# DS-GA 3001: Tools and Techniques for Machine Learning (Spring 2021)

*Instructor: David S. Rosenberg*

## Course description

This course deals with a range of topics that come up when applying machine learning in practice. Roughly half the course will cover topics connected to machine learning with interventions, such as counterfactual learning, reinforcement learning, and causal inference. Inverse propensity methods for handling biased samples and control variate methods for reducing variance will be given special attention, as these form a common thread of techniques relevant to each of these topics. We will also cover calibrating probability forecasts, interpreting machine learning models, active learning, crowdsourcing and "data programming", as time permits.

## Prerequisites

- DS-GA 1003: Machine Learning or equivalent.
- DS-GA 1002: Probability and Statistics or equivalent.
- Comfort with conditional expectations, conditional probability modeling, basic Bayesian statistics, hypothesis testing and confidence intervals.
- Python programming required for most homework assignments.

## Topics and rough schedule

**DISCLAIMER:** We will cover the majority of the topics below, roughly one line per week. However, the topics of the first 8 weeks can be quite challenging, and if we need to take more time with them, we may **drop some of the topics at the end of the syllabus**.

- Conditional expectation and variance decomposition
- Missing responses: IPW and self-normalized estimators
- Missing responses: regression imputation, control variates, and doubly robust estimators

- Estimating average treatment effects (ATE) and conditional average treatment effects (CATE) in randomized control trials
- Bandits and Contextual bandits: policy gradient and Thompson sampling
- Batch learning from logged bandit feedback (1)
- Batch learning from logged bandit feedback (2)
- REINFORCE for reinforcement learning
- Calibrated probability predictions
- Methods for global feature importance
- Explaining black-box model predictions
- Crowdsourcing and "Data Programming"
- Active learning

More information about each topic can be found in a draft version of the syllabus here.

## Course Requirements and Evaluation

- **(50%) Homework**: 4 – 5 homework assignments; mix of model building and written mathematical exercises to reinforce the main concepts.
- **(20%) Weekly Quizzes**: Concept-check quizzes that reinforce the main ideas from lectures and lab, which students may use any resources to complete.
- **(30%) Project**: In groups of 3–4; details forthcoming, but one possibility is to reproduce the experiments from a paper of relevance to the course and extend them in some way (e.g. an additional dataset, a new evaluation process, comparing to another method, etc.).

## Academic Integrity Policy:

The course conforms to NYU's policy on academic integrity for students.

## Moses Statement

Academic accommodations are available for students with disabilities. The Moses Center website is `http://www.nyu.edu/csd`. Please contact the Moses Center for Students with Disabilities (212-998-4980 or mosescsd@nyu.edu) for further information. Students who are requesting academic accommodations are advised to reach out to the Moses Center as early as possible in the semester for assistance.