Adding Optimization to Your Analytics Toolbox
Dr. Gwyneth Butera and Dr. Russell Halper
Agenda

• Basics of Optimization and MIP (Mixed Integer Programming)
• Identifying Optimization Problems Within Your Organization
• How MIP Complements Machine Learning
Basics of Optimization and MIP (Mixed Integer Programming)

Dr. Gwyneth Butera
Mathematical Optimization Primer

• What do we mean by mathematical optimization?
• What does “optimal” mean?
• Basic form of a mathematical programming problem (MP)

Objective: 
minimize $c^T x$

Constraints: 
$Ax = b$ (linear constraints)

$l \leq x \leq u$ (bound constraints)

some or all $x_j$ must take integer values (integrality constraints)

• Many types of MP problems including LP, MIP, QP
• Typical modeling applications
Basic Mixed-Integer Programming Example

“Modular” furniture manufacturing
- \( b \) – number of bar stools,
- \( c \) – number of chairs.

\[
\begin{align*}
\text{max } & \sum_{b,c} (10b + 11c) \\
\text{subject to } & b + c \leq 10, \\
& 3b + 4c \leq 36, \\
& 3b + 5c \leq 41, \\
& 4b + 2c \leq 35, \\
& \text{integer } b, c \geq 0
\end{align*}
\]
Sample Formulation: Sports Scheduling Modeling

- Given a set of teams and time slots
- Produce a schedule for the matches between teams
- Goals
  - Fair and exciting schedule
- Constraints
  - Must obey league rules (e.g. number of home/away games)
  - No schedule conflicts (e.g. venue restrictions)
  - Potentially lots of others
Sports Scheduling: Basic Model

• Define a set of binary variables $X$
  • $X(i, j, k) = 1$ if team $i$ plays team $j$ in week $k$
  • $X(i, j, k) = 0$ otherwise

• Translate scheduling rules into linear constraints
  • Each team plays a game each week
    • $X(i, j_1, 1) + X(i, j_2, 1) + X(i, j_3, 1) + \ldots = 1$
  • Each team must eventually play every other team (in league, in division, ...)
    • $X(i, j, 1) + X(i, j, 2) + X(i, j, 3) + \ldots = 1$
Sports Scheduling: Basic Model

Prohibited games (teams 1 and 4 can’t play in week 3)
• $X(1,4,3) = 0$

Required games (teams 3 and 5 must play in week 7)
• $X(3,5,7) = 1$

No back-to-back division series (1 and 2 can’t play week 3 and week 4)
• $X(1, 2, 3) + X(1, 2, 4) \leq 1$

Objective
• Given projected TV audience for each game and each week, maximize the total audience
<table>
<thead>
<tr>
<th>Week</th>
<th>Games</th>
<th>Home Teams</th>
<th>Away Teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NE</td>
<td>CLE</td>
<td>BUF</td>
</tr>
<tr>
<td>2</td>
<td>CAR</td>
<td>IND</td>
<td>NJJ</td>
</tr>
<tr>
<td>3</td>
<td>NJJ</td>
<td>ATL</td>
<td>TB</td>
</tr>
<tr>
<td>4</td>
<td>BAL</td>
<td>NO</td>
<td>ATL</td>
</tr>
<tr>
<td>5</td>
<td>CLE</td>
<td>BAL</td>
<td>SF</td>
</tr>
<tr>
<td>6</td>
<td>NB</td>
<td>NO</td>
<td>PIT</td>
</tr>
<tr>
<td>7</td>
<td>MIA</td>
<td>BUF</td>
<td>NE</td>
</tr>
<tr>
<td>8</td>
<td>NO</td>
<td>NE</td>
<td>MIA</td>
</tr>
<tr>
<td>9</td>
<td>MIA</td>
<td>SD</td>
<td>GB</td>
</tr>
<tr>
<td>10</td>
<td>OAK</td>
<td>GB</td>
<td>MIN</td>
</tr>
<tr>
<td>11</td>
<td>ND</td>
<td>OAK</td>
<td>GB</td>
</tr>
<tr>
<td>12</td>
<td>NE</td>
<td>MIA</td>
<td>BAL</td>
</tr>
<tr>
<td>13</td>
<td>ATL</td>
<td>NJJ</td>
<td>CLE</td>
</tr>
<tr>
<td>14</td>
<td>TB</td>
<td>CLE</td>
<td>OAK</td>
</tr>
<tr>
<td>15</td>
<td>MIA</td>
<td>BUF</td>
<td>CLE</td>
</tr>
<tr>
<td>16</td>
<td>NE</td>
<td>NJJ</td>
<td>BUF</td>
</tr>
</tbody>
</table>

- **NBC, CBS, FOX, ESPN, NFLN**
- Shaded=home
- Left NFC (E, N, S, W), Right AFC (E, N, S, W)
Partial List of 2013 Rules and Goals

• Teams playing in London home week prior, on BYE week after
• Bills game in Toronto Week 13 against ATL, BAL, CAR, KC, MIA or NYJ
• No team has earliest BYE in consecutive seasons
• No team plays more than two road games against team coming off their BYE
• All teams playing road Thursday games are home the previous week
• Every team plays one Thursday game
• All teams playing home Thursday games have limited travel previous week
• Minimize non-division games during hurricane season (in order to “swap” sites)
• Minimize 3-game road trips (and 3-game home stands for teams with ticket issues)
• Minimize number of division series that end in first half of season
• Minimize number of games that would conflict with MLB postseason
• Maximize the number of late-season division games
• Minimize early-season games 1PM games for teams with weather concerns
• Minimize number of Pacific time zone teams that play at 1PM ET
• CBS/FOX have at least 3 1PM games every week, preferably geographically diverse
• CBS/FOX have at least 5 total games
• Team can play no more than 6 prime time games, and only 3 teams per year can play 6. All other teams can play no more than 5 primetime games (max 4 NBC)
Gurobi MIP Library

(5927 models)

Copyright © 2019, Gurobi Optimization, LLC
Linear Programming (LP) by Picture

\[
\begin{align*}
\max_{x \in \mathbb{R}^2} & \quad (x_1 + x_2) \\
\text{subject to} & \quad x_1 + 2x_2 \leq 2, \\
& \quad x_1, x_2 \geq 0
\end{align*}
\]

- \( f(x) = x_1 + x_2 = 0 \)
- \( f(x) = x_1 + x_2 = 1/2 \)
- \( f(x) = x_1 + x_2 = 1 \)
- \( f(x) = x_1 + x_2 = 3/2 \)
- \( f(x) = x_1 + x_2 = 2 \)

\( x_1 + 2x_2 = 2 \)
Mixed-Integer Programming (MIP) by Picture

\[ \begin{align*}
\text{max} & \quad (10 \, b + 11 \, c) \\
\text{subject to} & \quad b + c \leq 10, \\
& \quad 3b + 4c \leq 36, \\
& \quad 3b + 5c \leq 41, \\
& \quad 4b + 2c \leq 35, \\
& \quad b, c \geq 0
\end{align*} \]

**Caveat:** although may look almost like an LP, algorithmically we are not that good at exploring the integer lattice,

- even a “simple” (0,1)-cube in n-dimensions would have $2^n$ points!
  (for instance, in dimension 100 it would have \(~1.26 \times 10^{30}\) points)
Algorithms for Solving MP Problems

- Simplex (and dual simplex) method for LP
- Interior point algorithms for LP and QP
- Branch and bound for MIP
- Relaxation methods
- Heuristics
- Cutting planes
- Lazy constraints
What Makes MIP Difficult?

- LP vs MIP – polynomial vs NP-complete
- Numerical issues
Importance of Solver Performance

Comparison of Gurobi Versions (PAR-10)

- Time limit: 10000 sec.
- Intel Xeon CPU E3-1240 v3 @ 3.40GHz
- 4 cores, 8 hyper-threads
- 32 GB RAM

Test set has 5656 models:
- 410 discarded due to inconsistent answers
- 1493 discarded that none of the versions can solve
- Speed-up measured on >100s bracket: 2012 models

~53x improvement (8+ years)
Identifying Optimization Problems Within Your Organization
Dr. Russell Halper
When Does Optimality Matter?

Designing a Distribution Network

**Challenge:** Where do I put my assets and how do I move products between them?

**Decision:** A selection of sites to build assets, type of asset, and a flow of products to/from each

**Cost of a Single Sub-Optimal Decision:** Often times, tens of millions of dollars… or more!

Music Recommendation

**Challenge:** What songs would a customer enjoy?

**Decision:** A selection of one or more songs to play to a customer

**Cost of a Single Sub-Optimal Decision:** Negligible
Optimization can often be used when the business has a high degree of control over a reasonably quantifiable business problem.
How MIP Complements Machine Learning

Dr. Russell Halper
Many ML Techniques Use Some Type of Optimization

**PREDICTING A VALUE**

**Predict a numerical value**
- Is the predicted value linear in the data?

<table>
<thead>
<tr>
<th>Linear Methods</th>
<th>Nonlinear Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>Interpretability: Regression Trees</td>
</tr>
<tr>
<td>Prior knowledge and/or small data sets: Bayesian Linear Regression</td>
<td>Accuracy: Random Forests</td>
</tr>
<tr>
<td>Feature reduction: LASSO, Ridge</td>
<td>Large data sets: Neural Network Regression</td>
</tr>
</tbody>
</table>

**Predict a categorical value**
- Is the predicted value linear in the data?

<table>
<thead>
<tr>
<th>Linear Methods</th>
<th>Nonlinear Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic Approach: Logistic Regression</td>
<td>Interpretability: Classification Trees</td>
</tr>
<tr>
<td>Explicitly create boundaries: Support Vector Machines</td>
<td>Accuracy: Random Forests</td>
</tr>
<tr>
<td></td>
<td>Large data sets: Neural Network Classification</td>
</tr>
</tbody>
</table>

**SEGMENTING**

**Group together like data**
- Is the number of groups known?

<table>
<thead>
<tr>
<th>Known</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each data point in its own cluster: K-Means, K-Median</td>
<td>Always gives the same result: Hierarchical clustering, Affinity Propagation</td>
</tr>
</tbody>
</table>
Many ML Techniques Use *Some* Type of Optimization

<table>
<thead>
<tr>
<th>PREDICTING A VALUE</th>
<th>SEGMENTING</th>
</tr>
</thead>
</table>
| **Predict a numerical value**
Is the predicted value linear in the data? | **Group together like data**
Is the number of groups known? |
| Linear Methods
Linear Regression
Prior knowledge and/or small data sets
Bayesian Linear Regression
Feature reduction: LASSO, Ridge | Nonlinear Methods
Interpretability: Regression Trees
Accuracy: Random Forests
Large data sets: Neural Network Regression |
| Linear Methods
Interpretability: Logistic Regression
Express boundaries: Support Vector Machines |
| Known
Each data point in its own cluster: K-Means, K-Median | Unknown
Always gives the same result:
Hierarchical clustering
Affinity Propagation |

### Regression
Minimize the sum of squared errors when fitting the model

### Neural Networks
Stochastic Gradient Descent is used to minimize the average error over the training data set

### K-Median/K-Means
Assign a data point to a cluster to minimize some aggregate distance metric
K-Median Clustering as a MIP

From a set $P$ of data points, select a subset of $k$ points (i.e., “centers”) that minimizes the distance from each point in $P$ to its closest center.

**Index Sets**

$P =$ the set of points in the model

**Data**

$Dist_{ij} =$ Distance between points $i$ and $j$

**Variables**

$y_j = \begin{cases} 
1 & \text{if point } j \text{ is a Center} \\
0 & \text{otherwise} 
\end{cases}$

$x_{ij} = \begin{cases} 
1 & \text{if point } i \text{ is assigned to center } j \\
0 & \text{otherwise} 
\end{cases}$

Minimize:

$$\sum_{i,j \in P} Dist_{ij} x_{ij}$$

Minimize the Distance

Subject To:

$$\sum_{j \in P} y_j = k$$  Select $k$ centers

$$x_{ij} \leq y_j \text{ for all } i, j \in P$$  Only assign points to a selected center

$$\sum_{j \in P} x_{ij} = 1 \text{ for all } i \in P$$  Each point must be assigned to exactly one center

$$x_{ij}, y_j \in \{0, 1\} \text{ for all } i, j \in P$$  Variables are either 0 or 1
Predictive Models Often Require Actions to Be Taken Before Value is Realized....

Predictive ML
- Many excellent options!
- Interpretable:
  - Lasso
  - Random Forests
  - Classification Trees
- Non-Interpretable:
  - Neural Networks
  - Reservoir Computing

Prescriptive
- **Mathematical Programming (Optimization):** Provably optimal results. High degree of control over problem.
- **AI/Neural Networks:** Evolving technology and useful for many problems. Black box.
- **Simulation:** Useful for non-convex problems with multiple local minima or highly-linear problems
- **Heuristics:** Obtainable but can be inflexible when new constraints added to a business problem
Should I Use MIP, AI, or Another Approach?

- How important is optimality?
  - Is the cost of sub-optimal solutions high?
  - Most optimization models are NP-Complete problems. Polynomial runtime algorithms cannot find an optimal solution.

- Does data exist to train a model? Is it a new situation? Will the business problem change?

- What type of scenarios analysis needs to be done? What type of sensitivity analysis needs to be done?

- Do I need more than one solution?

- How big is the problem?
Retail chain w/ ~200 stores
+1400 Employees
Demand varies throughout the day
Complicated local labor laws
Problem statement: How much staff should we hire for each store?
Developing a Predictive Model

So how does one predict how much labor is needed?

- Register sales provided a key piece of data that help to understand the level of staffing over time.
- Other key demand drivers: Day of week, Time of Day, Holidays, Seasonality, Weather, ....
- What does the business need out of a predictive model:

Your favorite predictive ML model
What Does the Optimization Model need to Consider?

Determine how many employees to hire (minimize)
Schedule (what days, what hours, when to break)

Employees are of 2 types:
- Regular employees
  - Fixed monthly salary
  - Can be scheduled at most 6 consecutive days
  - Can be scheduled at most 2 consecutive Sundays
  - Same starting hour during weekdays
  - Fixed number of hours per day (9 hours)
  - Overtime is available (at additional cost)
  - Min and max number of hours per week
  - During a day, cannot work more than 6 hours without a 1-hour break
- Part-time employees
  - Salary determined by assigned hours
  - Less restrictions of consecutive days or Sundays
  - Flexible starting hours
  - No overtime is available

Model Penalties
- The model penalizes if there are fewer employees are assigned than the requirement (by hour, day)
- It is assumed that fewer employees will result in loss of sales

Total Cost = Salaries + Penalties + Overtime costs
Resources

• Access [www.gurobi.com](http://www.gurobi.com) and browse through Resources on our website
  • This includes code examples, webinar recordings, documentation and information on how to get started with optimization.

• Get a free 30-day trial of Gurobi
  • [www.gurobi.com/eval](http://www.gurobi.com/eval)

• Join us in next week’s webinar (May 21 & 22):
  • Optimization Application Demos - This presentation will walk through two live demos implemented by the Gurobi team and deployed on Amazon Web Services using Docker and Gurobi Instant Cloud.

• Need help leveraging Optimization for your business?
  • Contact [russell@e2eanalytics.com](mailto:russell@e2eanalytics.com).