

# Algorithms for Data Science

EECS 498.004 - Special Topics

**Instructor:** Euiwoong Lee

This course will introduce algorithmic and theoretical aspects of data science. With the emergence of machine learning and data science, providing theoretical foundations for them will become increasingly important. The course will cover several important algorithms in data science and see how their performances can be analyzed. While fundamental ideas covered in EECS 376 (e.g., design and analysis of algorithms, NP-hardness, etc.) will be still important, some topics will introduce new concepts and ideas, including sublinear time algorithms, algorithms for continuous domains, and average-case analysis.

**A tentative list of topics:** Randomized methods for big data (streaming algorithms, dimensionality reduction), spectral methods (spectral graph theory and spectral clustering), continuous optimization (convex optimization, its applications to discrete optimization, and gradient descent), average-case problems (compressed sensing, stochastic block model).

**Prerequisites:** EECS 376 and linear algebra.

**Course description:** Design and analysis of algorithms for problems arising in data science, with an emphasis on differences and connections between the traditional discrete algorithms. Topics include randomized methods for big data (hashing, streaming, and dimension reduction), spectral methods (singular value decomposition, spectral clustering/partitioning), continuous optimization (convex analysis, gradient descent), and average-case problems (compressed sensing, stochastic block model).