Machine Learning for Business Analytics

Assignment 5

Due: January/ 16 23:59 (GMT+8)

Instruction

- There are two parts in this assignment. You only have to hand in Part 1 regarding the Julia Programming. **Part 2 will not be graded**; of notice is that similar questions may show up in midterm or final.
- Please hand in your Julia code with a **PDF file**. The wrong format will not be graded.
- **Copying the assignment will result in zero points.**
- Late submission will be graded according to the following rule:
  Your grade = original grade × (1 − 0.05*h*), if you submit *h* hours after the deadline.

Part 1 Programming Exercises (100%)

*** In this assignment, you need to download "MarketData" and "Dates" packages. For stock data import, please refer to

https://docs.juliahub.com/MarketData/vKuMd/0.13.1/downloads/

1. (30%) (Unexpected Income) Barnaby is an employeet in Everblue Airline Corporation, which just announced an unprecedented 40-month year-end bonus. With this unexpected fortune, he has no idea of how to allocate the money. Barnaby's average income this year is $100,000 NTD per month and he wants to allocate his money on large TECH companies, such as Apple, Tesla, Amazon, TSMC(TSM), Intel, AMD, NVIDIA, Alphabet, Netflix and Facebook. Moreover, he would like to keep $88,888 NTD for the new-year red envelope for his parents this year. (This should be included as one asset on your portfolio) For simplicity, we denote μ_t as the expected return "vector" from year *t* to 2020, where the i^{th} element will be the expected return for asset *i* on that period (Expected return = average daily return* trading day), and Σ_t as covariance matrix from year *t* to 2020, where $(\Sigma_t)_{i,j}$ is the expected covariance for return of asset *i* and asset *j* (Expected covariance = the covariance of daily return*trading day). Notably, year *t* to 2020 refers to t/1/1~2020/12/31.

$$\text{Return} = \frac{\text{Close Price}_t - \text{Close Price}_{t-1}}{\text{Close Price}_{t-1}}$$

(1) Use functions in "MarketData" and "Dates" to import the stock data above. First, write a function called "Daily_return" that outputs daily return for the given stocks. Then, write a function called "Stock_info" that outputs the

expected returns on that period and the return covariance matrix. Print out μ_{2016}, μ_{2018} and Σ_{2016}, Σ_{2018} (round your result to 4 digits). For your inference, the code for each stock are listed below: AAPL, TSLA, AMZN, TSM, INTC, AMD, NVDA, GOOGL, NFLX, FB.

(2) Consider the data from 2016 to 2020, then using the optimization framework below to find the optimal allocation that Barnaby will hold until 2021-2025. The targets are 50%, 80%, 110%, 140%, 170%, respectively. You may think that you will reconstruct your portfolio at the beginning of the year, and the performance of the portfolio will be evaluated at the end of the year. Moreover, the expected covariance matrix and return vector evaluated by 2020 (2019-2020) data, that is, $\mu_{2020}, \Sigma_{2020}$ ($\mu_{2019}, \Sigma_{2019}$) will be applied when you construct the portfolio Barnaby will hold until 2021 (2022) and so on. (Reminder: Don't forget that Barnaby need to pay his parents.)

$$\min \quad w^T \Sigma w$$

$$\text{s.t.} \quad \begin{bmatrix} 1^T \\ \mu^T \end{bmatrix} w = \begin{bmatrix} 1 \\ \rho \end{bmatrix}$$

(3) After Barnaby finds the optimal asset allocation by the above formulation, he thinks there may be a deficiency by doing so. The objective above treats portfolio returns that exceed our target value the same as returns that fall short of our target value, whereas in fact we should be delighted to have a return that exceeds our target value. Therefore, he decides to apply the "downside risk" framework.

$$\text{Downside risk} = \frac{1}{T}\sum_{t=1}^{T}(\max\{\rho^{tar} - r_t, 0\})^2$$

In "downside risk" framework, *T* is the amount of trading date, ρ^{tar} is your target return and *r* is a *T*-vector, where r_t represents the return of your portfolio from *t*-1 to *t*. (You might need the last day of 2015 to compute r_1 when you construct the portfolio that Barnaby will hold until 2025) Reformulate the optimization problem, then solve it by implementing Levenberg-Marquardt algorithm (Don't forget the constraints). Set the tolerance to 10^{-5} and maximum iteration at least to 100. As for the initial point, you can consider the solution of (2) or evenly distributed portfolio.

(4) Use stacked bar plot or line plot to show the value of each stock in your portfolio at the end of each year (2021-2025) for (2) and (3) respectively. Compare two trading strategies and make some simple conclusions for the plots.

2. (25%) (An astronaut, Algae) Algae has a dream of becoming an impressive astronaut one day. Nowadays, she is interested in planning the operational orbit of a satellite, DTM (Doge To the Moon). We know that the orbit should follow

$$\hat{f}(t;\theta) = \begin{bmatrix} c_1 + r\cos(\alpha + t) + \delta\cos(\alpha - t) \\ c_2 + r\sin(\alpha + t) + \delta\sin(\alpha - t) \end{bmatrix}$$

where t ranges from 0 to 2π. Through the orbit, the satellite will assist scientists observe some asteroids around the moon. Therefore, Algae should plan an orbit that minimize the distance between asteroids and the orbit. The parameters we mention below is $(c_1, c_2, r, \delta, \alpha)$. Hint. All of the parameters should take value between 0 to 2. You can eliminate those that are out of this range first.

(a) Given the coordinate of the asteroids in "orbit.csv". Apply the Levenberg-Marquardt algorithm to fit the data "X" and "Y" and find the orbit. Try at least 100 initial setting (you can generate it randomly, don't forget to set seed before you generate the random initial values), then pick up reasonable parameter bundles. Simply explain how you "reasonably" pick up the bundles and report the result of your bundle (take average to each parameter if there are several bundles). (Please round your answer to 4 digits)

(b) Plot the orbit you found in (a) and mark the location for each asteroid (set the color to red) and the corresponding nearest points on Algae's orbit (set the color to blue).

(c) However, due to universal gravitation, if the mass of the asteroid is large, then the satellite should not be too close to the it. Otherwise, the satellite will be drag out of the orbit and collide into the asteroids. Therefore, Algae will take the mass of asteroid into account when she plans the orbit, that is, instead of using least square estimation, she will use weighted least square estimation. Setting the weight for each asteroid to $1/\mathrm{Mass}$, $1/\mathrm{Mass}^2$, Mass, Mass^2. Report the parameters for different weight respectively. (Please round your answer to 4 digits)

(d) Plot the orbits in (a) and (b) with labels that indicate the weight. Moreover, add scatters(markers) to mark the location of each asteroid and adjust the size of the scatters(markers) according to the mass of asteroids ($\mathrm{size} = \mathrm{Mass} \times 5$).

3. (25%) We consider a constrained nonlinear least squares problem with three variables $x = (x_1, x_2, x_3)$ and two equations:

$$\begin{aligned} \min \quad & (x_1 - 1)^2 + (x_2 - 2)^2 + (x_3 - 3)^2 \\ \text{s.t.} \quad & x_1^2 + 0.5x_2^2 + x_3^2 - 3 = 0 \\ & 0.75x_1^2 + 2.5x_2^2 + x_3^2 + 2x_1x_3 - x_1 - x_2 - x_3 - 1 = 0 \end{aligned}$$

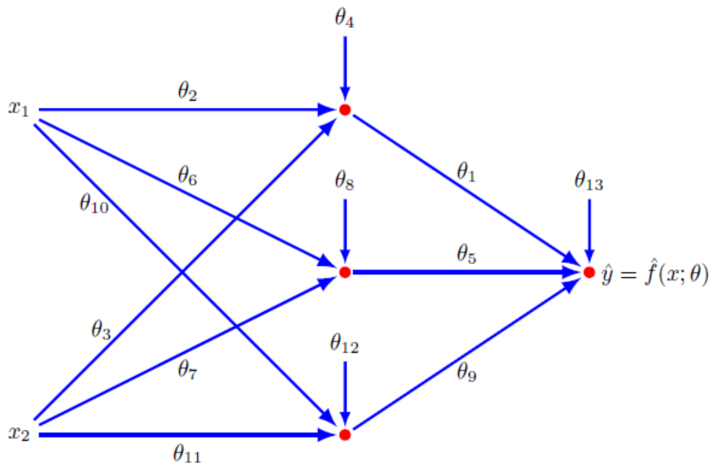
Please solve the question with mentioned algorithms below with the stopping condition of both feasibility residual and optimality condition residual smaller than 10^{-5}.

(a) Solve the problem using the augmented Lagrangian method.
(b) Solve the problem using the penalty method.
(c) Make the plots of two residuals and of the penalty parameter versus the cumulative number of Levenberg-Marquardt iterations for (a) and (b).

4. (20%) A neural network is a widely used model of the form $\hat{y} = \hat{f}(x;\theta)$, where the *n*-vector *x* is the feature vector and the *p*-vector θ is the model parameter. In a neural network model, the function $\hat{f}$ is not an affine function of the parameter θ. We now consider two layers, three internal nodes and two inputs in our neural network model, that is, 13 parameters should be evaluated and is given by

$$\hat{f}(x;\theta) = \theta_1\varphi(\theta_2 x_1 + \theta_3 x_2 + \theta_4) + \theta_5\varphi(\theta_6 x_1 + \theta_7 x_2 + \theta_8) + \\ + \theta_9\varphi(\theta_{10} x_1 + \theta_{11} x_2 + \theta_{12}) + \theta_{13}$$

where $\varphi: \mathrm{R} \to \mathrm{R}$, is a sigmoid function. The function is shown as a signal flow graph in the below graph. In this graph each edge from an input to an internal node, or from an internal node to the output node, corresponds to multiplication by one of the parameters. At each node (shown as the small filled circles) the incoming values and the constant offset are added together, then passed through the sigmoid function, to become the outgoing edge value.



(a) Import the data given in "neural_data.csv". Use the Levenberg-Marquardt algorithm to try to minimize

$$f(\theta) = ||r(\theta)||^2 + \gamma||\theta||^2$$

with $\mathrm{r}(\theta)_i = \hat{f}(x^{(i)}; \theta) - \mathrm{y}^{(i)}$, $\gamma = 10^{-5}$. Plot the value of f and the norm of its gradient versus iteration. Report the RMS fitting error achieved by the neural network model.

(b) Experiment with choosing different starting points to see the effect on the final model found in (a). (At least try 3 different starting points)

(c) Fit the dataset with a regression model $f^{\mathrm{reg}}(x; \beta, v) = x^T\beta + v$ and report the RMS fitting error achieved. Then, compare the result with the result you found in (b).

Part 2 Mathematical Exercises (0%)

Chapter 16 Exercises

1. Smallest right inverse. Suppose the $m \times n$ matrix A is wide, with linearly independent rows. Its pseudo-inverse $A^\dagger$ is a right inverse of A. In fact, there are many right inverses of A and it turns out that $A^\dagger$ is the smallest one among them, as measured by the matrix norm. In other words, if X satisfies $AX = I$, then $\|X\| \geq \|A^\dagger\|$. You will show this in this problem.

(a) Suppose $AX = I$, and let $x_1, \cdots, x_m$ denote the columns of X. Let b_j denote the jth column of $A^\dagger$. Explain why $\|x_j\|^2 \geq \|b_j\|^2$. Hint. Show that $z = b_j$ is the vector of smallest norm that satisfies $Az = e_j$ for $j = 1, \cdots, m$.

(b) Use the inequalities from part (a) to establish $\|X\| \geq \|A^\dagger\|$.

2. Matrix least norm problem. The matrix least norm problem is

$$\begin{array}{ll} \text{minimize} & \|X\|^2 \\ \text{subject to} & CX = D, \end{array}$$

where the variable to be chosen is the $n \times k$ matrix X; the $p \times n$ matrix C and the $p \times k$ matrix D are given. Show that the solution of this problem is $\hat{X}$=$C^\dagger D$, assuming the rows of C are linearly independent. Hint. Show that we can find the columns of X independently, by solving a least norm problem for each one.

Chapter 17 Exercises

1. A variation on the portfolio optimization formulation. Show that the two problem formulations are the equivalent.

Formulation (A).

$$\text{minimize} \qquad ||Rw||^2$$

$$\text{subject to} \qquad \begin{bmatrix} \mathbf{1}^T \\ \mu^T \end{bmatrix} w = \begin{bmatrix} 1 \\ \rho \end{bmatrix}.$$

Formulation (B).

$$\begin{aligned} &\text{minimize} \qquad && w^T \Sigma w \\ &\text{subject to} \qquad && \begin{bmatrix} \mathbf{1}^T \\ \mu^T \end{bmatrix} w = \begin{bmatrix} 1 \\ \rho \end{bmatrix}. \end{aligned}$$

2. A simple portfolio optimization problem.

(a) Find an analytical solution for the portfolio optimization problem with $n = 2$ assets. You can assume that $\mu_1 \neq \mu_2$, i.e., the two assets have different mean returns. Hint. The optimal weights depend only on μ and ρ, and not (directly) on the return matrix R.

(b) Find the conditions under which the optimal portfolio takes long positions in both assets, a short position in one and a long position in the other, or a short position in both assets. You can assume that $\mu_1 < \mu_2$, i.e., asset 2 has the higher return. Hint. Your answer should depend on whether $\rho < \mu_1$, $\mu_1 < \rho < \mu_2$, or $\mu_2 < \rho$, i.e., how the required return compares to the two asset returns.

Chapter 18 Exercises

1. A common form for the residual. In many nonlinear least-squares problems, the residual function $f: \mathbb{R}^n \to \mathbb{R}^m$ has the specific form

$$f_i(x) = \phi_i(a_i^T x - b_i), i = 1, \cdots m,$$

where a_i is an n-vector, b_i is a scalar, and ϕ_i: $\mathbb{R} \to \mathbb{R}$ is a scalar-valued function of a scalar. In other words, $f_i(x)$ is a scalar function of an affine function of x. In this case, the objective of the nonlinear least-squares problem has the form

$$\|f(x)\|^2 = \sum_{i=1}^{m} (\phi_i(a_i^T x - b_i)\ )^2.$$

We define the $m \times n$ matrix A to have rows $a_1^T, \cdots, a_m^T$, and the m-vector b to have entries $b_1, \cdots, b_m$. Note that if the functions ϕ_i are the identity function, i.e., $\phi_i(u) = u$ for all u, then the objective becomes $\|Ax - b\|^2$, and in this case the nonlinear least squares problem reduces to the linear least squares problem. Show that the derivative matrix $Df(x)$ has the form

$$Df(x) = \mathbf{diag}(d)\mathrm{A},$$

where $d_i = \phi_i'(r_i)$ for $i = 1, \cdots, m$, with $r = Ax - b$.