

Βαθιά Μάθηση

Εξαμηνιαίες εργασίες 2022

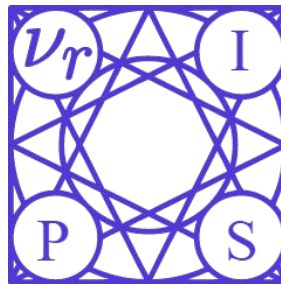


NIPS @NipsConference · 4m
#NIPS2018 The main conference sold
out in 11 minutes 38 seconds



Τα δύο σημαντικότερα συνέδρια ML & AI

[NIPS](#) (Neural Information Processing Systems) had 5,000 registered participants in 2016 and 8,000 in 2017 making it the largest conference in Artificial Intelligence



[ICML](#) (International Conference on Machine Learning).
One of the oldest conferences on ML. Along with NIPS it is one of the two primary conferences of high impact in ML and AI.



Επιλογή θέματος και παραδοτέα

- Επιλέξτε ένα θέμα (paper) με Βαθιά Μηχανική Μάθηση από το NIPS ή το ICML των τριών τελευταίων ετών
- Papers του NIPS [2021](#), [2020](#), [2019](#)
- Papers του ICML [2021](#), [2020](#), [2019](#)
- Ομάδες των 2 ή 3
- Διαφορετικό paper ανά ομάδα, όχι απαραίτητα διαφορετικό dataset
- [Φόρμα επιλογής paper](#)
- Θα βρείτε τις ομάδες (καθώς σχηματίζονται) και τα ήδη επιλεγμένα papers [εδώ](#)
- Παραδοτέα: κώδικας, logfiles τρεξίματος και αναφορά μέχρι 12 σελίδες

Στόχος εργασίας

- Στόχος εργασίας
 - Αναπαραγωγή αποτελεσμάτων.
 - Η αναπαραγωγή των αποτελεσμάτων ενός paper είναι μια καλή μορφή εξάσκησης.
 - Δοκιμάστε τροποποιήσεις που αναφέρονται στο future work section των papers
 - Αναλύστε τα δομικά στοιχεία της μεθόδου και δοκιμάστε λχ άλλες μεθόδους ή άλλες υπερπαραμέτρους
 - Επεκτείνετε την τεχνική που προτείνετε
 - δοκιμάζοντας την σε άλλα δεδομένα
 - δοκιμάζοντας παρεμφερείς τεχνικές στα ίδια δεδομένα

Computing & datasets

- Υπολογιστική ισχύς
 - Αρχικά πειραματίζεστε σε subset του συνόλου δεδομένων σας σε “παραδοσιακά” μέσα (Google Colaboratory, Cloud, δικές σας GPUs)
 - Όταν η ιδέα σας έχει ωριμάσει, την εκτελείτε στο [GRNET HPC](#), στο Google Cloud ή σε δικές σας GPU, για όλα τα δεδομένα και περισσότερα epochs εκπαίδευσης
- Σύνολα δεδομένων
 - Βεβαιωθείτε ότι τα δεδομένα του paper είναι διαθέσιμα.
 - Αν πρόκειται να επιλέξετε το GRNET HPC ως περιβάλλον εκτέλεσης και δείτε ότι δύο ή περισσότερες ομάδες σκοπεύουν να χρησιμοποιήσουν το ίδιο πολύ μεγάλο dataset, επικοινωνήστε με τους διδάσκοντες για τον τρόπο διαμοιρασμού τους

HPC ARIS

- Εργασία στο HPC ARIS

- Μέσα στο home directory φτιάχνετε φάκελο της ομάδας σας με ονομασία teamN, όπου N ο αριθμός της ομάδας σας.
- Μέσα στο φάκελο της ομάδας σας θα πρέπει να υπάρχουν τα αρχεία του κώδικά σας, το slurm script καθώς και απαραίτητα τα logfiles των εκτελέσεων στο ARIS
- Μπορείτε να συμβουλευέστε τον φάκελο team0 (η ομάδα των διδασκόντων), για να δείτε λχ πως γράφεται το slurm script
- Επίσης, όταν εκπαιδεύετε μοντέλο στο HPC ARIS, συνίσταται να κρατάτε συχνά checkpoints, έτσι ώστε να μπορείτε να συνεχίζετε την εκπαίδευση από “μεταγενέστερη” εποχή και όχι κάθε φορά από την αρχή

Google Cloud ή δική σας GPU

- Εργασία στο Google Cloud ή σε δικές σας GPU
 - Η εργασία αυτή νοείται μέσω των GPU που παρέχονται από το Google Cloud και όχι μέσω των δωρεάν GPU του Google Colaboratory
 - Αν είστε υποψήφιοι διδάκτορες δικαιούστε δωρεάν credits 1.000 δολαρίων. [Οδηγίες εδώ](#)
 - Προσοχή: Θα πρέπει να δημιουργήσετε λογαριασμό χρέωσης (billing account) και να τον συνδέσετε με τον τραπεζικό σας λογαριασμό ή με χρεωστική/πιστωτική κάρτα για να διεκδικήσετε τα δωρεάν credits.
 - Στο τελικό παραδοτέο θα πρέπει να συμπεριλάβετε, εκτός της αναφοράς, τον κωδικά σας, τα logfiles της εκτέλεσης καθώς και αρχείο requirements.txt με τις απαιτούμενες βιβλιοθήκες της Python. Θα πρέπει επίσης να μας ενημερώσετε ποια έκδοση της Python χρησιμοποιήσατε
 - Δεν γίνονται δεκτές υποβολές notebooks

Σημειώσεις

Η πληρότητα στη βιβλιογραφική μελέτη είναι εξίσου σημαντική με τα πειραματικά αποτελέσματα. Βεβαιωθείτε ότι έχετε συμπεριλάβει τις σημαντικότερες εργασίες στο task που εξετάζετε.

Ακόμα και αν δεν βγάλετε καλύτερα αποτελέσματα, θα αξιολογηθεί η σκέψη, η λογική πορεία που ακολουθήσατε (αν υποστηρίζεται θεωρητικά ή από παρατηρήσεις των ιδιοτήτων των δεδομένων) και φυσικά τα πειράματα που κάνατε

Παράδοση εργασίας

- Το τελικό report πρέπει να είναι μια εργασία σε μορφή “σχεδόν δημοσιεύσιμη”. Δείτε [εδώ γενικές συμβουλές](#) για τη σύνταξη επιστημονικών δημοσιεύσεων και [εδώ ειδικά για Deep Learning και την εργασία](#). Μέγιστο μήκος 12 σελίδες χωρίς την βιβλιογραφία και τυχόν παραρτήματα.
- Για την αναφορά σας επιλέξτε ένα από τα ακόλουθα templates:
 - [IEEE Conference Proceedings](#)
 - [Springer Computer Science Proceedings](#)
- **Τα τελικά πειράματα πρέπει να τρέξουν απαραίτητα είτε στο GRNET είτε στο Google Cloud (όχι στο Google Colaboratory) ή σε δικές σας GPU**
- Παράδοση zip με αναφορά και κώδικα εδώ στο Helios
 - Αν έχετε εκτελέσει των κωδικά σας στο Google Cloud ή σε δικές σας GPU, θα συμπεριλάβετε στο zip τα logfiles εκτέλεσης καθώς και το αρχείο requirements.txt
- Deadline παράδοσης: όπως αναγράφεται στο Helios