Business Analytics

Session 12b. Convex Optimization

Renyu (Philip) Zhang

New York University Shanghai

Spring 2019

Ordinary Least Squares

Linear Regression Recap

- Training data: Outcomes Y_i ($i = 1, 2, \cdots, n$); covariates X_{ij} ($1 \leq i \leq n$ and $1 \leq j \leq p$)
- Linear Regression:
$$Y_i \approx \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j X_{ij}$$
- In matrix form: $Y \approx X\hat{\beta}$, where
$$\begin{aligned} Y &= (Y_1, Y_2, \cdots, Y_n)' \\ X &= \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots, & X_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} \\ \hat{\beta} &= (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2 \cdots, \hat{\beta}_p)' \end{aligned}$$
- What do we mean by "$\approx$"?

Ordinary Least Squares

- To fit a linear regression model, we try to find the coefficients $\hat{\beta}$ to minimize the total sum of squared errors:

$$\min_{\hat{\beta}} \mathcal{L}(\hat{\beta}) = \sum_{i=1}^{n} (Y_i - X_i\hat{\beta})^2$$

where $X_i = (X_{i0} = 1, X_{i1}, X_{i2}, \cdots, X_{ip})$

- Need to specify the gradient of the objective function.
- $\text{Grad}[\mathcal{L}(\hat{\beta})] = \left(\frac{\partial \mathcal{L}(\hat{\beta})}{\partial \beta_0}, \frac{\partial \mathcal{L}(\hat{\beta})}{\partial \beta_1}, \frac{\partial \mathcal{L}(\hat{\beta})}{\partial \beta_2}, \cdots, \frac{\partial \mathcal{L}(\hat{\beta})}{\partial \beta_p} \right)$, where

$$\frac{\partial \mathcal{L}(\hat{\beta})}{\partial \beta_j} = 2 \sum_{i=1}^{n} X_{ij}(X_i\hat{\beta} - Y_i)$$

- Python demonstration of OLS with the data wine.csv
  - Outcome: *Price*
  - Covariates: *AGST*, *HarvestRain*, *WinterRain*, and *Age*

Gradient of Multi-Variable Functions

- Gradient is the generalization of derivative for multi-variable functions:

$$\text{Grad}[f(x)] = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n} \right)$$

- Examples:
  - $\text{Grad}[(3x_1 + 2x_2)^2] = (18x_1 + 12x_2, 12x_1 + 8x_2)$
  - $\text{Grad}[\exp(2x_1 + x_2)] = (2\exp(2x_1 + x_2), \exp(2x_1 + x_2))$
  - $\text{Grad}[(\log(x_1 + 2x_2))^2] = \left(\frac{2\log(x_1+2x_2)}{x_1+2x_2}, \frac{4\log(x_1+2x_2)}{x_1+2x_2} \right)$

$$\text{Grad}[MV(x)] = (\mu_1 - 2\lambda(\sum_{j=1}^{6} \sigma_1\sigma_j\rho_{1j}x_j), \mu_2 - 2\lambda(\sum_{j=0}^{5} \sigma_2\sigma_j\rho_{2j}x_j), \cdots, \mu_6 - 2\lambda(\sum_{j=1}^{6} \sigma_6\sigma_j\rho_{6j}x_j))$$

Jacobian of Multi-Variable Vector-Valued Functions

- Jacobian is the generalization of Gradient for vector-valued functions $g(x) = (g_1(x), g_2(x), \cdots, g_m(x))'$:

$$\mathrm{Jac}[g(x)] = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x_1} & \frac{\partial g_1(x)}{\partial x_2} & \cdots & \frac{\partial g_1(x)}{\partial x_n} \\ \frac{\partial g_2(x)}{\partial x_1} & \frac{\partial g_2(x)}{\partial x_2} & \cdots & \frac{\partial g_2(x)}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_m(x)}{\partial x_1} & \frac{\partial g_m(x)}{\partial x_2} & \cdots & \frac{\partial g_m(x)}{\partial x_n} \end{pmatrix}$$

- Examples:
  - $\mathrm{Jac}[x_1 + x_2, 2x_1 - x_2]' = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$
  - $\mathrm{Jac}[\exp(x_1 + x_2), (x_1 - x_2)^2]' = \begin{pmatrix} \exp(x_1 + x_2) & \exp(x_1 + x_2) \\ 2x_1 - 2x_2 & 2x_2 - 2x_1 \end{pmatrix}$
  - $\mathrm{Jac}[B(x)] = (1, 1, 1, 1, 1, 1)$

Portfolio Optimization Problem Revisited

- Objective function: $\lambda \cdot x' \cdot Cov \cdot x - \mu' \cdot x$
  - Cov is the covariance matrix for the annual returns of the assets.
- Gradient of objective function: $2\lambda \cdot x' \cdot Cov - \mu'$
- Constraint(s): $\sum_{i=1}^{6} x_i = 1$
- Jacobian of constraint(s): $(1, 1, 1, 1, 1, 1)$

Non-Linear Programming in General

General Formulation in NLP

- A general formulation of NLP:

$$\min_{x} f(x)$$

Subject to

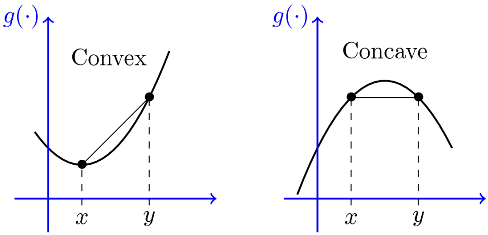
$$\begin{cases} g_1(x) \geq B_1 \\ g_2(x) \geq B_2 \\ \cdots\cdots \\ g_{m_1}(x) \geq B_{m_1} \\ h_1(x) = b_1 \\ h_2(x) = b_2 \\ \cdots\cdots \\ h_{m_2}(x) = b_{m_2} \\ x_1 \in [\underline{x}_1, \bar{x}_1], x_2 \in [\underline{x}_2, \bar{x}_2], \cdots, x_n \in [\underline{x}_n, \bar{x}_n] \end{cases}$$

Local Optimum and Global Optimum

- Local minimum:
  - $f(x^*) \leq f(x)$ for all $|x - x^*| \leq r$.
- Global minimum:
  - $f(x^*) \leq f(x)$ for all x satisfying the constraints.
- A big issue in NLP:
  - In general, an NLP algorithm can only obtain a local optimum, but not a global one.
  - A global minimum is a local optimum, but not vice versa.
- Resolving the local optimality issue:
  - Start the algorithm at multiple initial points.
  - Convex optimization.

Convex Optimization

- $f(x)$ is convex if $\lambda f(x) + (1 - \lambda) f(y) \geq f(\lambda x + (1 - \lambda) y)$ $(0 \leq \lambda \leq 1)$.
  - Line joining any two points is on or above the curve.
  - $f(x)$ is concave if $-f(x)$ is convex.



- Local minimum of a convex function is also a global minimum.
- Local maximum of a concave function is also a global maximum.
- In optimization, non-linear models are not considered difficult, but non-convex models are difficult.

Convex Functions

- Specific functions:
  - $f(x) = \sum_{i=1}^{n} a_i x_i$
  - $f(x) = ax^2 + bx + c$ for $a \geq 0$
  - $f(x) = \exp(x)$
  - $f(x) = -\log(x)$
  - $f(x) = -\sqrt{x}$
- General functions:
  - For a uni-variable function $f(x)$, if $f''(x) \geq 0$ for all x, $f(x)$ is convex.
  - If $f(x)$ is convex, so is $\lambda f(x)$ for $\lambda \geq 0$.
  - If $f(x)$ and $g(x)$ are convex, so is $f(x) + g(x)$.
  - If $f(\cdot)$ is convex and uni-variable, $f(Ax)$ is convex, where $A \in \mathbb{R}^{1\times n}$ and $x \in \mathbb{R}^n$.
  - If $f(x, Y)$ is convex in x, $\mathbb{E}_Y[f(x, Y)]$ is convex in x.
  - If $f(x, y)$ is convex in (x, y), $\min_y f(x, y)$ is convex in x.

Homework

- Submit your final project progress report on NYU Classes.
- Finish Homework 12 (NO need to submit it).