Lagrange multiplier and KKT condition

Hung-Hsuan Chen

Toy example

- Maximize $5x_1x_2$ subject to $2x_1 + x_2 = 100$
- Sol 1

  $$f\left(x_1, x_2\right) = 5x_1x_2 = 5x_1\left(100 - 2x_1\right)$$

  $$\boxed{?}\,\frac{\partial f(x_1, x_2)}{\partial x_1} = 500 - 20x_1 := 0$$

  ➔$(x_1, x_2) = (25, 50)$

  ➔The maximum value of $5x_1x_2$ is 6250
- If the constraints are not complicated, such a method is manageable

Lagrange multiplier

- Lagrange multipliers is a strategy for finding the extreme value of a function ***subject to equality constraints***
- Finding the maximum/minimum value of $y = f(\mathbf{x})$ subject to $g_i(\mathbf{x}) = 0,\ i = 1,2,\ldots,m$
  – Lagrange function (a.k.a. Lagrangian)
    - $y_\lambda = f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \lambda_2 g_2(\mathbf{x}) + \ldots + \lambda_m g_m(\mathbf{x})$

  – λ_i's are called "Lagrange multipliers"

Solving the problem

- Lagrange function

$$\mathscr{L}\left(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m\right) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

  - Some textbook may write $f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$
    - This is also correct, since the signs can be "absorbed" by each $g_i(\mathbf{x})$
- Take the derivative of the Lagrange function to every variable (i.e., all the *x*'s and the λ's)

$$\Rightarrow \begin{cases} \dfrac{\partial \mathscr{L}(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m)}{\partial x_j} = 0 \ \forall j \\ \dfrac{\partial \mathscr{L}\left(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m\right)}{\partial \lambda_i} = 0 \ \forall i \end{cases}$$

- Test each solution set. Whichever gives the greatest (or the smallest) value is the maximum (or minimum) point
  - Lagrangian is a necessary but ***not a sufficient*** condition

Solving the toy example by Lagrange multiplier

- Maximize $5x_1x_2$ subject to $2x_1 + x_2 = 100$

$$f(x_1, x_2) = 5x_1x_2,\ g(x_1, x_2) = 2x_1 + x_2 - 100$$

$$\mathscr{L}(x_1, x_2, \lambda_1) = 5x_1x_2 + \lambda_1(2x_1 + x_2 - 100)$$

$$\begin{cases} \dfrac{\partial \mathscr{L}(x_1, x_2, \lambda_1)}{\partial x_1} = 5x_2 + 2\lambda_1 := 0 \\ \dfrac{\partial \mathscr{L}(x_1, x_2, \lambda_1)}{\partial x_2} = 5x_1 + \lambda_1 := 0 \\ \dfrac{\partial \mathscr{L}(x_1, x_2, \lambda_1)}{\partial \lambda_1} = 2x_1 + x_2 - 100 := 0 \end{cases}$$

こここ 解联立

$$\Rightarrow (x_1, x_2, \lambda_1) = \underline{(25, 50, -125)}$$

Generalized Lagrange multiplier

- Lagrange multipliers is generalized to include the ***inequality constraints*** under the Karush–Kuhn–Tucker (KKT) condition
- Standard form problem

  Minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \leq 0 \ (i = 1, \ldots, p)$ and $h_j(\mathbf{x}) = 0 \ \left(j = 1, \ldots, m\right)$

  - If the task is to maximize $f(\mathbf{x})$, transform the problem into minimize $-f(\mathbf{x})$
- Lagrangian

$$\mathscr{L}\left(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}\right) = f(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{m} \mu_j h_j(\mathbf{x})$$

Necessary (not sufficient!) optimal condition (KKT condition)

what is KKT

- If $\mathbf{x}^*$ is the optimal solution to the standard form problem, then there exist KKT multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ such that
  - Lagrangian optimality

$$\nabla \mathscr{L}\left(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}\right) = 0 ------ \quad (1)$$

  - Primal feasibility

$$g_i\left(\mathbf{x}^*\right) \leq 0 \ \forall i ------- \quad (2)$$

$$h_j(\mathbf{x}^*) = 0 \ \forall j ------- \quad (3)$$

  - **Dual feasibility**

$$\lambda_i \geq 0 \ \forall i ---------- \quad (4)$$

  - **Complementary slackness**

$$\lambda_i g_i\left(\mathbf{x}^*\right) = 0 \ \forall i ------ \quad (5)$$

Example

改成 ≤ 0

- Minimize $f(x_1, x_2) = x_1^2 + 2x_2^2$ subject to $x_1 + x_2 \geq 3$ and $x_2 - x_1^2 \geq 1$

  $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = x_1^2 + 2x_2^2 + \lambda_1(3 - x_1 - x_2) + \lambda_2(1 + x_1^2 - x_2)$

  $$\Rightarrow \begin{cases} \dfrac{\partial \mathcal{L}}{\partial x_1} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial x_2} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_1} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_2} = 0 \end{cases} \Rightarrow (x_1, x_2, \lambda_1, \lambda_2) = (-2, 5, 12, -8) \text{ or } (1, 2, 6, 2)$$

  ???

  $\Rightarrow (-2, 5, 12, -8)$ does not follow the KKT condition (4)

  $(\because \lambda_2 = -8 < 0)$

  $\Rightarrow \min f(x_1, x_2) = f(1,2) = 1^2 + 2 \cdot 2^2 = 9$

  - BTW, $(1, 2, 6, 2)$ also follows KKT condition (5) $\lambda_i g_i(\mathbf{x}^*) = 0$

    Check: $6(3 - 1 - 2) = 0; 2(1 + 1^2 - 2) = 0$

    $\lambda_1 \times g_1(\ )$ $\lambda_2 \times g_2(\ )$

Summary

- Lagrangian as a tool to find extreme values of a function with equality constraints
- Genralized Lagrangian and KKT condition together as a tool set to find extreme values of a function with equality and inequality constraints
- We will use KKT conditions to analyze the properties of SVM solutions

Quiz: find p_is to maximize entropy

- Suppose the probability of the events (a_1, a_2, a_3) are (p_1, p_2, p_3) respectively
  - $p_1 + p_2 + p_3 = 1$
- Entropy
  - $$\sum_{i=1}^{n} p_i \log_2(\frac{1}{p_i}) = \sum_{i=1}^{n} (-p_i \log_2(p_i))$$
- What are the values of p_i to maximize entropy?

Quiz: find p_is to maximize entropy

Maximize $-p_1\log(p_1)-p_2\log(p_2)-p_3\log(p_3)$ subject to $p_1+p_2+p_3=1$

$f(p_1,p_2,p_3)=-p_1\log(p_1)-p_2\log(p_2)-p_3\log(p_3)$

$g(p_1,p_2,p_3)=1-p_1-p_2-p_3$

$\mathscr{L}(p_1,p_2,p_3,\lambda)=-p_1\log(p_1)-p_2\log(p_2)-p_3\log(p_3)+\lambda(1-p_1-P_2-p_3)$

$$\Rightarrow \begin{cases} \frac{\partial\mathscr{L}}{\partial p_1}=0 \\ \frac{\partial\mathscr{L}}{\partial p_2}=0 \\ \frac{\partial\mathscr{L}}{\partial p_1}=0 \\ \frac{\partial\mathscr{L}}{\partial \lambda}=0 \end{cases} \Rightarrow \begin{cases} -\log(p_1)-1-\lambda=0 \\ -\log(p_2)-1-\lambda=0 \\ -\log(p_3)-1-\lambda=0 \\ 1-p_1-p_2-p_3=0 \end{cases} \Rightarrow \begin{cases} p_1=1/3 \\ p_2=1/3 \\ p_3=1/3 \\ \lambda=\log(3)-1 \end{cases}$$