

# Αλγοριθμική Επιστήμη Δεδομένων

Εισαγωγή και διαδικαστικά

Άνοιξη 2020

ΔΠΜΣ Επιστήμη Δεδομένων και Μηχανική Μάθηση (ε.δε.μ<sup>2</sup>)  
Εθνικό Μετσόβιο Πολυτεχνείο

---

# Περιεχόμενο μαθήματος

- Εντοπισμός κανόνων συσχέτισης και συνόλων στοιχείων που εμφανίζονται συχνά σε μεγάλα δεδομένα.
- Εντοπισμός παρόμοιων στοιχείων - locality sensitive hashing.
- Πολυδιάστατοι χώροι και μείωση διάστασης, τυχαίες προβολές και Johnson-Lindenstrauss λήμμα.
- Recommendation systems, singular value decomposition, εφαρμογές.
- Το web ως γράφημα και ως αλυσίδα Markov, ανάλυση συνδέσμων, PageRank. Ανάλυση κοινωνικών δικτύων, εντοπισμός κοινοτήτων.

# Περιεχόμενο μαθήματος

- Μηχανική μάθηση μεγάλης κλίμακας, PAC μοντέλο μάθησης, VC dimension, online μάθηση, support vector machines, boosting.
- Ροές δεδομένων, δειγματοληψία και sketching σε ροές δεδομένων.
- Συσταδοποίηση δεδομένων (clustering), k-means, spectral clustering.
- Ηλεκτρονική διαφήμιση, δημοπρασίες, Generalized Second Price.
- Ασφάλεια και ιδιωτικότητα δεδομένων, differential privacy.

# Προαπαιτούμενα

- **Αλγόριθμοι και πολυπλοκότητα:** ανάλυση χρονικής πολυπλοκότητας αλγορίθμων, βασικές δομές δεδομένων, βασικοί αλγόριθμοι γράφων, NP-πληρότητα.
- **Μαθηματικά εργαλεία:** πιθανότητες, concentration inequalities, Markov chains, γραμμική άλγεβρα, θεωρία γραφημάτων

# Διαδικαστικά

- ▶ Διδάσκοντες: Δ. Φωτάκης, Α. Παγουρτζής, Δ. Σούλιου
- ▶ Βοηθοί διδασκαλίας: Σ. Πετσαλάκης, ...
  
- ▶ Βαθμολογία:
  - ▶ 8 μον. από ΤΔ («βάση»: 3)
  - ▶ 3 μον. από ασκήσεις

# Βιβλιογραφία

- ▶ [Mining of Massive Datasets](#), Jure Leskovec, Anand Rajaraman, and Jeff Ullman
  - ▶ [www.mmds.org](http://www.mmds.org) (book, slides, videos, etc.)
- ▶ [Foundations of Data Science](#), Avrim Blum, John Hopcroft, and Ravindran Kannan
  - ▶ <https://www.cs.cornell.edu/jeh/book.pdf>
- ▶ .... και πολλά ακόμη βιβλία, άρθρα και ιστοσελίδες!