INDR 450/550

Spring 2022

Course Introduction
Feb. 14, 2022

Fikri Karaesmen

Course Description

- This course is an introduction to data science and prescriptive analytics in the context of operations. We take as examples typical operational problems of demand and sales forecasting, inventory management and pricing and revenue management. We review predictive analytics approaches from machine learning and statistics in the context of time-series forecasting and feature-based prediction. We then investigate prescriptive analytics approaches focusing on data-driven optimization and joint estimation and optimization. We will work with real data as much as possible and the implementations will be in Python and MATLAB.

Learning Outcomes

- use tools of predictive analytics for time series demand modeling and optimization
- understand and apply the concepts of model validation
- understand the data driven optimization framework for inventory and price optimization
- understand the basic stochastic dynamic programming framework, approximations and reinforcement learning tools
- combine and use predictive analytics tools for solving optimization problems from operations

Course Evaluation Methods

Method	Description	Weight %
Midterm Exam	Midterm Exam	30.00
Project	Course Project (Groups of 2 or 3)	35.00
Homework	4 Homework Assignments	30.00
Participation	Lecture attendance, in-class exercises, discussion	5.00
	Total:	100.00

Homeworks

- Groups of 2 or 3 (or alone)

- We’ll provide help and examples on implementation but you will have to implement your own computation.

Participation

- To get full participation grade, you have to be present synchronously (i.e. At the lecture time) and complete the polls/ class exercises.
- You can get a partial grade for watching lecture/PS videos on line.

- This course is more fun with your live participation ☺

Blackboard page

- Please follow course blackboard page for lecture content
  - Lecture slides
  - Notes
  - Implementation videos
  - Problem Solutions
  - Resources
- Please follow announcements / e-mails: that’s our only way to reach you.

Course Schedule (subject to change)

Meeting Times	Subject
FEB. 14	Time series and ARIMA Forecasting, Estimation issues.
FEB. 21	Simple Regression based methods and tradeoffs
FEB. 28	Multi-feature regression and bias/variance tradeoff
MARCH 7	Regularization: ridge and lasso regression
MARCH 14	Advanced regression based methods: non-linear approaches
MARCH 21	Advanced regression based methods: non-linear approaches
MARCH 28	Single-period random demand inventory problem and optimization perspectives
APRIL 4	Joint Estimation and Optimization Framework: modeling extensions
APRIL 11	Capacity Allocation for Revenue Management: Finite Horizon Discrete Time Stochastic Dynamic Programming
APRIL 25	The reinforcement learning framework
MAY 2	Reinforcement learning and capacity allocation
MAY 9	Single Item Dynamic Pricing: Stochastic Dynamic Programming formulation.
MAY 16	Reinforcement learning and dynamic pricing
MAY 23	Reinforcement learning and dynamic pricing

Implementations

- We’ll work with real data as much as possible
  - sometimes we’ll combine real data with synthetic data for optimization purposes
- I’ll try to provide documentation and videos for implementation examples.
- Some MATLAB and some Python
  - Most of the time nothing requiring advanced knowledge or expertise in programming
  - If you have advanced skills, you can easily do better on what I can offer you in terms of implementation.

Books

- Two books that are available as free texts on line:
  - Hyndman and Athanasapoulos, *Forecasting: Principles and Practices,* https://otexts.com/fpp2/
  - James, Witten, Hastie, Thibshirani, *An Introduction to Statistical Learning 2nd ed.*, https://www.statlearning.com/

Warnings and disclaimers ☻

- I am offering this course for the first time. The schedule is subject to change and I have no prior documentation (slides etc.) to provide in advance.
  - No examples of former homeworks or exams
- My skills in python programming are very limited
  - Please don’t ask me technical questions (how to download jupyter notebook, why is my compiler not working etc.)
  - Same for data manipulation: I’ll prepare most of my data on a spreadsheet but if you are able to manipulate data within python you can make your life simpler
- It would be great to have a platform to share best practices (in terms of data manipulation, appropriate libraries etc.)

Warnings and disclaimers ☻

- There are excellent resources in predictive analytics
  - including the two recommended textbooks but also others
- There is no good book yet on Prescriptive Analytics that specializes in operations (inventory, capacity, price optimization)
  - Most of the developments are very recent
  - Please try to follow the lectures and read the papers
- We’ll look at Reinforcement Learning and DP approximations in a special setting (most books treat the general setting which is more complicated)

On the positive side

- While I have limited knowledge of implementations, I do know the methods and the theory (and the surrounding areas in probability, statistics and optimization)
- I have supervised completed master's and PhD theses in prescriptive analytics applied to inventories
  - A recent publication co-authored with former PhD student Davood Pirayesh Neghab
- Tons of experience in stochastic DP for inventory management, revenue management, pricing etc.

Prerequisites and other requirements

- Some prerequisites: INDR 262, ENGR 200 and INDR 252 (or some similar background in optimization and statistics)
- And to a certain extent: INDR 343, INDR 372 (or some basic knowledge on inventory and capacity management)

- I do not have time to review topics from these courses, but I will use them as needed. Please make sure to review them.

Positioning

- This course includes an overview of predictive ML methods but the end goal is to use them in the context of optimization.
  - This means there is some overlap with ENGR 421
- The predictive tools apply to many contexts but we'll maintain a focus on demand uncertainty especially on time-series data
  - This means there is some overlap with INDR 372
  - And maybe with courses in Econometrics
- We'll look at stochastic dynamic programming in a limited context (simple discrete-time models) in the context of revenue management.
  - There is some overlap with INDR 564 and INDR 475/575
  - But in these applications, we can have an entry point to approximation schemes fpr reinforcement learning

Prescriptive Analytics

- This is likely to be the right perspective for Industrial Engineering and Operations Research
  - Combine predictive analytics with optimization to solve decision problems involving uncertainties
  - Joint estimation (prediction through ML) and optimization (OR tools) is a new and exciting framework

Operational Problems

- Designing inventories, capacities, prices to minimize costs or maximize profits.
- There are many uncertainties, demand, yield, demand response to price etc.
- This requires looking at formulation that minimize an expected cost or maximizes an expected profit.
    - Or some more advanced risk measures taking into account the structure of the uncertainty.

A typical operational problem

- A standard optimization problem in operations looks like

$$\min_{\mathbf{z}} E[c(\mathbf{Y}, \mathbf{z})]$$

  where **z** is a decision variable and **Y** is a random variable. In addition, there could be constraints on the decision variable (i.e. $\mathbf{z} \in \mathcal{Z}$).
- To consider a concrete problem we can consider inventory planning at two stores with random demands (Y_1, Y_2) and the decisions could be the order quantities (z_1, z_2) that minimize the expected cost. This problem becomes interesting if inventory transshipments can take place between the stores.
  - We then need to consider the simultaneous decisions for (z_1, z_2), taking into account the correlation structure of (Y_1, Y_2).

A typical operational problem

- If we start with the assumption that the probability distribution of $\mathbf{Y}$ is known, then we have optimization frameworks (e.g. stochastic programming) to address such problems even at large scale.

- Some smaller scale problems can be solved analytically (the single-period random demand newsvendor problem is an example).

$$\min_q c_u E[(D-q)^+] + c_o E[(q-D)^+]$$

where D is the random demand, q is the order quantity and c_u and c_o are the underage and overage costs.

A typical operational problem

- In practice (reality), the probability distribution of $\mathbf{Y}$ is not known with certainty but we may have some past observations on hand for $\mathbf{Y}$: $(\mathbf{y_1}, \mathbf{y_2}, \ldots \mathbf{y_n})$.
- We may have observed demands of (28,43) at the two stores on day 1, (52, 25) on day 2 and so on.
- We then have options to "fit" a joint probability distribution using the observations or use the demand observations as scenarios that become inputs to the optimization problem.
  - For instance, we may fit a bivariate normal distribution to the data that specifies, the means, the standard deviations and the correlation.
  - A little more on this later.

A typical operational problem

- In practice, we might also have other relevant data is on external observations that may have an impact on $\mathbf{Y}$. Assume that in addition to $(\mathbf{y_1}, \mathbf{y_2}, ...\mathbf{y_n})$ we have recorded observations of other factors $(\mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_n})$
- In statistics, such observations are called covariates, in Machine LearningL the term features is used.
- For inventory optimization at the two stores, we may suspect that the demand in each store depends on factors such weather conditions at store locations, day of the week, week of the month, promotions, competitor's actions etc.
- Assume that we are planning inventories for the next day, which is a Tuesday, the second week of the month, weather sunny and 13 degrees etc. Then we can devise an order quantity that takes this information into account:

$$\min_{\mathbf{z}} E[c(\mathbf{Y}|\mathbf{X} = \mathbf{x}, \mathbf{z})]$$

A typical operational problem

- Machine learning (in the supervised learning framework) starts with data $(\mathbf{y_1}, \mathbf{x_1}), (\mathbf{y_2}, \mathbf{x_2})...(\mathbf{y_n}, \mathbf{x_n})$ and focuses on the prediction problem of $\mathbf{Y}|\mathbf{X}$
- and proposes a number of effective tools.
- On the other hand, prescriptive analytics focuses on:

$$\min_{\mathbf{z}} E[c(\mathbf{Y}|\mathbf{X} = \mathbf{x}, \mathbf{z})]$$

- and of course also on finding the minimizer $\mathbf{z}^*$.
- Note that the typical ML-based problem is also an optimization problem where some error function is minimized.
- Prescriptive analytics therefore considers such nested optimization problems one for estimation, the other on operational cost minimization.