New York University Shanghai

Problem Set 12

Solutions

BUSF-SHU 210: Business Analytics (Spring 2019)

1. Pricing Strategy of Starbucks

Starbucks has offered two coffee choices: Latte and Cappuccino. Assume Starbucks charges p_l for Latte and p_c for Cappuccino (we assume Starbucks only has one cup size). Starbucks estimated that the hourly demand for Latte would be

$$\frac{1500\exp(25-p_l)}{1+\exp(25-p_l)+\exp(28-p_c)}$$

And the hourly demand for Cappuccino would be

$$\frac{1500\exp(28-p_c)}{1+\exp(25-p_l)+\exp(28-p_c)}$$

This demand model is called the Multi-Nomial Logit (MNL) discrete choice model, which is very widely used to analyze customer behaviors and optimize pricing and assortment strategies.

Assume that the cost of making one Latte is 10 RMB and that of making one Cappuccino is 13 RMB. You goal is to help Starbucks optimize their pricing strategy so as to maximize the total hourly profit.

(a) Formulate the pricing problem of Starbucks as a non-linear programming model. Solve the optimization model in Python. What is the optimal pricing strategy (p_l^*, p_c^*) that maximizes the total hourly profit of Starbucks?

The decision variables are (p_l, p_c). The objective function is the hourly profit of Starbucks:

$$\pi(p_l,p_c) = (p_l-10)\times\frac{1500\exp(25-p_l)}{1+\exp(25-p_l)+\exp(28-p_c)} + (p_c-13)\times\frac{1500\exp(28-p_c)}{1+\exp(25-p_l)+\exp(28-p_c)}.$$

The constraints are that $p_l \geq 10$ and $p_c \geq 13$. We solve this non-linear program using Python.

(b) For some marketing concerns, Starbucks do not want the price difference between Latte and Cappuccino to be larger than 2 RMB. What is the optimal pricing strategy $(\hat{p}_l^*, \hat{p}_c^*)$ that maximizes the total hourly profit of Starbucks under this constraint?

The decision variables and the objective function are the same as those in part (a). The constraints should be adjusted to that $p_l \geq 10$, $p_c \geq 13$, $p_l - p_c - 2 \leq 0$ and $p_c - p_l - 2 \leq 0$.

2. Convex Functions and Non-Linear Programming

Please briefly answer the following questions:

(a) Plot the figure of $h(x) = \frac{1}{1+\exp(-x)}$. Is $h(x)$ a convex or concave function when (i) $x \leq 0$ and (ii) $x \geq 0$?

We calculate the second-order derivative of $h(x)$:

$$h''(x) = \frac{\exp(-x)(\exp(-x) - 1)}{(1 + \exp(-x))^3}$$

Hence, $h''(x) \geq 0$ for $x \leq 0$ and $h''(x) \leq 0$ for $x \geq 0$. Thus, $h(x)$ is convex for $x \leq 0$, and $h(x)$ is concave for $x \geq 0$. The plot is given in Figure 1.

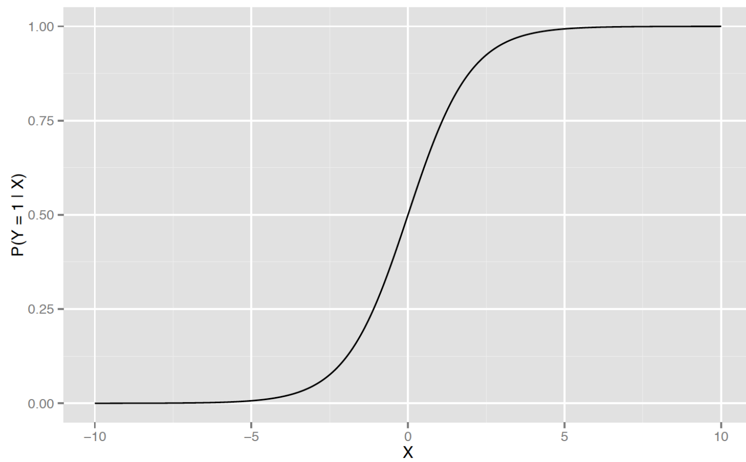

Figure 1: Logistic Function

(b) Assume that $f(x)$ is a uni-variable convex function from $\mathbb{R}$ to $\mathbb{R}$. Let $x_1 = 1$ and $x_2 = 2$ be two different local minima: $f(x) \geq f(x_1)$ for all $|x - x_1| \leq 0.1$; and $f(x) \geq f(x_2)$ for all $|x - x_2| \leq 0.1$. Show that $f(x_1) = f(x_2)$ and $f(x) = f(x_1) = f(x_2)$ for all $x \in [x_1, x_2]$. This actually implies that x_1 and x_2 are global minima of the function $f(\cdot)$.

Without loss of generality, assume that $f(x_1) \leq f(x_2)$. Define $x_3 = x_2 - 0.1 = 1.9 = 0.1x_1 + 0.9x_2$. Since $f(x)$ is convex,

$$f(x_3) = f(0.1x_1 + 0.9x_2) \leq 0.1f(x_1) + 0.9f(x_2) \leq f(x_2),$$

where the first inequality follows from the convexity of $f(x)$ and the second from that $f(x_1) \leq f(x_2)$. Since x_2 is a local minimum, $f(x_3) \geq f(x_2)$, which implies that $f(x_3) = 0.1f(x_1) + 0.9f(x_2) = f(x_2)$, i.e., $f(x_3) = f(x_2) = f(x_1)$. Therefore, for any $x \in [x_1, x_2]$, $x = \lambda x_1 + (1-\lambda)x_2$ for some $\lambda \in [0, 1]$, $f(x) = f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) = f(x_1) = f(x_2)$.

On the other hand, if $f(x) < f(x_1)$ for some $x \in (x_1, x_2)$, then for all $y \in (x_1, x)$, $y = \lambda x_1 + (1-\lambda)x$ for some $\lambda \in (0, 1)$. We have $f(y) = f(\lambda x_1 + (1-\lambda)x) \le \lambda f(x_1) + (1-\lambda) f(x) < f(x_1)$, where the the first inequality follows from the convexity of $f(\cdot)$, and the second from $f(x) < f(x_1)$. On the other hand, x_1 is a local minimum, so $f(x_1) \ge f(y)$ for $y \le 1.1 = x_1 + 0.1$. This forms a contradiction. Hence, $f(x) \ge f(x_1)$ for all $x \in (x_1, x_2)$. Since we have also shown that $f(x) \le f(x_1)$ for all $x \in (x_1, x_2)$, we have that $f(x) = f(x_1)$ for all $x \in (x_1, x_2)$. Therefore, $f(x_1) = f(x_2) = f(x)$ for all $x \in [x_1, x_2]$.

(c) Consider the same 5 stocks and the bond as in Session 12. The mean annual return and the standard deviation of each stock are summarized in the following table:

Index	Stock	Expected Annual Return	Standard Deviation of Annual Return
1	AAPL	0.114	0.039
2	AMZN	0.103	0.030
3	DIS	0.092	0.032
4	WFM	0.085	0.029
5	WMT	0.078	0.022
0	Bond	0.05	0

The correlations between the assets are shown in the following table:

	AAPL	AMZN	DIS	WFM	WMT	Bond
AAPL	1	0.160	0.163	-0.260	0.399	0
AMZN	0.160	1	0.029	0.272	-0.193	0
DIS	0.163	0.029	1	0.173	0.124	0
WFM	-0.260	0.272	0.173	1	0.125	0
WMT	0.399	-0.193	0.124	0.125	1	0
Bond	0	0	0	0	0	1

For a portfolio $x = (x_1, x_2, x_3, x_4, x_5, x_6)$, Ms. Liu requests that the mean annual return is at least 0.095. What is the portfolio x^* that minimizes the variance of the return given that the mean annual return is at least 0.095? To normalize, we set an additional constraint that $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 1$.

The variance of return for a portfolio x is $\sum_{i=0}^{5} \sum_{i=0}^{5} \sigma_i \sigma_j \rho_{ij} x_i x_j$, whereas the mean return is $\sum_{i=0}^{5} \mu_i x_i$. Therefore, we can formulate the portfolio selection problem as:

$$\min \sum_{i=0}^{5} \sum_{i=0}^{5} \sigma_i \sigma_j \rho_{ij} x_i x_j$$

Subject to

$$\sum_{i=0}^{5} \mu_i x_i \ge 0.095$$

$$\sum_{i=0}^{5} x_i = 1$$

Using Python, we find that the variance-minimization portfolio is $x^* = (0.173, 0.301, 0.150, 0.166, 0.210, 0.00)$.

3. Maximum Likelihood Estimation

In logistic regression, we use the maximum likelihood estimation (MLE) to train the coefficients $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)$. Specifically, given the data $\mathcal{D} = \{Y_i \in \{0,1\}, X_{ij} : 1 \leq i \leq n, 1 \leq j \leq p\}$, we define the likelihood function as the probability of the outcomes $(Y_1, Y_2, \cdots, Y_n)$ conditioned on the covariates $\{X_{ij}, 1 \leq i \leq n, 1 \leq j \leq p\}$. Conditioned on $X_i = (X_{i1}, X_{i2}, \cdots, X_{ip})$, the probability of the outcome $Y_i = 1$ is

$$\mathbb{P}(Y_i = 1|X_i) = \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}$$

And the probability of the outcome $Y_i = 0$ is

$$\mathbb{P}(Y_i = 0|X_i) = \frac{1}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}$$

We define the likelihood function of data point i is

$$l(Y_i|X_i, \beta) = \begin{cases} \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}, & \text{if } Y_i = 1; \\ \frac{1}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}, & \text{if } Y_i = 0 \end{cases}$$

We can write the likelihood function of data point i in a more compact form as

$$l(Y_i|X_i, \beta) = \left(\frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})} \right)^{Y_i} \left(\frac{1}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})} \right)^{1-Y_i}$$

The likelihood of the whole training data set can be written as

$$L(Y|X, \beta) = l(Y_1|X_1, \beta) \times l(Y_2|X_2, \beta) \times \cdots \times l(Y_n|X_n, \beta)$$

The estimated coefficients $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)$ are chosen to maximize the likelihood of the whole training data set $\mathcal{D}$, i.e.,

$$\hat{\beta} = \text{argmax}_\beta L(Y|X, \beta)$$

The likelihood itself is difficult to compute and to maximize. Instead, we focus on maximizing the log-likelihood function of the training data set, i.e.,

$$\hat{\beta} = \text{argmax}_\beta \log(L(Y|X, \beta))$$

This question asks you to formulate a non-linear programming model and estimate the logistic regression model using the data set `quality.csv`, which we have used in the lecture of Session 3. This data set has a lot of variables, but we only use $PoorCare \in \{0, 1\}$ (which indicates whether the patient receives poor care; $PoorCare = 1$ means poor care whereas $PoorCare = 0$ means good care) as the outcome, and $OfficeVisits$ (which refers to the total number of office visits to the healthcare provider) and $Narcotics$ (which refers to the amount of narcotics this patient has taken) as the covariates. Our goal is to use MLE to fit a logistic regression model. Please use the entire data set as the training data.

Please briefly answer the following questions:

(a) Given the data set $\mathcal{D}$, calculate the log-likelihood $\log(L(Y|X))$. Write the log-likelihood function in Python.

It can be calculated that

$$
\begin{aligned}
\log(L(Y|X,\beta)) &= \sum_{i=1}^{n} \log(l(Y_i|X_i,\beta)) \\
&= \sum_{i=1}^{n} Y_i \log\left(\frac{\exp(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij})}{1+\exp(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij})}\right) \\
&\quad + \sum_{i=1}^{n} (1-Y_i) \log\left(\frac{1}{1+\exp(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij})}\right)
\end{aligned}
$$

(b) Find the coefficient $\hat{\beta}$ that maximizes the log-likelihood $\log(L(Y|X))$. If we have a new patient for whom the total number of office visits is 17, and the amount of Narcotics is 4. What is the probability that this patient receives good healthcare quality based on the logistic regression model you trained?

We fit the logistic regression model and find that $\hat{\beta}_0 = -2.54021$, $\hat{\beta}_{OfficeVisits} = 0.06273$, and $\hat{\beta}_{Narcotics} = 0.10990$. Plugging in that the total number of office visits is 17, and the amount of Narcotics is 4, we have the probability that this patient receives good healthcare quality is

$$\frac{1}{1+\exp(-2.54021 + 0.06273 \times 17 + 0.10990 \times 4)} = 0.7377$$

<u>*Hint*</u>: Recall the property of the $\log(\cdot)$ function: $\log(ab) = \log(a) + \log(b)$ and $\log(a^b) = b\log(a)$.