INDR 450/550

Spring 2022

Lecture 22: Prescriptive analytics 4

May 11, 2022

Fikri Karaesmen

Announcements

- Class Exercise at the end of lecture today. If you are participating online, please upload your document under Course Contents/Class Exercises
- HW 4 is now available (model reduction, trees, forests etc.)
- Exam on May 13.
  - Review exercises are available
  - Make sure that you also review the class exercises, the labs, and the homeworks

The Problem

- Inventory or Capacity Management under Uncertain Demand
- Matching supply (order quantity) to the random demand to minimize expected underage and overage costs
  - Short selling season
  - Decision made in advance of the season
  - No replenishments or capacity additions during the season (purchasing in advance is required)
  - Demand forecasts have considerable uncertainty
  - Items lose value significantly after the season

Empirical Risk Minimization

Here's a stochastic optimization formulation that is a linear program:

$$\min_Q R(Q) = \frac{1}{n}\sum_{i=1}^{n} c_u\,(d_i - Q)^+ + c_o(Q - d_i)^+$$

$$\equiv \min_Q \frac{1}{n}\sum_{i=1}^{n} c_u\, {z_i}^+ + c_o {z_i}^-$$

$$s.t.$$

$$\begin{array}{lcll} {z_i}^+ & \geq & d_i - Q & i = 1,2,\ldots,n \\ {z_i}^- & \geq & Q - d_i & i = 1,2,\ldots,n \\ {z_i}^+, {z_i}^- & \geq & 0 & i = 1,2,\ldots,n \end{array}$$

Empirical Risk Minimization

- The stochastic optimization formulation has recently received a lot of attention.
- We'll next look at a more recent formulation that combines predictive analytics with prescriptive analytics.

Gah-Yi Ban and Cynthia Rudin, "The Big Data Newsvendor: Practical Insights From Machine Learning", *Operations Research*, Vol. 67, pp. 90-108, 2019.

Forecasting demand using predictive analytics

- Ice-cream store: Daily demand depends on
  - Day of week
  - Temperature
  - Weather condition (sunny, cloudy, rainy etc.)
- What is the demand for tomorrow?
  - Friday
  - Forecasted temperature 19^{o}C.
  - Partly sunny
- We have seen many approaches to handle such predictors (features): simple regression, non-linear regressions, random forests etc.

Forecasting demand

- Ice-cream store
- Naive model: ignore the dependence on the predictors.
- Then optimal order quantity for Friday (or any other day):

$$Q^{*}=F_{D}^{-1}(c_{u}/c_{u}+c_{o})$$

where F_D is the empirical distribution coming from past demand data.

The separated optimization framework

- Ice-cream store: better model
- Model the dependence on the predictors:
- $D=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\varepsilon$ where ε is Normal$(0,\sigma_\varepsilon)$.
- Estimate coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ (by regression)
- Then, optimal order quantity for Friday:

$$Q^*=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+z^*\sigma_\varepsilon.$$

where (x_1, x_2, x_3) are the observed predictors.

Separated vs. Joint Optimization

- The above approach is standard we first use predictive analytics to estimate the demand and then solve an optimization problem for the optimization approach.
- We can call this separated estimation and optimization (estimation and optimization stages are clearly separated).
- Ban and Rudin (2019) take an alternative approach and propose a joint estimation and optimization approach.

Feature based newsvendor

- Ice-cream store
- Demand depends on observable features. Let $\boldsymbol{x}$ be the feature vector.
- Then the cost minimization problem is

$$\min_{Q(\boldsymbol{x})} E[C(Q(\boldsymbol{x}); D(\boldsymbol{x}) \mid \boldsymbol{x})]$$

Feature based newsvendor: data

- Now assume that past demand data as a function of the feature is available: we have observations: $S_n = (d_1, \boldsymbol{x}_1), (d_2, \boldsymbol{x}_2), \ldots, (d_n, \boldsymbol{x}_n))$.
- For the ice cream vendor:

Data Point	Day	Temp.	Weather	Sales
1	Monday	12	Rainy	23kg
2	Tuesday	14	Cloudy	28kg
3	Wed.	14	Sunny	35kg
4	Thu.	17	Sunny	30kg
…	…	….	…	…

Feature based newsvendor: machine learning ideas

- Formulate and find a way to solve the following problem:

$$\min_{Q(\cdot)} R(Q(\cdot), S_n) = \frac{1}{n}\sum_{i=1}^{n} c_u\,(d_i - Q(\mathrm{x}_i))^+ + c_o(Q(\mathrm{x}_i) - d_i)^+$$

- To solve the above, we need to guess the functional form of $Q(\boldsymbol{x})$.
- Reasonable guess: a linear decision rule

$$Q(\mathrm{x}) = Q((x^1, x^2, \ldots, x^p)) = q^0 + \sum_{j=1}^{p} q^j\, x^j$$

- Note that this is supported by what we have seen before: the optimal order quantity is a linear function of the predictors and the standard deviation of the estimation error

Feature based newsvendor: machine learning ideas

- Here's the ML – optimization formulation:

$$\min_{Q(\cdot)} R(Q(\cdot), S_n) = \frac{1}{n}\sum_{i=1}^{n} c_u\,(d_i - Q(\mathrm{x}_i))^+ + c_o(Q(\mathrm{x}_i) - d_i)^+$$

$$\equiv \min_{Q=(q^1,q^2,\ldots q^p)} \frac{1}{n}\sum_{i=1}^{n} c_u\, {z_i}^+ + c_o {z_i}^-$$

$$s.t.$$

$$\begin{aligned}
{z_i}^+ &\geq d_i - \left(q^0 + \sum_{j=1}^{p} q^j x_i^j\right) && i = 1,2,,\ldots,n \\
{z_i}^- &\geq q^0 + \sum_{j=1}^{p} q^j x_i^j - d_i && i = 1,2,,\ldots,n \\
{z_i}^+, {z_i}^- &\geq 0 && i = 1,2,,\ldots,n
\end{aligned}$$

This is a linear program and can therefore be efficiently solved at large scale.

Feature based newsvendor: big data

- We can shrink the model by using lasso or ridge

$$\min_{Q(\cdot)} R(Q(\cdot), S_n) = \frac{1}{n}\sum_{i=1}^{n} c_u \left(d_i - Q(\mathrm{x}_i)\right)^+ + c_o (Q(\mathrm{x}_i) - d_i)^+ + \lambda \sum_{l=1}^{p} |q^l|$$

$$\equiv \min_{Q=(q^1,q^2,\ldots q^p)} \frac{1}{n}\sum_{i=1}^{n} c_u\, {z_i}^+ + c_o {z_i}^-$$

$$s.t.$$

$$\begin{aligned}
{z_i}^+ &\geq d_i - \left(q^0 + \sum_{j=1}^{p} q^j x_i^j\right) && i = 1,2,,\ldots,n \\
{z_i}^- &\geq q^0 + \sum_{j=1}^{p} q^j x_i^j - d_i && i = 1,2,,\ldots,n \\
{z_i}^+, {z_i}^- &\geq 0 && i = 1,2,,\ldots,n
\end{aligned}$$

No longer an LP but still a convex optimization problem, therefore solvable.

Feature based newsvendor: machine learning ideas

- For linear decision rules and without regularization, this is an LP
- For lasso or ridge regularization, it is a convex optimization problem.
- We have seen that with a large number of predictors, regularization is crucial.
- Other ML tricks: Kernel Optimization

Feature based newsvendor: Kernel-based methods

- Ban and Rudin also consider a Kernel-based formulation based on a Nadaraya-Watson estimator.
- This is a k Nearest Neighbours approach with a sophisticated weight function.
- Consider that the k-nearest neigbours are: $d_1, d_2, \ldots, d_k$ and their corresponding weights are $w_1, w_2, \ldots, w_k$.
- The resulting problem is equivalent to solving a discrete demand newsvendor problem with a probability mass function: $w_1, w_2, \ldots, w_k$.

Feature based newsvendor: Kernel-based methods

- Assume that Friday is expected to be sunny, temperature 15 degrees.
- It is somewhat similar to the first four days but more similar to days 3 and 4.
- Assume that the weights are: $w_1 = 0.1$, $w_2 = 0.2$, $w_3 = 0.4$, $w_4 = 0.3$.

Data Point	Day	Temp.	Weather	Sales
1	Monday	12	Rainy	23kg
2	Tuesday	14	Cloudy	28kg
3	Wed.	14	Sunny	35kg
4	Thu.	17	Sunny	30kg
…	…	….	…	…

$$Q^* = \min\left\{F_D(Q) \geq \frac{c_u}{c_u + c_0},\ Q \in \mathrm{Z}^+\right\}$$

Feature based newsvendor: Kernel-based methods

- Assume that the weights are: $w_1 = 0.1$, $w_2 = 0.2$, $w_3 = 0.4$, $w_4 = 0.3$.
- The point prediction for Friday is then:
- $\hat{y}_5 = (0.1)(23) + (0.2)(28) + (0.4)(35) + (0.3)(30) = 30.9$
- If we take this point estimator as our basis for approximation then:

$$Q^* = 30.9$$

Data Point	Day	Temp.	Weather	Sales
1	Monday	12	Rainy	23kg
2	Tuesday	14	Cloudy	28kg
3	Wed.	14	Sunny	35kg
4	Thu.	17	Sunny	30kg
…	…	….	…	…

Feature based newsvendor: Kernel-based methods

- But we can do better by taking the weights as a probability distribution.
- Assume that the weights are: $w_1 = 0.1$, $w_2 = 0.2$, $w_3 = 0.4$, $w_4 = 0.3$.

Data Point	Day	Temp.	Weather	Sales
1	Monday	12	Rainy	23kg
2	Tuesday	14	Cloudy	28kg
3	Wed.	14	Sunny	35kg
4	Thu.	17	Sunny	30kg
…	…	….	…	…

Sales	CDF
23kg	0.1
28kg	0.3
30kg	0.6
35kg	1

$$Q^* = \min\left\{F_D(Q) \geq \frac{c_u}{c_u + c_0},\ Q \in \mathrm{Z}^+\right\}$$

This time we get a different order quantity when the critical fraction changes.

Feature based newsvendor: new vs. old

- Contrast the ML approach with standard separated predictive analytics and optimization
  - In predictive analytics, we estimate the demand distribution as a function of the parameters and then solve the standard newsvendor problem.
  - In joint estimation and optimization, we do not attempt to forecast the demand but go directly for the optimal decision rule.
- Note that the predictive analytics problem is also an optimization problem: (i.e. Least squares optimization, lasso, finding the optimal tree approximation).
- The objective function of the estimation problem is not aligned with that of the operational optimization problem.
- For high margin items (c_u>> c_o), overestimation of demand might be better than underestimation of demand but regression does not take this into account.

Feature based newsvendor: new vs. old

- It turns out that the joint estimation-optimization approach may perform better than separated estimation and optimization with small and medium size data samples and a large number of predictors.
- This is subject of ongoing research.

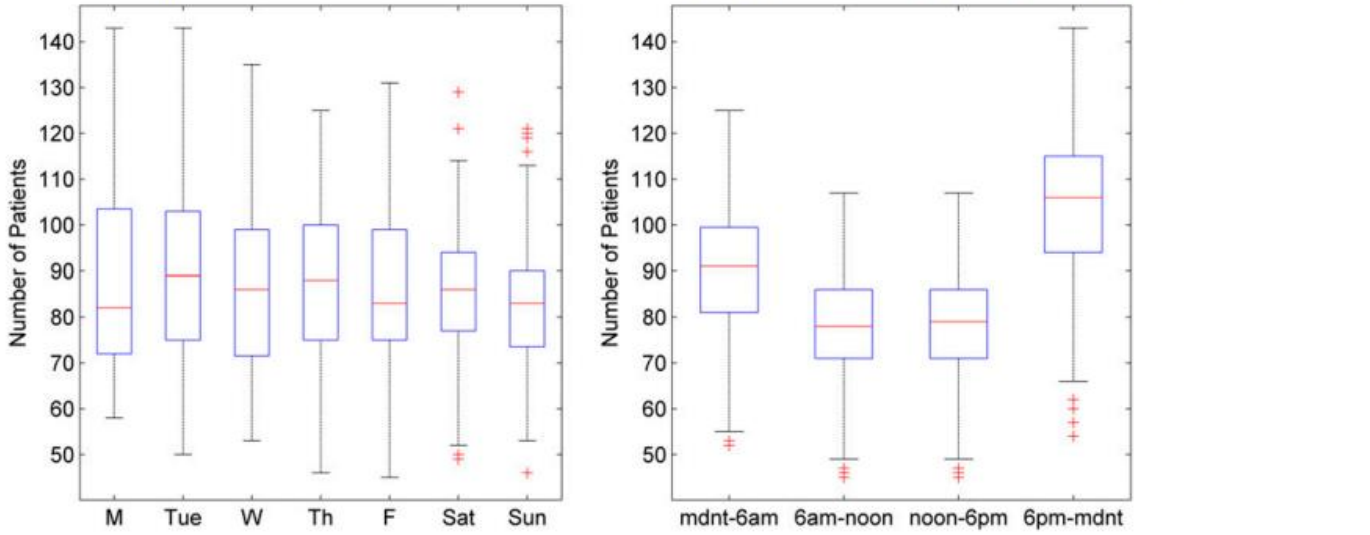
The Nursevendor Case from Ban and Rudin

- Ban and Rudin (2019) take the example of an emergency room nurse staffing problem: find the number of nurses to staff on a given day and time.
- ‘Our data comes from the emergency room of a large teaching hospital in the United Kingdom from July 2008 to June 2009. The data set includes the total number of patients in the emergency room at two-hour intervals. We provide box plots of the number of patients by day and by time periods in Figure 1. We assumed a nurse-to-patient ratio of one to five; hence, the demand is the total number of patients divided by five. We do not require the staffing level to be an integer in our predictions as multiskilled workers could be used for part-time work. We also assumed that the hourly wage of an agency nurse is 2.5 times that of a regular nurse, that is b = 2.5/3.5 and h= 1/3.5, resulting in a target fractile of r= b/(b + h) 2.5/3.5. Although the exact agency nurse rate differs by location, experience, and agency, our assumption is a modest estimate (Donnelly and Mulhern 2012).’

The Nursevendor Case from Ban and Rudin

- Boxplots for Number of Patients

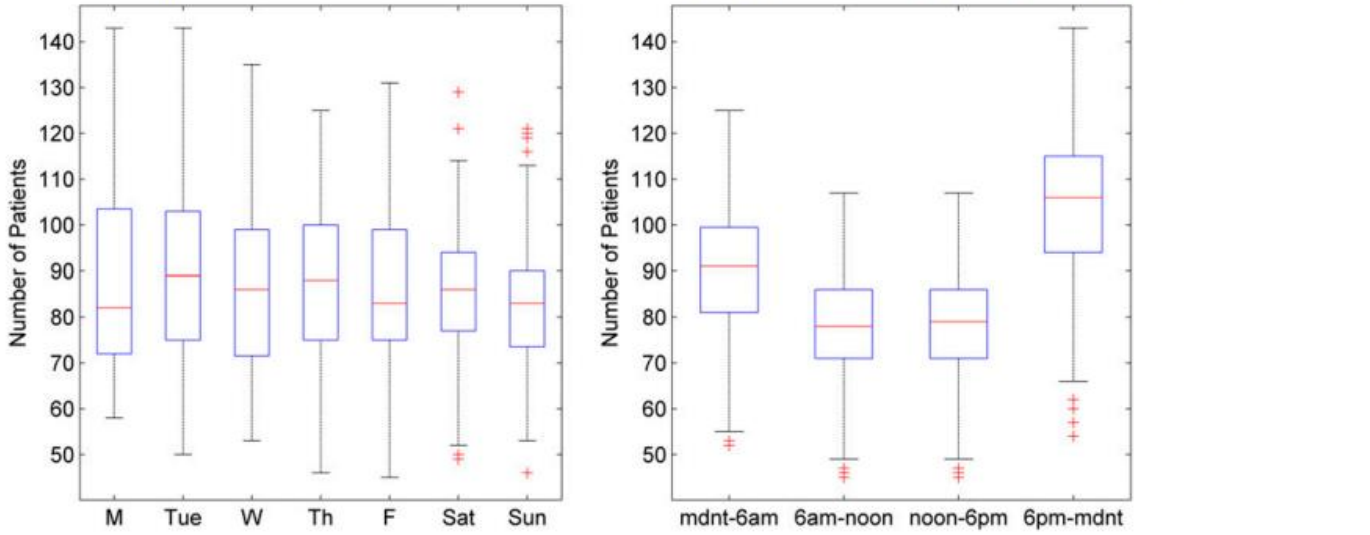
Figure 1. (Color online) A Boxplot of the Number of Patients in the Emergency Room (left) by Day and (right) by Time Period



The Nursevendor Case from Ban and Rudin

- Boxplots for Number of Patients

Figure 1. (Color online) A Boxplot of the Number of Patients in the Emergency Room (left) by Day and (right) by Time Period



The Nursevendor Case from Ban and Rudin

- Features and validation:
- ‘We considered two sets of features: the first set being the day of the week, time of the day, and m number of days of past demands and the second set being the first set plus the sample average of past demands and the differences in the order statistics of past demands, which is inspired by the analysis in Liyanage and Shanthikumar (2005) as described in Section 2.4.1. We refer to these features as operational statistics features. We used n 1,344 past demand observations (16 weeks) as training data and computed the critical staffing level 3 periods ahead. We then recorded the out-of-sample newsvendor cost of the predicted staffing level on 1,344/2 = 672 validation data on a rolling horizon basis, following the-rule-of thumb in Friedman et al. (2009) for choosing the size of the validation data set. Any parameter that needs calibration was calibrated on the validation data set. We then applied the algorithms to a test set of 672 unseen observations

The Nursevendor Case from Ban and Rudin

- Methods

- 1. SAA by day of the week: take a sample average of the training data set by day of the week (because our training data set consists of 16 weeks of demand, there are 1344/16 84 observations for each day of the week). We note that this is reflective of nurse staffing done in practice.
- 2. Cluster + SAA: we take the vector of features, then first classify them into k 2, ..., 12 clusters before applying SAA. This is an intuitive and alternative method to use the feature data. For clustering, we use the k-means clustering algorithm.
- 3. Solve (NV-KO) with the Gaussian kernel with the day of the week and time of the day features and an increasing number of days of past demands (for up to two weeks) with and without operational statistics features. These constitute a total of two algorithms.
- 4. Solve (NV-ERM1) with the day of the week and time of the day features and an increasing number of days of past demands (for up to two weeks) with and without operational statistics features, which are explained in Section 2.4.1. These constitute a total of two algorithms.

The Nursevendor Case from Ban and Rudin

- Methods
- 5. Solve (NV-ERM2) with the day of the week and time of the day features and two weeks of past demands with and without OS features for a range of regularization parameters. We investigate both $\ell 1$ and $\ell 2$ regularizations. These constitute a total of four algorithms.
- 6. Separated estimation and optimization: a commonsense approach to incorporating features in newsvendor decision making is by first regressing the demand on the features assuming a normally distributed error term (estimation) then applying the appropriate formula for the optimal quantile using the assumption of normality for the demand (optimization). We use day of the week and time of the day features and an increasing number of days of past demands (for up to two weeks).We consider two cases: one without and one with OS features.
- 7. Separated estimation and optimization approach with $\ell 1$ or $\ell 2$ regularization: we apply OLS regression with $\ell 1$ or $\ell 2$ or with no regularization to first estimate a demand model, then choose the optimal quantile under the assumption of normally distributed demand. For the demand estimation step, we use day of the week and time of the day features and two weeks of past demands and consider using and not using OS features.
- 8. We also consider Scarf's Minimax approach (Scarf et al. 1958).

The Nursevendor Case from Ban and Rudin

- Results

Table 3. A Summary of Results

Method	Calibrated parameter	Avg. computation time (per iteration)	Mean (95 % CI)	% savings relative to SAA-day	Annual cost savings rel. to SAA-day
1a. SAA-day	—	14.0 s	1.523 (± 0.109)	—	—
1b. Cluster + SAA	—	14.9 s	1.424 (± 0.102)	—	—
2a. Ker-0	$w = 0.08$	0.0444 s	1.208 (± 0.146)	20.7%	£39,915 ($63,864)
2b. Ker-OS	$w = 1.62$	0.0494 s	1.156 (± 0.140)	24.1%	£46,555 ($74,488)
3a. NV-0	12 days	325 s	1.326 (± 0.100)	12.9%	£24,909 ($39,854)
3b. NV-OS	Four days	360 s	1.463 (± 0.144)	—	—
4a. NVreg1	1×10^{-7}	84.5 s	1.336 (± 0.100)	—	—
4b. NVreg1-OS	1×10^{-7}	114 s	1.174 (± 0.113)	22.9%	£44,219 ($70,750)
5a. NVreg2	5×10^{-7}	79.6 s	1.336 (± 0.110)	—	—
5b. NVreg2-OS	1×10^{-7}	107 s	1.215 (± 0.111)	20.2%	£39,065 ($62,503)
6a. SEO-0	One day	10.8 s	1.279 (± 0.099)	16.0%	£30,952 ($49,523)
6b. SEO-OS	Six days	16.1 s	12.57 (± 10.63)	—	—
7a. SEOreg1	5×10^{-1}	22.1 s	1.417 (± 0.106)	—	—
7b. SEOreg1-OS	5×10^{-3}	25.9 s	11.95 (± 6.00)	—	—
8a. SEOreg2	1×10^{-1}	26.6 s	1.392 (± 0.105)	—	—
8b. SEOreg2-OS	5×10^{-3}	27.1 s	12.57 (± 10.63)	—	—
9. Scarf	12 days	20.8 s	1.593 (± 0.114)	—	—

Notes. We assume the hourly wage of an agency nurse is 2.5 times that of a regular nurse. We report the calibrated parameter (if any), the average computational time taken to solve one problem instance, and the mean and the 95% confidence interval for the out-of-sample staffing cost in normalized units. In the last column, we report the annual cost savings of the method relative to SAA-day in instances in which there is a statistically significant net cost saving, assuming a regular nurse salary of £25,000 (which is the Band 4 nurse salary for the National Health Service in the United Kingdom in 2014) and standard working hours. A dashed line represents cost differential that is not statistically significant. Cost savings in USD are also reported, assuming an exchange rate of £1: USD 1.6.