IE 529
Statistics of Big Data and Clustering
MWF 1-1:50pm
MEB 153

Instructor: Prof. Carolyn Beck, room 159 CSL, *beck3@illinois.edu*
TA: Philip Paré

Reading material: Journal articles, lectures notes, slides and other readings will be posted to the course website; students are expected to check the website regularly and read the posted material.

Prerequisites: Probability and statistics at or above the level of IE 300, and linear algebra at the level of MATH 415.

Tentative List of Course Topics:

Weeks 1-3 **Introduction and Preliminaries:** (1) introduction to big data and clustering; (2) general statistical concepts and principles; (3) review of hypothesis testing and basic statistical inference; (4) linear algebra concepts.

Week 4 **Dimension reduction:** (1) principal component analysis; (2) canonical analysis; (3) multi-dimensional scaling.

Weeks 5-6 **Regression:** (1) linear regression (model and interpretation; model fitting; model diagnosis); (2) logistic regression; (3) local polynomial regression; (4) model evaluations using AIC and BIC.

Weeks 6-7 **Basic Clustering and Classification:** (1) k-means, k-centers, k-medians, k-means++; (2) spectral clustering; (3) expectation-maximization.

Week 8 **Additive models:** (1) generalized additive models; (2) boosting; (3) neural networks.

Weeks 9-11 **Student presentations**

Weeks 12-15 **Advanced Topics:** (1) regularization; (2) networks and graphs; (3) recommender systems.

Assignments:

- There will be 4-5 homework assignments (30% of grade).
- There will be 3-4 computational (i.e., programming) assignments (40% of grade).
- There will be 1 **team** project and presentation (30% of grade)