UIC BUSINESS

Linear Regression and Subset Selection in Gurobi

IDS 435

Guest Lecturer: Parshan Pakiman

Discussion Outline

UIC BUSINESS

- Quick review of last session (recording posted on Mar 11th)
  - A procedure to write an optimization problem in Gurobi.
  - A toy example
- New material
  - Simple linear regression in Gurobi
  - Subset selection for linear regression using l_0 norm constraint in Gurobi

Quick review of last session

Elements of a formulation (completed version)

UIC BUSINESS

1. Sets and indices
2. Parameters (i.e., data)
3. Decision variables
4. Constraints
5. Objective function
6. Optimize
7. Analyze results (*Gurobi solved the model*)
8. Troubleshooting (*Gurobi could not solve the model*)

Let's look at Section "*A Procedure to Model an Optimization Problem in Gurobi*" from the Jupyter Notebook.

Applying elements of a formulation to a toy example

UIC BUSINESS

We use the following optimization problem to illustrate using Gurobi and the aforementioned procedure for using Gurobi:

$$\min_{x_1,x_2} \quad -x_1 - x_2 \longrightarrow \text{Linear}$$

$$x_1 + 2x_2 \leq 1 \longrightarrow \text{Linear}$$

$$2x_1 + x_2 \leq 1 \longrightarrow \text{Linear}$$

$$x_1, x_2 \geq 0$$

Let's look at Section "A Toy Example " from the Jupyter Notebook for the Gurobi model.

Simple linear regression

Simple Linear Regression (Single Feature)

UIC BUSINESS

- Consider a dataset of N points $\{(x^i, y^i): i = 1,2, \dots, N\}$.
- We want to fit a linear model with intercept β_0 and slope β_1 to approximate response variable y^iusing feature x^i.
- Specifically, we want the following constraints to hold as close as possible:

$$y^i \approx \beta_0 + \beta_1 x^i \qquad \forall i = 1,2, \dots, N.$$

- For each i, $\beta_0 + \beta_1 x^i$ is the approximate value for y^i.
- Our objective is to compute $\beta = (\beta_0, \beta_1)$ such that an "error" is minimized

Optimizing Mean Absolute Deviation (MAD)

UIC BUSINESS

MAD optimization can be written as the following linear optimization problem:

$$\min_{\beta_0,\beta_1} \quad \frac{1}{N}\sum_{i=1}^{N}(u_i + v_i) \longrightarrow \text{Linear}$$

$$y^i = \beta_0 + \beta_1 x^i + u_i - v_i, \qquad \forall i = 1, 2, \ldots, N, \longrightarrow \text{Linear}$$

$$u_i, v_i \geq 0, \qquad \forall i = 1, 2, \ldots, N,$$

$$\beta_0, \beta_1 \ \text{unrestricted.}$$

Optimizing Mean Squared Error (MSE)

UIC BUSINESS

MSE optimization can be written as the following optimization problem:

$$
\begin{aligned}
\min_{\beta_0,\beta_1} \quad & \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i^2 && \longrightarrow \text{Quadratic} \\
& \varepsilon_i = y^i - \beta_0 - \beta_1 x^i, \qquad \forall i = 1, 2, \ldots, N, && \longrightarrow \text{Linear} \\
& \varepsilon_i \ \text{ unrestricted}, \qquad \forall i = 1, 2, \ldots, N, \\
& \beta_0, \beta_1 \ \text{ unrestricted}.
\end{aligned}
$$

Let’s look at Section “Simple Linear Regression” from the Jupyter Notebook for the Gurobi model.

Feature Selection in Regression

Linear Regression with Multiple Features

UIC BUSINESS

- Building on the previous example, we study a linear regression problem in which the optimized linear model should only use a small subset of features.
- In other words, we need to select the best subset of features that their linear combinations have the lowest training error.
- Consider training set with $\{(x^i, y^i): i = 1,2, \dots, N\}$ with N observations:
  - $x^i = (x_1^i, x_2^i, \dots, x_d^i)$ is the d-dimensional feature vector for the i-th observation
  - y^i response variable for the i-th observation

- Our objective is to compute coefficients $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_d)$ for the following linear model

$$y^i \approx \beta_0 + \sum_{j=1}^{d} \beta_j x_j^i \qquad \forall i = 1,2, \dots, N.$$

Ordinary least squares (OLS)

UIC BUSINESS

- Define the following error term:

$$\varepsilon_i := y^i - \left(\beta_0 + \sum_{j=1}^{d} \beta_j x_j^i\right), \qquad \forall i = 1, 2, \ldots, N.$$

- OLS is the following optimization problem:

$$\min_{\beta} \ \mathrm{MSE}(\beta)$$

$$\mathrm{MSE}(\beta) = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i^2 = \frac{1}{N}\sum_{i=1}^{N}\left[y^i - \left(\beta_0 + \sum_{j=1}^{d} \beta_j x_j^i\right)\right]^2$$

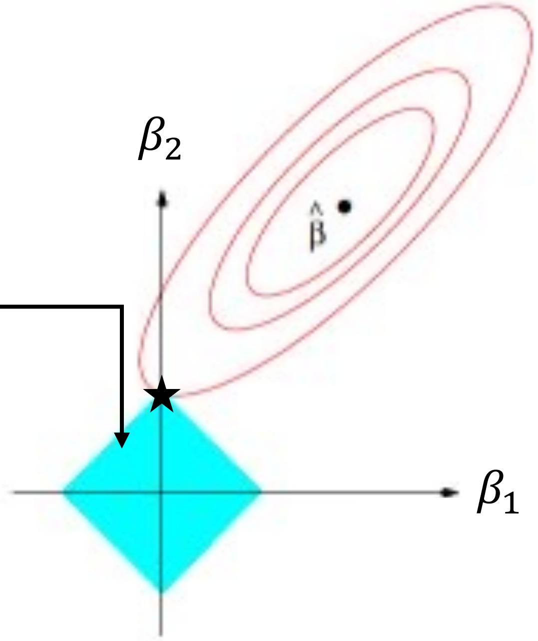
LASSO Regression

LASSO Regression

$$\min_{\beta} \ \mathrm{MSE}(\beta) \quad \text{s.t.} \quad \sum_{j=0}^{d} |\beta_j| \le T.$$

$$\min_{\beta} \ \mathrm{MSE}(\beta) + \lambda \sum_{j=0}^{d} |\beta_j|.$$

Indirect subset selection



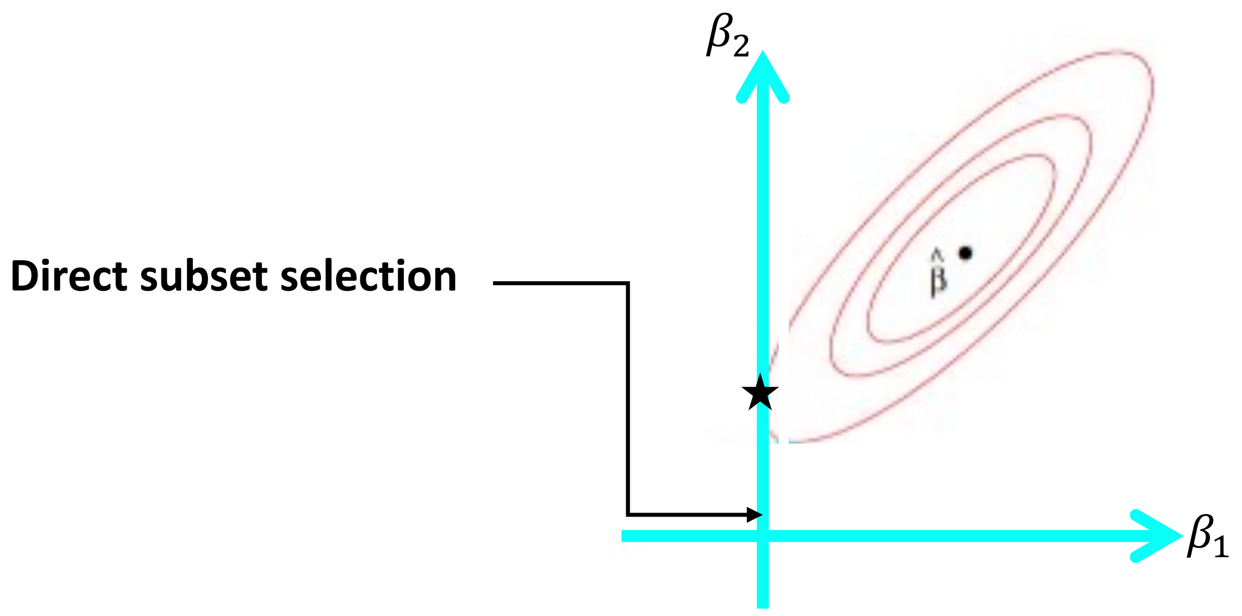
Cyan region is the feasible set

ℓ_0-Regression

UIC BUSINESS

$$\min_{\beta} \ \mathrm{MSE}(\beta) \quad \text{s.t.} \quad \|\beta\|_0 \le K$$

$$\|\beta\|_0 := \text{number of non-zero elements of } \beta.$$



Cyan region is the feasible set for $K = 1$

Comparison of Models

UIC BUSINESS

	OLS	LASSO	l_0-regression
Convex	Yes	Yes	No
Indirect subset selection	No	Yes	Yes
Direct subset selection	No	No	Yes
Constrained?	No	Yes	Yes
How to solve?	`LinearRegression` in sk-learn	`Lasso` in sk-learn	A model in Gurobi

Reformulating ℓ_0-Regression

UIC BUSINESS

- Let's focus on formulating the ℓ_0 constraint $||\beta||_0 \leq K$

$$\begin{cases} z_j := \begin{cases} 1 & \text{if } \beta_j \neq 0 \\ 0 & \text{if } \beta_j = 0 \end{cases}, & \forall j = 0, 1, \ldots, d; \\ \sum_{j=0}^{d} z_j = K \end{cases}$$

- Overall, we can write the ℓ_0-regression problem as the following MIQP:

$$\min_{\beta, z} \frac{1}{N} \sum_{i=1}^{N} \left[y^i - \left(\beta_0 + \sum_{j=1}^{d} \beta_j x_j^i \right) \right]^2 \longrightarrow \text{Quadratic}$$

$$\begin{aligned} & z_j = 0, \quad \text{if} \quad \beta_j = 0, && \forall j = 0, 1, \ldots, d, \\ & z_j = 1, \quad \text{if} \quad \beta_j \neq 0, && \forall j = 0, 1, \ldots, d, \end{aligned} \longrightarrow \text{If-then constraint (nonlinear)}$$

$$\sum_{j=0}^{d} z_j = K, \longrightarrow \text{Linear}$$

$$\begin{aligned} & \beta_j \quad \text{unrestricted}, && \forall j = 0, 1, \ldots, d, \\ & z_j \quad \text{binary}, && \forall j = 0, 1, \ldots, d. \end{aligned}$$

ℓ_0-Regression for California housing dataset

UIC BUSINESS

```
California Housing dataset
--------------------------

**Data Set Characteristics:**

    :Number of Instances: 20640

    :Number of Attributes: 8 numeric, predictive attributes and the target

    :Attribute Information:
        - MedInc        median income in block group
        - HouseAge      median house age in block group
        - AveRooms      average number of rooms per household
        - AveBedrms     average number of bedrooms per household
        - Population    block group population
        - AveOccup      average number of household members
        - Latitude      block group latitude
        - Longitude     block group longitude

    :Missing Attribute Values: None
```

Let’s look at Section “Trying function MIQP_version_1” from the Jupyter Notebook for the Gurobi model.

Why do we get error? Troubleshooting (Step 8).

UIC BUSINESS

Does not work in Gurobi	Works in Gurobi
$\frac{1}{N}\sum_{i=1}^{N}\left[y^i-\left(\beta_0+\sum_{j=1}^{d}\beta_j x_j^i\right)\right]^2$	$\beta^\top\left(\frac{1}{N}\hat{X}^\top\hat{X}\right)\beta-\left(\frac{2}{N}Y^\top X\right)\beta+\frac{1}{N}Y^\top Y$
Non-standard Quadratic From	Standard Quadratic From

$$\frac{1}{N}\sum_{i=1}^{N}\left[y^i-\left(\beta_0+\sum_{j=1}^{d}\beta_j x_j^i\right)\right]^2 = \beta^\top\left(\frac{1}{N}\hat{X}^\top\hat{X}\right)\beta-\left(\frac{2}{N}Y^\top X\right)\beta+\frac{1}{N}Y^\top Y$$

Corrected MIQP

UIC BUSINESS

$$
\begin{aligned}
\min_{\beta,z}\ & \beta^\top\left(\frac{1}{N}\,\hat{X}^\top\hat{X}\right)\beta-\left(\frac{2}{N}\,Y^\top X\right)\beta+\frac{1}{N}\,Y^\top Y \\
& z_j=0,\quad \text{if}\quad \beta_j=0, && \forall j=0,1,\dots,d, \\
& z_j=1,\quad \text{if}\quad \beta_j\neq 0, && \forall j=0,1,\dots,d, \\
& \sum_{j=0}^{d} z_j=K, \\
& \beta_j\quad \text{unrestricted}, && \forall j=0,1,\dots,d. \\
& z_j\quad \text{binary}, && \forall j=0,1,\dots,d.
\end{aligned}
$$

Let’s look at Section “Solving MIQP” from the Jupyter Notebook for the Gurobi model.

Final Comparison

UIC BUSINESS

```
OLS Testing MSE    : 0.52
LASSO Testing MSE : 0.54
MIQP Testing MSE   : 0.52

OLS Optimal Solution:
 [ 0.84  0.12 -0.28  0.32 -0.01 -0.04 -0.89 -0.86]

LASSO Optimal Solution:
 [ 0.74  0.12 -0.03  0.07 -0.   -0.02 -0.72 -0.67]

MIQP Optimal Solution:
 [ 2.07  0.83  0.12 -0.28  0.32  0.    0.   -0.89 -0.86]
```