αλγόριθμους βελτιστοποίησης α) Στοχαστική κάθοδο κλίσης ("Vanilla" SGD) β) Στοχαστική κάθοδο κλίσης με ορμή (SGD with momentum) και γ) έναν αλγόριθμο όπως RMSProp ή Adam.

Για κάθε αλγόριθμο χωριστά, περιγράψτε τις ομοιότητες και τις διαφορές που θα έχει η ελαχιστοποίηση του σφάλματος στην  $g$  σε σχέση με τη συμπεριφορά που είχε ο ίδιος αλγόριθμος στην  $f$ .

Προς διευκόλυνση της περιγραφής σας μπορείτε να θεωρήσετε ότι ο κάθε αλγόριθμος φτάνει στην  $f$  σε ένα ελάχιστο σφάλμα  $L_{min}$  μετά από  $N$  επαναλήψεις.

### Θέμα 17 (3)

Σε ένα δίκτυο LSTM (Long Short-Term Memory) ποια είναι η διαφορά μεταξύ της κρυφής κατάστασης (hidden state) και της κατάστασης κυττάρου (cell state) και πώς σχετίζονται μεταξύ τους;

### Θέμα 18 (3)

Για τον αλγόριθμο Word2Vec επιλέξτε ποια (μία) από τις ακόλουθες προτάσεις είναι η λανθασμένη:

α) Ο Word2Vec μοντελοποιεί την τοπική δομή μεταξύ μιας λέξης "στο κέντρο" και της γειτονιάς της στην ακολουθία εισόδου.

β) Τα μοντέλα Word2Vec μπορούν να υπολογιστούν με βαθιά νευρωνικά δίκτυα με σφάλμα εξόδου L2.

γ) Τα μοντέλα Word2Vec μπορούν να υπολογιστούν με βαθιά νευρωνικά δίκτυα με σφάλμα εξόδου διασταυρούμενης εντροπίας (cross-entropy).

δ) Ο Word2Vec υπολογίζει δύο πίνακες, έναν με τα εμφυτεύματα (embeddings) των λέξεων του κέντρου και έναν που εκφράζει τα "συμφραζόμενα" (context).

ReLU:  $Y = (0, S)$

Σιγμοειδής:  $Y = \frac{1}{1+e^{-S}}$

και συγκρίνοντας τις γραφικές τους παραστάσεις, υπάρχει μια συγκεκριμένη τιμή ή εύρος τιμών στο οποίο η συνεισφορά της ReLU στην αποφυγή του προβλήματος των αφανιζόμενων κλίσεων είναι πιο εμφανής; Αν ναι, ποια/-ο είναι αυτή/-ο;

(α)  $S = 0$

~~(β)  $S = 1$~~

(γ)  $0 < S < +\infty$

(δ)  $-\infty < S < 0$

(ε)  $-1 < S < 1$

(στ) Δεν υπάρχει τέτοια τιμή ή εύρος τιμών. Όλες οι τιμές του  $S$  συνεισφέρουν εξίσου.

### Θέμα 20 (3)

Η έξοδος ενός στρώματος 8 νευρώνων  $X(x_1, x_2, \dots, x_8)$  δίνεται ως είσοδος στο στρώμα  $Y(y_1, y_2, \dots, y_5)$ , το οποίο χρησιμοποιεί σιγμοειδή συνάρτηση ενεργοποίησης. Το  $Y$  συγκρίνεται με ένα διάνυσμα-στόχο  $T(t_1, t_2, \dots, t_5)$  με συνάρτηση σφάλματος την:

$$L = \frac{1}{2} \sum_{i=1}^5 (y_i - t_i)^2$$

Πιο πρόσφατο προς τα εμπρός πέρασμα (forward pass), έχουμε:  $[0.3, 0.8, 0.1, 0.1, 0.1, 0.4, 0.7, 0.2]$ ,  $Y = [0.9, 0.5, 0.2, 0.3, 0.1]$  και  $[1, 0, 0, 0, 0]$ .  
 Δω το βάρος της σύνδεσης μεταξύ  $x_1$  και  $y_1$ . Εάν ο ρυθμός μάθησης  $\beta = 0.2$ , νίστετε την τιμή του  $\Delta w$  στην οπισθοδιάδοση (backpropagation) με βάση το προηγούμενο forward pass.



### Θέμα 21 (6)

Ένας YouTuber που δημοσιεύει βίντεο με εκλαϊκευμένα μαθήματα βαθιάς μάθησης έχει δημιουργήσει ένα βαθύ νευρωνικό δίκτυο για ένα πρόβλημα ταξινόμησης σε 2 κλάσεις: τη θετική και την αρνητική. Τα αποτελέσματα δεν τον ικανοποιούν απόλυτα και επιθυμεί να τα «ωραιοποιήσει» ώστε να πείσει το κοινό του για τη μεγάλη επιτυχία της προσέγγισής του. Ποια ή ποιες από τις παρακάτω στρατηγικές θα τον βοηθήσουν σε αυτήν την απόπειρα «εξαπάτησης»; Σημειώστε Σ (Σωστό) ή Λ (Λάθος) δίπλα από κάθε εναλλακτική.

- (α) Να χρησιμοποιήσει σύνολο εκπαίδευσης (training set) που περιλαμβάνει 95% θετικά δείγματα και να εμφανίσει ως αποτέλεσμα την τελική ακρίβεια (precision) υπολογισμένη στο σύνολο εκπαίδευσης. Σ
- (β) Να χρησιμοποιήσει σύνολο εκπαίδευσης που περιλαμβάνει 95% θετικά δείγματα, σύνολο δοκιμής που περιλαμβάνει μόνο 25% θετικά δείγματα, και να εμφανίσει ως αποτέλεσμα την τελική ανάκληση (recall) υπολογισμένη στο σύνολο δοκιμής. Σ
- (γ) Να χρησιμοποιήσει σύνολο εκπαίδευσης που περιλαμβάνει 95% θετικά δείγματα, σύνολο δοκιμής που περιλαμβάνει 50% θετικά δείγματα, και να εμφανίσει ως αποτέλεσμα την τελική ορθότητα (accuracy) υπολογισμένη στο σύνολο δοκιμής. Λ
- (δ) Να χρησιμοποιήσει περισσότερους νευρώνες κρυμμένων επιπέδων από όσους χρειάζονται και να αναφέρει ως αποτέλεσμα μόνο την τελική ορθότητα (accuracy) στο σύνολο εκπαίδευσης και όχι την τελική ορθότητα (accuracy) στο σύνολο δοκιμής. Σ

### Θέμα 22 (6)

Χαρακτηρίστε καθεμία από τις παρακάτω προτάσεις για τα Graph Neural Networks (GNNs) ως Σωστή (Σ) ή Λανθασμένη (Λ):

- (α) Για να εκπαιδευτεί ένα GNN είναι απαραίτητο να γνωρίζουμε τις ετικέτες (labels) των κόμβων του γράφου. Λ
- (β) Στα GNNs μπορούν να ενσωματωθούν δομικά στοιχεία ή τεχνικές κλασικών βαθύων νευρωνικών δικτύων όπως το dropout και το batch normalization, όμως πάντοτε με διαφορετικό τρόπο και για διαφορετικό σκοπό. Λ
- (γ) Για να αντιμετωπιστεί το πρόβλημα του oversmoothing στα GNNs, δύο πιθανές εναλλακτικές είναι (i) η χρήση shortcuts μεταξύ μη διαδοχικών GNN στρωμάτων και (ii) η προσθήκη στρωμάτων MLP πριν ή/και μετά τα GNN στρώματα. Σ