Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Εαρινό Εξάμηνο 2023-2024

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΒΑΘΙΑ ΜΑΘΗΣΗ

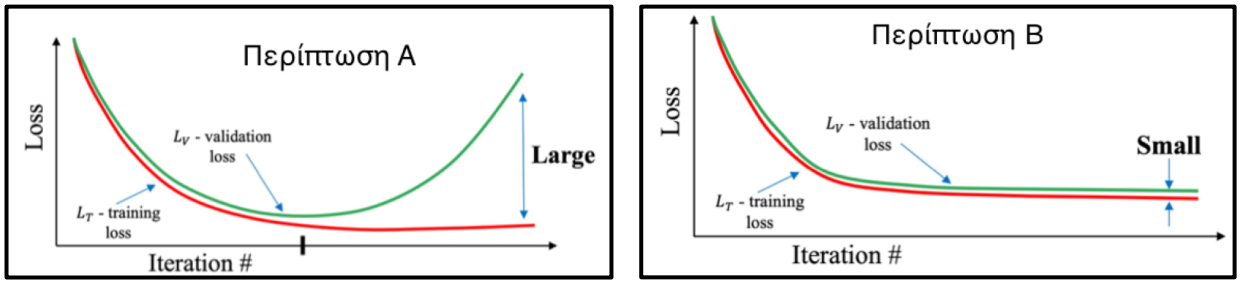
Σειρά Αναλυτικών Ασκήσεων

Ιωάννης (Χουάν) Τσαντήλας
03120883

Contents

Άσκηση 1 (Multilayer Perceptron – Regularization)

Για να εκπαιδεύσουμε ένα βαθύ νευρωνικό δίκτυο, χωρίζουμε τα δεδομένα που έχουμε στη διάθεσή μας σε τρία σύνολα: το σύνολο εκπαίδευσης (training set), το σύνολο επικύρωσης (validation set) και το σύνολο ελέγχου (testing set). Σε κάθε επανάληψη (εποχή) καταγράφουμε την τιμή της συνάρτησης κόστους (loss function) και έτσι λαμβάνουμε, για δύο διαφορετικές περιπτώσεις Α και Β, τα αντίστοιχα διαγράμματα που φαίνονται παρακάτω.



1. Τι συμπέρασμα βγάζετε από το διάγραμμα για την αρχιτεκτονική του μοντέλου σας για καθεμία από τις περιπτώσεις Α και Β;
2. Ποια τιμή επαναλήψεων (εποχών) n είναι πιο πιθανό να οδηγεί στο καλύτερο δυνατό μοντέλο μάθησης για καθεμία από τις περιπτώσεις Α και Β;
3. Για την περίπτωση Α, προτείνετε δύο τεχνικές που θα μπορούσαν να βελτιώσουν την επίδοση του μοντέλου.
4. Εξηγήστε για ποιο λόγο είναι απαραίτητο το σύνολο ελέγχου (testing set) πέρα από τα σύνολα εκπαίδευσης και επικύρωσης.

Λύση

Περιγραφή Εικόνων

- **Εικόνα Α**:
  - Η απώλεια εκπαίδευσης L_T μειώνεται σταθερά και σταθεροποιείται σε χαμηλή τιμή.
  - Η απώλεια επικύρωσης L_V αρχικά μειώνεται, αλλά στη συνέχεια αρχίζει να αυξάνεται μετά από ορισμένο αριθμό επαναλήψεων, υποδεικνύοντας υπερπροσαρμογή.
- **Εικόνα Β**:
  - Τόσο η απώλεια εκπαίδευσης L_T όσο και η απώλεια επικύρωσης L_V μειώνονται σταθερά και εξομαλύνονται σε χαμηλές τιμές, υποδεικνύοντας ένα μοντέλο καλής απόδοσης με ελάχιστη υπερπροσαρμογή.

Ερώτημα 1

- **Περίπτωση Α**: Το μοντέλο έχει υπερπροσαρμογή. Αυτό υποδεικνύεται από την απώλεια επικύρωσης που αρχίζει να αυξάνεται μετά από ένα ορισμένο σημείο, ενώ η απώλεια εκπαίδευσης συνεχίζει να μειώνεται. Το μοντέλο μαθαίνει πολύ καλά τα δεδομένα εκπαίδευσης, συμπεριλαμβανομένου του θορύβου και των ακραίων τιμών, γεγονός που βλάπτει την απόδοσή του σε αθέατα δεδομένα επικύρωσης.
- **Περίπτωση Β**: Το μοντέλο έχει καλή προσαρμογή. Τόσο οι απώλειες εκπαίδευσης όσο και οι απώλειες επικύρωσης μειώνονται και τελικά ισοπεδώνονται σε χαμηλές τιμές. Αυτό υποδεικνύει ότι το μοντέλο γενικεύει καλά σε νέα δεδομένα και δεν κάνει υπερβολική προσαρμογή.

Ερώτημα 2

- **Περίπτωση Α**: Η καλύτερη τιμή του n είναι στο σημείο όπου η απώλεια επικύρωσης βρίσκεται στο ελάχιστο πριν αρχίσει να αυξάνεται. Αυτό το σημείο σηματοδοτεί την αρχή της υπερπροσαρμογής. Εδώ μπορεί να εφαρμοστεί πρόωρη διακοπή.
- **Περίπτωση Β**: Η καλύτερη τιμή του n είναι προς το τέλος των απεικονιζόμενων επαναλήψεων, όπου τόσο οι απώλειες εκπαίδευσης όσο και οι απώλειες επικύρωσης είναι χαμηλές και έχουν εξομαλυνθεί, υποδεικνύοντας ότι το μοντέλο έχει συγκλίνει και δεν έχει υπερπροσαρμοστεί.

Ερώτημα 3

- **Αναγνώριση**: Τεχνικές όπως η κανονικοποίηση L2 (αποσύνθεση βάρους) μπορούν να βοηθήσουν στη μείωση της υπερπροσαρμογής, τιμωρώντας τα μεγάλα βάρη στο μοντέλο.
- **Αποβολή**: Η εφαρμογή στρωμάτων εγκατάλειψης κατά τη διάρκεια της εκπαίδευσης μπορεί να αποτρέψει την υπερπροσαρμογή με την τυχαία εγκατάλειψη μονάδων κατά τη διάρκεια της διαδικασίας εκπαίδευσης, καθιστώντας το μοντέλο λιγότερο ευαίσθητο στα συγκεκριμένα δεδομένα εκπαίδευσης.

Ερώτημα 4

Το σύνολο δοκιμών χρησιμοποιείται για την αξιολόγηση της απόδοσης του τελικού μοντέλου αφού αυτό έχει εκπαιδευτεί και ρυθμιστεί χρησιμοποιώντας τα σύνολα εκπαίδευσης και επικύρωσης. Παρέχει μια αμερόληπτη αξιολόγηση της απόδοσης του μοντέλου σε εντελώς αφανή δεδομένα, διασφαλίζοντας ότι οι μετρικές απόδοσης του μοντέλου αντικατοπτρίζουν την ικανότητά του να γενικεύει σε νέα δεδομένα του πραγματικού κόσμου. Το σύνολο δοκιμών βοηθά στην αξιολόγηση της πραγματικής προβλεπτικής ισχύος και της ευρωστίας του μοντέλου.

Άσκηση 2 (Representation Learning- Autoencoders)

Έστω ότι έχουμε πρόσβαση σε αυξημένες υπολογιστικές υποδομές και εκπαιδεύουμε ένα skipgram μοντέλο για ένα μεγαλύτερο λεξιλόγιο $\mathcal{V}'$. Το $\mathcal{V}'$ περιέχει αναπαραστάσεις 1500 λέξεων (μαζί με τα special tokens) και η διάσταση των διανυσμάτων u_o και u_c είναι (256×1). Στη συνέχεια χρησιμοποιούμε τα skipgram vectors και εκπαιδεύουμε έναν αυτοκωδικοποιητή (auto-encoder) με 5 κρυμμένο στρώμα διαστάσεων $[500,250,50,250,500]$ αντιστοίχως. Απαντήστε στα παρακάτω ερωτήματα:

a) Ποια είναι η διάσταση των χαρακτηριστικών εισόδου x_i στον auto-encoder;
b) Ποια είναι η διάσταση των χαρακτηριστικών εξόδου y_i του auto-encoder;
c) Ποια είναι η διάσταση της λανθάνουσας αναπαράστασης (latent representation) του auto-encoder;

Λύση

Ερώτημα (α)

- **Μέγεθος λεξιλογίου ($\mathcal{V}'$)**: 1500 λέξεις
- **Διάσταση διανύσματος**: Κάθε διάνυσμα λέξεων έχει μέγεθος 256×1.
- **Διανύσματα πλαισίου και στόχου**: Χρησιμοποιούνται τόσο τα διανύσματα u_o (έξοδος) όσο και τα διανύσματα u_c (πλαίσιο), δηλαδή 2 διανύσματα ανά λέξη.

Έτσι, για κάθε λέξη, έχουμε:

$$\text{Total Dimension per Word} = 256 \ (\text{από } u_o) + 256 \ (\text{από } u_c) = 512$$

Δεδομένου ότι το μέγεθος του λεξιλογίου είναι 1500, η συνολική διάσταση των χαρακτηριστικών εισόδου x_i είναι:

$$\text{Dimension of } x_i = 1500 \cdot 512 = 768{,}000$$

Ερώτημα (β)

Σε έναν αυτόματο κωδικοποιητή, τα χαρακτηριστικά εξόδου y_i είναι ανακατασκευασμένες είσοδοι. Επομένως, η διάσταση των χαρακτηριστικών εξόδου y_i θα είναι η ίδια με τη διάσταση των χαρακτηριστικών εισόδου x_i.

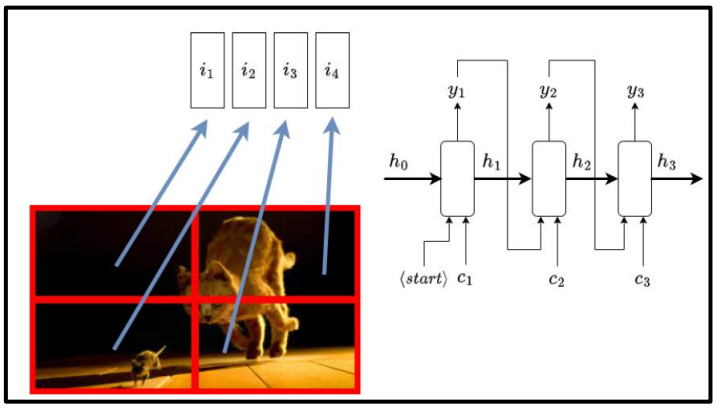
Συνεπώς, η διάσταση των χαρακτηριστικών εξόδου y_i είναι επίσης 768,000.

Ερώτημα (γ)

Η λανθάνουσα αναπαράσταση του αυτοκωδικοποιητή καθορίζεται από το μέγεθος του μικρότερου κρυμμένου στρώματος του δικτύου, το οποίο είναι συνήθως το στρώμα συμφόρησης. Δεδομένων των διαστάσεων του κρυμμένου στρώματος: $[500,250,50,250,500]$. Η μικρότερη διάσταση είναι 50. Συνεπώς, η διάσταση της λανθάνουσας αναπαράστασης είναι 50.

Άσκηση 3 (Recurrent Neural Networks)

Εξετάστε το πρόβλημα που φαίνεται στο σχήμα, όπου δίνεται μια εικόνα και το αντικείμενο είναι να δημιουργηθεί μια περιγραφική λεζάντα φυσικής γλώσσας για την εικόνα.



Την εικόνα επεξεργάζεται ένα CNN (δεν αναπαρίσταται και δε χρειάζεται για την επίλυση), με αποτέλεσμα 4 αναπαραστάσεις χαρακτηριστικών i_1, i_2, i_3, i_4, όπου κάθε $i_j \in R^2$ όπως φαίνεται στο σχήμα. Στη συνέχεια, η λεζάντα δημιουργείται αυτόματα με παλινδρόμηση από έναν αποκωδικοποιητή που βασίζεται σε RNN, που εξαρτάται από τις αναπαραστάσεις των χαρακτηριστικών της εικόνας. Το λεξιλόγιο εξόδου περιέχει μόνο 6 λέξεις, συμπεριλαμβανομένων των συμβόλων (start) και (stop) με το ακόλουθο indexing: $V = [(start), (stop), mouse, cat, staring, hunting]$ και με embedding vectors:

$$y_{<start>} = [0,0,0]^T, y_{<stop>} = [1,1,1]^T, y_{mouse} = [-1,2,0]^T, y_{cat} = [1,-2,0]^T,$$

$$y_{staring} = [0,-1,-1]^T, y_{hunting} = [0,2,1]^T$$

Ερώτημα 1

Έστω εικόνα εισόδου αυτή που φαίνεται στο σχήμα, με τους ακόλουθους χάρτες χαρακτηριστικών εικόνας:

$$i_1 = [4,0]^T, \quad i_2 = [0,4]^T, \quad i_3 = [0,0]^T, \quad i_4 = [0,0]^T$$

Σε κάθε χρονικό βήμα t, ο decoder που βασίζεται σε RNN λαμβάνει ως είσοδο $\boldsymbol{x_t} \in \boldsymbol{R^5}$, η οποία είναι **concatenation** της προηγούμενης εξόδου που ενσωματώνει $\boldsymbol{y_{t-1}} \in \boldsymbol{R^3}$ και μια αναπαράσταση εικόνας $\boldsymbol{c_t} \in \boldsymbol{R^2}$ (με αυτή τη σειρά) και χρησιμοποιεί αυτήν την είσοδο και την προηγούμενη κρυφή κατάσταση $\boldsymbol{h_{t-1}}$ για να υπολογίσει τη νέα κατάσταση $\mathbf{h_t}$. Αυτό ακολουθείται από ένα γραμμικό επίπεδο εξόδου (linear output layer) με πίνακα hidden to output:

$$W_{gh} = \begin{bmatrix} -5 & -5 \\ 0 & 3 \\ 1 & 2 \\ 2 & 2 \\ 3 & -1 \\ 2.9 & 0 \end{bmatrix} \begin{matrix} \#<start> \\ \#<stop> \\ \#mouse \\ \#cat \\ \#staring \\ \#hunting \end{matrix}$$

Όπου κάθε σειρά αυτού του πίνακα αντιστοιχεί στις λέξεις που αναφέρονται παραπάνω. Ο πίνακας input to hidden και ο πίνακας recurrence δίνονται αντίστοιχα από:

$$W_{hx} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}, \qquad W_{hh} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Ας υποθέσουμε ότι το RNN χρησιμοποιεί ενεργοποιήσεις ReLU, ότι όλα τα bias vectors είναι μηδενικά διανύσματα και ότι η αρχική κρυφή κατάσταση (hidden state) $\mathbf{h_0}$ είναι ένα μηδενικό διάνυσμα.

Σε αυτήν την ερώτηση, υποθέστε ότι το $\boldsymbol{c_t} \coloneqq \boldsymbol{c}$ είναι **σταθερό για όλα τα χρονικά βήματα** και προκύπτει από τη μέση συγκέντρωση των αναπαραστάσεων των χαρακτηριστικών εικόνας, $c = \frac{1}{4}\sum_{j=1}^{4} i_j$, χωρίς μηχανισμό προσοχής. Υποθέστε τις 3 πρώτες λέξεις της λεζάντας χρησιμοποιώντας **greedy coding** (επιλογή της πιο πιθανής λέξης σε κάθε βήμα).

Λύση

Υπολογισμός της μέσης αναπαράστασης χαρακτηριστικών της εικόνας c

Για να βρούμε το c , παίρνουμε το μέσο όρο των τεσσάρων αναπαραστάσεων χαρακτηριστικών:

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

Υπολογισμός της κρυφής κατάστασης και της εξόδου σε κάθε χρονικό βήμα

- *Χρονικό βήμα t=1*

Είσοδος

$$x_1 = [0,0,0,1,1]^T$$

Hidden state

$$h_1 = ReLU(W_{hx} \cdot x_1 + W_{hh} \cdot h_0 + bias_{in}) = ReLU\left(\begin{bmatrix}1 & 0 & 1 & 1 & 0\\ 0 & 1 & 1 & 0 & 1\end{bmatrix}\begin{bmatrix}0\\0\\0\\1\\1\end{bmatrix} + \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \cdot 0 + 0\right) =$$

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

Έξοδος

$$y_1 = \left(W_{gh} \cdot h_1 + bias_{out}\right) = \begin{bmatrix}-5 & -5\\ 0 & 3\\ 1 & 2\\ 2 & 2\\ 3 & -1\\ 2.9 & 0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}-10\\3\\3\\4\\2\\2,9\end{bmatrix}$$

Η λέξη με τη μεγαλύτερη πιθανότητα είναι η **cat**.

- *Χρονικό βήμα t=2*

Είσοδος

$$x_2 = [1,-2,0,1,1]^T$$

Hidden state

$$h_2 = ReLU(W_{hx} \cdot x_2 + W_{hh} \cdot h_1 + bias_{in}) = ReLU\left(\begin{bmatrix}1 & 0 & 1 & 1 & 0\\ 0 & 1 & 1 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\ -2\\ 0\\ 1\\ 1\end{bmatrix} + \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \cdot \begin{bmatrix}1\\ 1\end{bmatrix} + 0\right) =$$

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

Έξοδος

$$y_2 = \left(W_{gh} \cdot h_2 + bias_{out}\right) = \begin{bmatrix}-5 & -5\\ 0 & 3\\ 1 & 2\\ 2 & 2\\ 3 & -1\\ 2.9 & 0\end{bmatrix}\begin{bmatrix}3\\ 0\end{bmatrix} = \begin{bmatrix}-15\\ 0\\ 3\\ 6\\ 9\\ 8{,}7\end{bmatrix}$$

Η λέξη με τη μεγαλύτερη πιθανότητα είναι η **staring**.

- *Χρονικό βήμα t=2*

Είσοδος

$$x_3 = [0, -1, -1, 1, 1]^T$$

Hidden state

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

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

Έξοδος

$$y_3 = \left(W_{gh} \cdot h_3 + bias_{out}\right) = \begin{bmatrix}-5 & -5\\ 0 & 3\\ 1 & 2\\ 2 & 2\\ 3 & -1\\ 2.9 & 0\end{bmatrix}\begin{bmatrix}0\\ 2\end{bmatrix} = \begin{bmatrix}-10\\ 6\\ 4\\ 4\\ -2\\ 0\end{bmatrix}$$

Η λέξη με τη μεγαλύτερη πιθανότητα είναι η **(stop)**.

Ερώτημα 2

Ας υποθέσουμε τώρα ότι, αντί να χρησιμοποιεί ένα σταθερό c_t για όλα τα χρονικά βήματα, ο αποκωδικοποιητής που βασίζεται σε RNN έχει έναν μηχανισμό προσοχής scaled dot-product που παρακολουθεί τις αναπαραστάσεις των χαρακτηριστικών της εικόνας. Για κάθε χρονικό βήμα, το διάνυσμα ερωτήματος είναι h_{t-1} και οι αναπαραστάσεις των χαρακτηριστικών εικόνας i_1, i_2, i_3, i_4, χρησιμοποιούνται τόσο ως κλειδιά (**key**) όσο και ως τιμές (**value**).

Υποθέστε ξανά ότι το h_0 είναι ένα μηδενικό διάνυσμα. Για το 1ο χρονικό βήμα (t=1), υπολογίστε τις πιθανότητες προσοχής (attention probabilities) και το διάνυσμα εικόνας που προκύπτει c_1. Η 1η λέξη θα είναι ίδια ή διαφορετική από αυτήν στην προηγούμενη ερώτηση;

Λύση

Attention Scores

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

Scaled Scores

$$Scaled\ Score = \frac{Score}{\sqrt{d_k}} = \frac{[0 \quad 0 \quad 0 \quad 0]}{\sqrt{2}} = [0 \quad 0 \quad 0 \quad 0]$$

Attention Weights

$$P = softmax(ScaledScore) = softmax([0 \quad 0 \quad 0 \quad 0]) =$$

$$= [e^0/4 \quad e^0/4 \quad e^0/4 \quad e^0/4] = [0.25 \quad 0.25 \quad 0.25 \quad 0.25]$$

Τελικά:

$$z = P \cdot V = [0.25 \quad 0.25 \quad 0.25 \quad 0.25] \cdot \begin{bmatrix}4 & 0\\0 & 4\\0 & 0\\0 & 0\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$$

Καταλήξαμε στο ίδιο ακριβώς διάνυσμα. Επομένως, η λέξη παραμένει **cat**.

Άσκηση 4 (Convolutional Neural Networks)

Ας υποθέσουμε ότι έχουμε ένα δίκτυο σαν το AlexNet (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", NeurIPS 2012). Θεωρήστε πως έχουμε εικόνες διαστάσεων 227x227x3 (έγχρωμες με RGB channels) και φίλτρο 11x11x3 στο πρώτο convolutional layer. Το δίκτυο έχει συνολικά 96 φίλτρα, stride ίσο με 4 και μηδενικό padding.

a) Υπολογίστε τις διαστάσεις στην έξοδο του πρώτου convolutional layer.
b) Υπολογίστε τον αριθμό των layer στο πρώτο convolutional layer.
c) Υπολογίστε τον αριθμό των εκπαιδεύσιμων παραμέτρων του πρώτου convolutional layer με διαμοιρασμό βαρών.
d) Αν αντικαθιστούσαμε το CNN με ένα FeedForward layer με 256 units πόσες εκπαιδεύσιμες παραμέτρους θα είχαμε;

Λύση

- Διαστάσεις εικόνας εισόδου: $227 \times 227 \times 3$
- Μέγεθος φίλτρου: $F = 11 \times 11 \times 3$
- Αριθμός φίλτρων: $K = 96$
- Stride: $S = 4$
- Padding: $P = 0$
- Πλάτος, ύψος, βάθος εισόδου: $w_1 = 227, h_1 = 227, d_1 = 3$
- Πλάτος, ύψος, βάθος εξόδου: w_2, h_2, d_2

Ερώτημα (α)

Ο τύπος για τον υπολογισμό του πλάτους και του ύψους εξόδου για ένα συνελικτικό στρώμα είναι:

$$w_2 = \left(\frac{w_1 - F + 2 \cdot P}{S}\right) + 1 = \left(\frac{227 - 11 + 2 \cdot 0}{S}\right) + 1 = 55$$

$$h_2 = \left(\frac{h_1 - F + 2 \cdot P}{S}\right) + 1 = 55$$

$$d_2 = K = 96$$

Επομένως, οι διαστάσεις εξόδου του πρώτου επιπέδου συνελίξεων είναι $55 \times 55 \times 96$.

Ερώτημα (β)

Για να βρούμε τον αριθμό των μονάδων (ή ενεργοποιήσεων) στον όγκο εξόδου, πολλαπλασιάζουμε τις διαστάσεις μαζί:

$$Units = w_2 \cdot h_2 \cdot d_2 = 55 \cdot 55 \cdot 96 = 290.400$$

Ερώτημα (γ)

Ο αριθμός των εκπαιδεύσιμων παραμέτρων σε ένα συνελικτικό στρώμα υπολογίζεται ως εξής (ο όρος *+1* αντιπροσωπεύει τον όρο μεροληψίας σε κάθε φίλτρο):

$$Parameters = (F \cdot F \cdot d_1 + 1) \cdot K$$

Αντικαθιστώντας τις τιμές:

$$Parameters = (F \cdot F \cdot d_1 + 1) \cdot K = (11 \cdot 11 \cdot 3 + 1) \cdot 96 = 34{,}944$$

Ερώτημα (δ)

Σε ένα πλήρως συνδεδεμένο επίπεδο, κάθε μονάδα εισόδου συνδέεται με κάθε μονάδα εξόδου. Ο αριθμός των εκπαιδεύσιμων παραμέτρων υπολογίζεται ως εξής (ο όρος *+1* αντιπροσωπεύει τον όρο μεροληψίας):

$$\text{Parameters} = (\text{Input Units} + 1) \cdot \text{Output Units}$$

Οι μονάδες εισόδου σε αυτή την περίπτωση είναι όλα τα εικονοστοιχεία της εικόνας εισόδου:

$$\text{Input units} = 227 \cdot 227 \cdot 3 = 154{,}588$$

Οι μονάδες εξόδου δίνονται 256. Επομένως, ο υπολογισμός είναι ο εξής:

$$\text{Parameters} = 154{,}588 \cdot 256 = 39{,}574{,}528$$

Άσκηση 5 (Generative models)

Έχοντας μελετήσει τα βασικά χαρακτηριστικά των ακόλουθων τριών τύπων παραγωγικών μοντέλων: Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Diffusion Models να προβείτε σε μια αναλυτική σύγκριση των τριών τύπων μοντέλων. Μπορείτε να εξετάσετε θέματα όπως οι θεμελιώδεις αρχές κάθε μοντέλου, η διαδικασία εκπαίδευσης, η αποτελεσματικότητα, οι απαιτήσεις σε μνήμη/χρόνο/δεδομένα εκπαίδευσης, κ.α., και να αναφερθείτε σε ομοιότητες και διαφορές, πλεονεκτήματα και μειονεκτήματα του κάθε τύπου μοντέλου.

Λύση

Πηγές στις οποίες βασίστηκε η απάντηση μου:

- **VAEs**: https://arxiv.org/abs/1312.6114
- **GANs**: https://arxiv.org/abs/1406.2661
- **Diffusion**: https://arxiv.org/abs/2006.11239

Variational Autoencoders (VAEs)

Βασικές αρχές

- **Αναπαράσταση λανθάνουσας περιοχής**: Οι VAE κωδικοποιούν τα δεδομένα εισόδου σε έναν συνεχή λανθάνων χώρο, ο οποίος επιτρέπει την ομαλή παρεμβολή μεταξύ των σημείων δεδομένων.
- **Μεταβλητή συμπερασματολογία**: Αντί να μαθαίνουν μια ντετερμινιστική συνάρτηση, οι VAE μαθαίνουν μια κατανομή πιθανοτήτων, χρησιμοποιώντας ένα τέχνασμα επαναπαραμετροποίησης για να χειριστούν τη στοχαστική φύση αυτής της κατανομής.
- **Ανακατασκευή και κανονικοποίηση**: Οι VAEs στοχεύουν στην ανακατασκευή των δεδομένων εισόδου, ενώ τακτοποιούν τον λανθάνων χώρο ώστε να ακολουθεί μια γνωστή εκ των προτέρων κατανομή (συνήθως Γκαουσιανή).

Διαδικασία εκπαίδευσης

- **Συνάρτηση απώλειας**: Η συνάρτηση απώλειας στις VAE περιλαμβάνει δύο μέρη: την απώλεια ανακατασκευής (π.χ. μέσο τετραγωνικό σφάλμα) και τον όρο απόκλισης KL που μετρά πόσο στενά ο λανθάνων χώρος ακολουθεί την προηγούμενη κατανομή.
- **Βελτιστοποίηση**: Το μοντέλο εκπαιδεύεται χρησιμοποιώντας τεχνικές βελτιστοποίησης βασισμένες στην κλίση.

Αποτελεσματικότητα και απαιτήσεις

- **Μνήμη/χρόνος**: Τα VAE είναι σχετικά αποδοτικά στην εκπαίδευση λόγω της απλής αρχιτεκτονικής τους. Ωστόσο, ο συμβιβασμός μεταξύ ανακατασκευής και κανονικοποίησης μπορεί μερικές φορές να κάνει τη σύγκλιση αργή.
- **Απαιτήσεις δεδομένων**: Οι VAE απαιτούν σημαντικό όγκο δεδομένων για να μάθουν ουσιαστικές αναπαραστάσεις, αλλά είναι γενικά πιο αποδοτικοί σε δεδομένα από τους GAN.

Πλεονεκτήματα

- **Δυνατότητες δημιουργίας**: Οι VAE μπορούν να παράγουν ομαλές παρεμβολές μεταξύ σημείων δεδομένων.

- **Σταθερότητα**: Η εκπαίδευση είναι γενικά σταθερή και λιγότερο επιρρεπής σε προβλήματα όπως η κατάρρευση λειτουργίας που παρατηρείται στα GAN.

Μειονεκτήματα

- **Θολότητα**: Οι παραγόμενες εικόνες μπορεί να είναι θολές λόγω του στόχου ανακατασκευής που εστιάζει στις μέσες ιδιότητες.
- **Πολύπλοκη επιλογή προτεραιοτήτων**: Η επιλογή της σωστής προηγούμενης κατανομής μπορεί να είναι μη τετριμμένη.

Generative Adversarial Networks (GANs)

Βασικές αρχές

- **Αντιθετική εκπαίδευση**: Τα GAN αποτελούνται από δύο νευρωνικά δίκτυα, μια γεννήτρια και έναν διαχωριστή, τα οποία εκπαιδεύονται ταυτόχρονα σε ένα παιχνίδι minimax. Η γεννήτρια προσπαθεί να δημιουργήσει ρεαλιστικά δεδομένα, ενώ ο διαχωριστής προσπαθεί να διακρίνει μεταξύ πραγματικών και παραγόμενων δεδομένων.
- **Ισορροπία Nash**: Η εκπαίδευση στοχεύει να φτάσει σε ένα σημείο όπου η γεννήτρια παράγει δεδομένα που ο διαχωριστής δεν μπορεί να διακρίνει από τα πραγματικά δεδομένα.

Διαδικασία εκπαίδευσης

- **Συνάρτηση απώλειας**: Οι GAN χρησιμοποιούν συνήθως αντιθετικές συναρτήσεις απώλειας, όπου η γεννήτρια ελαχιστοποιεί την πιθανότητα ο διαχωριστής να αναγνωρίσει σωστά τα ψεύτικα δεδομένα και ο διαχωριστής μεγιστοποιεί αυτή την πιθανότητα.
- **Βελτιστοποίηση**: Η εκπαίδευση περιλαμβάνει εναλλασσόμενες ενημερώσεις της γεννήτριας και του διαχωριστή, οι οποίες μπορεί να είναι ασταθείς και απαιτούν προσεκτικό συντονισμό.

Αποτελεσματικότητα και απαιτήσεις

- **Μνήμη/χρόνος**: Η εκπαίδευση των GAN είναι υπολογιστικά εντατική και απαιτεί σημαντική μνήμη και χρόνο, ιδίως για εικόνες υψηλής ανάλυσης.
- **Απαιτήσεις δεδομένων**: Οι GAN απαιτούν μεγάλα σύνολα δεδομένων για να μάθουν να παράγουν ρεαλιστικά δεδομένα και η απόδοσή τους είναι ιδιαίτερα ευαίσθητη στην ποιότητα και την ποσότητα των δεδομένων.

Πλεονεκτήματα

- **Δείγματα υψηλής ποιότητας**: Τα GAN μπορούν να παράγουν εξαιρετικά ρεαλιστικές και υψηλής ποιότητας εικόνες, ξεπερνώντας συχνά άλλα παραγωγικά μοντέλα.
- **Ευελιξία**: Τα GAN μπορούν να προσαρμοστούν για διάφορες εφαρμογές, συμπεριλαμβανομένης της δημιουργίας εικόνων, της μεταφοράς στυλ και της υπερ-ανάλυσης.

Μειονεκτήματα

- **Αστάθεια εκπαίδευσης**: Τα GAN είναι γνωστό ότι είναι δύσκολο να εκπαιδευτούν λόγω προβλημάτων όπως η κατάρρευση τρόπων λειτουργίας, η μη σύγκλιση και η ευαισθησία στις υπερπαραμέτρους.
- **Εντατική σε πόρους**: Οι υψηλές απαιτήσεις υπολογισμού και μνήμης τις καθιστούν λιγότερο προσιτές για εφαρμογές μικρότερης κλίμακας.

Diffusion Models

Βασικές αρχές

- **Επαναληπτική βελτίωση**: Τα μοντέλα διάχυσης παράγουν δεδομένα μέσω μιας διαδικασίας επαναληπτικής προσθήκης και στη συνέχεια αφαίρεσης θορύβου από μια αρχική κατανομή θορύβου.
- **Στοχαστική διαδικασία**: Η διαδικασία δημιουργίας μοντελοποιείται ως στοχαστική διαφορική εξίσωση, καθιστώντας την ανάλογο της αλυσίδας Markov σε συνεχή χρόνο.

Διαδικασία εκπαίδευσης

- **Συνάρτηση απώλειας**: Η εκπαίδευση συνήθως περιλαμβάνει την ελαχιστοποίηση της διαφοράς μεταξύ των θορυβωδών δεδομένων και των αποθορυβοποιημένων δεδομένων σε κάθε βήμα.
- **Βελτιστοποίηση**: Παρόμοια με τις VAE και τις GAN, χρησιμοποιούνται μέθοδοι βελτιστοποίησης με βάση τη διαβάθμιση.

Αποτελεσματικότητα και απαιτήσεις

- **Μνήμη/χρόνος**: Τα μοντέλα διάχυσης μπορεί να είναι αργά στην εκπαίδευση και τη δημιουργία δεδομένων, καθώς απαιτούν πολλά επαναληπτικά βήματα για την παραγωγή δειγμάτων υψηλής ποιότητας.
- **Απαιτήσεις δεδομένων**: Όπως τα GAN, έτσι και τα μοντέλα διάχυσης χρειάζονται μεγάλα σύνολα δεδομένων για να μάθουν αποτελεσματικά βήματα αποθορυβοποίησης.

Πλεονεκτήματα

- **Ποιότητα δείγματος**: Τα μοντέλα διάχυσης μπορούν να παράγουν δείγματα πολύ υψηλής ποιότητας, ξεπερνώντας ενδεχομένως ορισμένες από τις αδυναμίες των VAE.
- **Σταθερότητα εκπαίδευσης**: Γενικά πιο σταθερά στην εκπαίδευσή τους σε σύγκριση με τα GAN, καθώς δεν υποφέρουν από προβλήματα αντιφατικής εκπαίδευσης.

Μειονεκτήματα

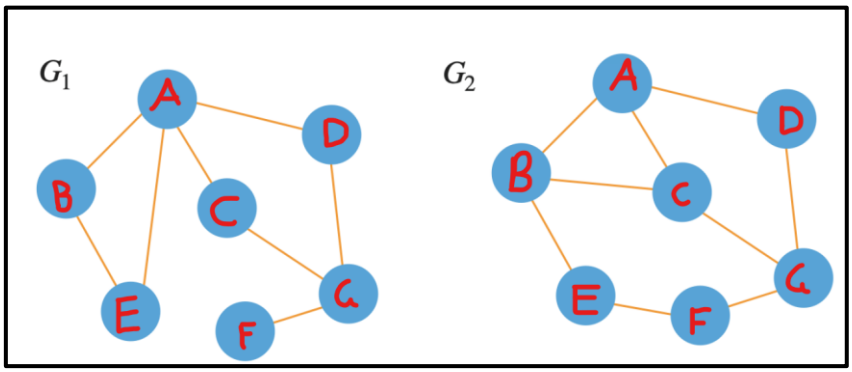
- **Ταχύτητα δημιουργίας**: Η επαναληπτική φύση καθιστά τη διαδικασία δειγματοληψίας πιο αργή σε σύγκριση με τα GAN και τα VAE.
- **Πολύπλοκοτητα**: Το μαθηματικό πλαίσιο και η υλοποίηση μπορεί να είναι πιο πολύπλοκα, περιλαμβάνοντας περίπλοκες διαδικασίες προγραμματισμού θορύβου και διάχυσης.

Σύνοψη

Feature	**VAEs**	**GANs**	**Diffusion Models**
Principle	Variational Inference	Adversarial Training	Iterative Refinement
Loss Function	Reconstruction + KL-Divergence	Adversarial Loss	Noise Matching
Training Stability	Stable	Unstable	Stable
Sample Quality	Moderate, often blurry	High, realistic	High, realistic
Efficiency	Moderate	Computationally Intensive	Slow sampling
Data Requirements	Moderate	High	High
Advantages	Smooth Latent Space, Stability	High-Quality Samples, Flexibility	High-Quality Samples, Stability
Disadvantages	Blurry Samples, Prior Selection	Training Instability, Resource Intensive	Slow Sampling, Complexity

Άσκηση 6 (Graph Neural Networks)

Δίνονται οι γράφοι G1 και G2 του παρακάτω σχήματος. Να υπολογιστεί η ομοιότητά τους με βάση τον πυρήνα Weisfeiler-Lehman (WL), δίνοντας όλα τα βήματα του αλγόριθμου.



Λύση

Graph G_1						
Node	**Neighbors**	**i=0**	**i=1**	**Hash 1**	**i=2**	**Hash 2**
A	B,C,D,E	1	1; 1,1,1,1	5	5; 3,3,3,3	15
B	A,E	1	1; 1,1	3	3; 4,5	10
C	A,G	1	1; 1,1	3	3; 3,5	8
D	A,G	1	1; 1,1	3	3; 4,5	10
E	A,B	1	1; 1,1	3	3; 3,5	8
F	G	1	1; 1	2	2; 4	6
G	C,D,F	1	1; 1,1,1	4	4; 2,3,3	11
Graph G_2						
Node	**Neighbors**	**i=0**	**i=1**	**Hash 1**	**i=2**	**Hash 2**
A	B,C,D	1	1; 1,1,1	4	4; 3,4,4	12
B	A,C,E	1	1; 1,1	3	3; 4,4	9
C	A,B,G	1	1; 1,1,1	4	4; 3,4,4	12
D	A,G	1	1; 1,1,1	4	4; 4,4,4	14
E	B,F	1	1; 1,1	3	3; 3,4	7
F	E,G	1	1; 1,1	3	3; 3,4	7
G	C,D,F	1	1; 1,1,1	4	4; 3,3,4	13

Στο Hash 1, κοιτώ τις unique ετικέτες του Iteration 1 και τις ονοματίζω. Με βάση αυτά προκύπτει το Iteration 2. Οι unique ετικέτες που έχουμε είναι οι:

- (1;-): 1
- (1; 1): 2
- (1; 1,1): 3
- (1; 1,1,1): 4
- (1; 1,1,1,1): 5

Στο Hash 2, κοιτώ τις unique ετικέτες του Iteration 2 και τις ονοματίζω. Με βάση αυτά προκύπτει το Iteration 3. Οι unique ετικέτες που έχουμε είναι οι:

- (2; 4): 6
- (3; 3,4): 7
- (3; 3,5): 8
- (3; 4,4): 9

- (3; 4,5): 10
- (4; 2,3,3): 11
- (4; 3,4,4): 12
- (4; 3,3,4): 13
- (4; 4,4,4): 14
- (5; 3,3,3,3): 15

Παρατηρώ πως οι γράφοι δεν έχουν καμία hash ετικέτα κοινή μεταξύ τους, επομένως σταματώ τα iterations.

Κρατώ στον παρακάτω πίνακα τις unique ταμπέλες και τις φορές που εμφανίστηκαν σε κάθε γράφο αντίστοιχα. Στην τελευταία γραμμή έχω το γινόμενο της αντίστοιχης στήλης (παρατηρούμε κιόλας πως για τις ετικέτες 6-15 ότι τα γινόμενα είναι 0, επιβεβαιώνοντας πως δεν χρειάζεται άλλη επανάληψη):

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

Το άθροισμα των γινομένων είναι 65, δηλαδή:

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