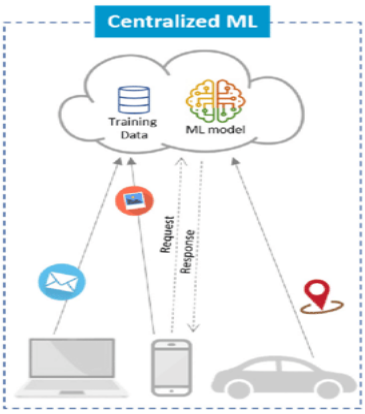
Πανεπιστήμιο Πατρών
Τμήμα Μηχ. Η/Υ & Πληροφορικής

Στατιστική Επεξεργασία Σήματος και Μάθηση

Δεύτερη Εργαστηριακή Άσκηση
Ακαδημαϊκό Έτος 2024/25

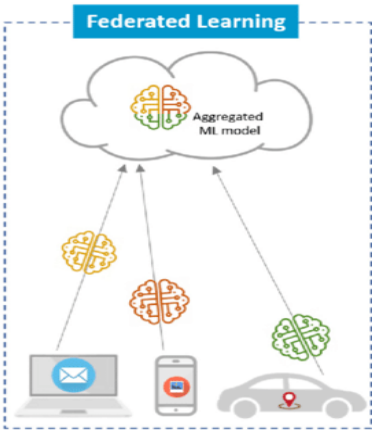
Ομοσπονδιακή Μάθηση (Federated Learning)

Στην παρούσα εργασία θα ασχοληθούμε με την ομοσπονδιακή μάθηση (που είναι τεχνική κατανεμημένης μάθησης).



Εικόνα 1:Κεντρικοποιημένη Μηχανική Μάθηση

Κατ' αρχάς πρέπει να σημειωθεί ότι η κεντρικοποιημένη μηχανική μάθηση [Εικόνα 1] αντιμετωπίζει σημαντικές προκλήσεις όταν εφαρμόζεται σε πραγματικές συνθήκες. Ένα κρίσιμο ζήτημα είναι η ασφάλεια των δεδομένων. Η συγκέντρωση μεγάλου όγκου δεδομένων σε ένα κεντρικό σημείο, όπως συμβαίνει συχνά στις εφαρμογές μηχανικής μάθησης, τα καθιστά ευάλωτα σε κυβερνοεπιθέσεις. Τομείς όπως οι υπηρεσίες υγείας και οικονομίας απαιτούν ιδιαίτερα υψηλά επίπεδα ασφαλείας και προστασίας της ιδιωτικότητας. Επιπλέον, η εκπαίδευση πολύπλοκων μοντέλων μηχανικής μάθησης απαιτεί τεράστια υπολογιστική ισχύ και ενέργεια, γεγονός που μπορεί να περιορίσει την εφαρμογή της σε μεγάλη κλίμακα.



Εικόνα 2: Ομοσπονδιακή Μάθηση

Τέλος, πολύπλοκες εφαρμογές μηχανικής μάθησης, όπως η αυτόνομη οδήγηση ή η αυτόνομη επικοινωνία συσκευών στο πεδίο στις παρυφές του δικτύου (Edge Computing) , απαιτούν λήψη αποφάσεων σε πραγματικό χρόνο. Η κεντρική επεξεργασία των δεδομένων μπορεί να επιβραδύνει σημαντικά αυτόν τον χρόνο απόκρισης, θέτοντας σε κίνδυνο την αποτελεσματικότητα και την ασφάλεια τέτοιων συστημάτων.

Μία εναλλακτική προσέγγιση σε αυτά τα προβλήματα προσφέρει η κατανεμημένη μηχανική μάθηση, όπως η λεγόμενη Ομοσπονδιακή Μάθηση (Federated Learning) που φαίνεται σχηματικά στην [Εικόνα 2]. Αντί να συγκεντρώνονται όλα τα δεδομένα σε ένα κεντρικό σημείο, η κατανεμημένη μάθηση επιτρέπει την εκπαίδευση μοντέλων μηχανικής μάθησης σε πολλαπλούς υπολογιστικούς κόμβους που βρίσκονται σε διαφορετικά γεωγραφικά σημεία. Με αυτόν τον τρόπο, προστατεύεται η ιδιωτικότητα των δεδομένων, αποφεύγοντας την ανταλλαγή των δεδομένων αυτών καθ'αυτών και μειώνεται ο κίνδυνος παραβίασης της ασφάλειας, μιας και αν δεχθεί επίθεση ένας κόμβος, όλο το υπόλοιπο δίκτυο παραμένει ασφαλές. Επιπλέον, η εκπαίδευση πολύπλοκων μοντέλων μηχανικής μάθησης μπορεί να διευκολυνθεί μιας και πλέον η διαδικασία μάθησης κατανέμεται σε πολλούς κόμβους έχοντας ως απότοκο την επεξεργασία μεγαλύτερων όγκων δεδομένων, με πιο φθηνή υπολογιστική ισχύ χωρίς να απαιτείται ένας υπερσύγχρονος υπολογιστικός εξοπλισμός σε ένα μόνο σημείο.

Στην συνέχεια, θα ορίσουμε το πρόβλημα της κεντρικοποιημένης μηχανικής μάθησης (centralized machine learning) και έπειτα θα δούμε πως διαμορφώνεται το πρόβλημα στην κατανεμημένη εκδοχή του.

Στατιστική Επεξεργασία Σήματος και Μάθηση

Δεύτερη Εργαστηριακή Άσκηση
Ακαδημαϊκό Έτος 2024/25

Θεωρούμε το πρόβλημα ταξινόμησης πολλαπλών κλάσεων, με διανύσματα εισόδου $x \in \mathcal{X} \subseteq \mathrm{R}^d$ *με αντίστοιχα διανύσματα-ετικέτες* $y \in \mathcal{Y} = \{1,2,\cdots,C\}$*, όπου* C *είναι το σύνολο των κλάσεων.* Αρχικά, το πρόβλημα βελτιστοποίησης που θέλουμε να λύσουμε κατά την κεντρικοποιημένη μηχανική μάθηση ορίζεται ως εξής,

$$\min_{\theta} F(\theta) \coloneqq E_{(x,y)\sim D}\left[l(\phi(x;\theta), y)\right],$$

όπου $(x, y) \sim D$ είναι τα διαθέσιμα δεδομένα τα οποία ακολουθούν μία κατανομή D, με $l(\phi(x;\theta), y)$ συμβολίζουμε τη συνάρτηση κόστους που μετρά την απόσταση που έχει η πρόβλεψη της παραμετρικής συνάρτησης $\phi(\cdot;\theta)$ (όπου με $\theta \in \Theta$ συμβολίζουμε τις παραμέτρους) για την είσοδο x, σε σχέση με την πραγματική τιμή y. Ιδανικά θέλουμε να γνωρίζουμε τις βέλτιστες παραμέτρους θ^* για τις οποίες η συνάρτηση $\phi(x;\theta^*)$ αντιστοιχεί την είσοδο x στην επιθυμητή έξοδο y. Όταν συμβαίνει αυτό η τιμή της συνάρτησης κόστους είναι ελάχιστη. Με $E[\cdot]: Z \rightarrow R$, συμβολίζουμε την αναμενόμενη τιμή, με Z συμβολίζουμε την τυχαία μεταβλητή του προβλήματός μας. Ο τελεστής $E[\cdot]$ εφαρμόζεται διότι τα δεδομένα μας είναι τυχαία (δηλαδή το Z αντικαθίσταται με τα τυχαία δείγματα (x, y)).

Για να μπορέσουμε να βρούμε αναλυτικά το Expectation $E[\cdot]$ χρειάζεται να γνωρίζουμε την συνάρτηση πυκνότητας πιθανότητας της κατανομής D, ωστόσο αυτό δεν καθίσταται δυνατό. Κάνουμε επομένως, μία εμπειρική εκτίμηση της μέσης τιμής χρησιμοποιώντας τα πειραματικά δεδομένα, θεωρώντας πως έχουμε αρκούντως πολλά δεδομένα (θεωρητικώς άπειρα δεδομένα) για να προσεγγίσουμε την πραγματική μέση τιμή.

Έτσι λοιπόν το πρόβλημα πλέον παίρνει την εμπειρική του μορφή,

$$\min_{\theta} \sum_{i=1}^{N} \frac{1}{N}\left[l(\phi(x_i;\theta), y_i)\right].$$

Δηλαδή το Expectation $E[\cdot]$ έχει αντικατασταθεί από εμπειρική μέση τιμή, η οποία για λόγους απλότητας θεωρούμε πως δίνεται με βάρος $\frac{1}{N}$ για κάθε δείγμα, όπου N είναι το πλήθος των δειγμάτων.

Το παραπάνω πρόβλημα μπορεί να λυθεί με κάποιο επαναληπτικό αλγόριθμο πρώτης τάξης, λόγου χάρη με την χρήση του Gradient Descent (GD), ή κάποιας παραλλαγής του. Οι παράμετροι θ εξελίσσονται ως εξής,

$$\theta^{t+1} \leftarrow \theta^t - \eta \nabla_\theta F(\theta^t),$$

όπου $t \in [0, T]$ είναι το πλήθος των βημάτων που θα κάνει ο αλγόριθμος μέχρι να συγκλίνει, δηλαδή $\theta^T \rightarrow \theta^*$.

Κατά την κατανεμημένη μηχανική μάθηση το πρόβλημα διαφοροποιείται ως εξής:

$$\min_{\theta} \frac{1}{K}\sum_{i=1}^{K} F_i(\theta) \coloneqq \frac{1}{K}\sum_{i=1}^{K} E_{(x,y)\sim D_i}[l_i(\phi(x;\theta), y)] \coloneqq \frac{1}{K}\sum_{i=1}^{K}\frac{1}{N_i}\sum_{j=1}^{N_i}\left[l_i\left(\phi\left(x_j;\theta\right), y_j\right)\right],$$

όπου K το πλήθος των χρηστών. Ένας ευρύτατα αποδεκτός τρόπος για να λυθεί το πρόβλημα της κατανεμημένης μάθησης είναι με χρήση του αλγορίθμου federated averaging (FedAvg). Πιο αναλυτικά, έστω ένα υποσύνολο από χρήστες S των K χρηστών που ενορχηστρώνονται από έναν Server [Εικόνα 2] για να εκπαιδεύσουν από κοινού ένα μοντέλο μηχανικής μάθησης. Θεωρούμε πως κάθε χρήστης διαθέτει ένα ιδιωτικό σύνολο δεδομένων αποτελούμενο από N_i δείγματα (ζεύγη (x, y)). Στην συνέχεια θεωρούμε πως οι χρήστες συνεργατικά προσπαθούν να ελαχιστοποιήσουν την συνάρτηση κόστους.

Τα ακριβή βήματα του αλγορίθμου είναι τα εξής:

1. Ο Server μεταδίδει (broadcast) σε όλους τους χρήστες (ή σε ένα υποσύνολο των χρηστών) το global μοντέλο θ^t.
2. Κάθε χρήστης που συμμετέχει στο συγκεκριμένο iteration χρησιμοποιεί κάποιο αλγόριθμο μάθησης (όπως το Stochastic Gradient Descent (SGD)), για να εκπαιδεύσει το μοντέλο του στα δεδομένα που έχει στην διάθεσή του, για κάποιες τοπικές επαναλήψεις.
3. Έπειτα, ο κάθε χρήστης στέλνει στον Server το μοντέλο του θ_i^{t+1}.
4. Τέλος, ο Server κάνει aggregate την πληροφορία των χρηστών,
$$\theta^{t+1} \leftarrow \sum_{i=1}^{K} a_i\, \theta_i^{t+1}$$
όπου οι συντελεστές a_i, έχουν την ιδιότητα, $\sum_{i=1}^{K} a_i = 1$. Συνήθως, $a_i = \frac{1}{K}$.
Αφότου υπολογιστεί το νέο global model γίνεται πάλι broadcast στους χρήστες.
5. Τα παραπάνω βήματα επαναλαμβάνονται για ορισμένα iterations, έως ότου κάποιο κριτήριο σύγκλισης ικανοποιηθεί.

Ζητούμενα

1. (**Πειραματικό Ερώτημα**) Για τη συγκεκριμένη άσκηση θα χρησιμοποιήσουμε το dataset MNIST. Αποτελείται από 60.000 δείγματα στο training set και 10.000 δείγματα στο testing set. Περιέχει 10 διαφορετικές κλάσεις χειρόγραφων ψηφίων με διαστάσεις $1 \times 28 \times 28$. Αρχικά, επιλύστε το πρόβλημα της κεντρικοποιημένης μάθησης με την χρήση ενός κατάλληλου μοντέλου (π.χ. CNN, MLP). Δοκιμάστε για 2

διαφορετικές τιμές της κάθε παραμέτρου [**batch size**, **learning rate**], κρατώντας σταθερό το πλήθος των τοπικών επαναλήψεων [**local epochs**] και των ολικών επαναλήψεων [**global iterations**]. Τυπώστε την καμπύλη του training και testing loss, καθώς επίσης και την καμπύλη του testing accuracy.

2. (**Πειραματικό Ερώτημα**) Στην συνέχεια, στα πλαίσια της αποκεντρωμένης μάθησης, θεωρήστε πως υπάρχουν 10 χρήστες, στους οποίους θα πρέπει να μοιράσετε τα δεδομένα με IID (identically independently distributed) τρόπο. Αυτό σημαίνει πως κάθε χρήστης θα λάβει ίδιο περίπου ποσοστό δεδομένων από όλες τις διαθέσιμες κλάσεις (patterns), καθώς επίσης και ότι όλοι οι χρήστες θα διαθέτουν ίδιο πλήθος δεδομένων.
   Πραγματοποιήστε πάλι τα ζητούμενα του προηγούμενου ερωτήματος.
3. (**Πειραματικό Ερώτημα**) Στο τέλος θα εστιάσουμε στην πιο απαιτητική πρόκληση της αποκεντρωμένης μάθησης, καθώς θα θεωρήσετε πως υπάρχουν 10 χρήστες, στους οποίους μοιράζετε τα δεδομένα με non-IID τρόπο. Αυτό σημαίνει πως κάθε χρήστης διαθέτει δεδομένα που ανήκουν αποκλειστικά σε 2 από τα 10 patterns.
   Πραγματοποιήστε πάλι τα ζητούμενα του πρώτου ερωτήματος.

Tip

Καθώς το MNIST θεωρείται toy dataset, μπορείτε για λόγους ταχύτητας να κάνετε **αρχικές δοκιμές** του κώδικά σας, σε ένα μικρό υποσύνολο του αρχικού dataset. Μπορείτε για παράδειγμα, να κρατήσετε μόνο το 10% των δεδομένων από κάθε pattern. Ασφαλώς, τα τελικά αποτελέσματα θα περιέχουν **όλα τα διαθέσιμα δεδομένα** (μιας και αυτή είναι απαραίτητη προϋπόθεση για να προσεγγίσουμε το expectation μέσω του empirical average!)

Διαδικαστικά

- Προθεσμία παράδοσης: 20/02/2025.
- Οι ασκήσεις είναι ατομικές.
- Η τεχνική αναφορά (με τα αποτελέσματα, τα σχόλια και τον κώδικα που γράψατε) παραδίδεται ηλεκτρονικά μέσω eclass.
- Μπορείτε να χρησιμοποιήσετε οποιαδήποτε γλώσσα προγραμματισμού επιθυμείτε.

Για ερωτήσεις σχετικά με την εργασία, μπορείτε να απευθύνεστε στον Περικλή Θεοδωρόπουλο (st1002474 at ceid ...).