BITS F464: Machine Learning

07 SSD Lagrange Multiplier

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

Feb 01, 2021 ONLINE (Campus @ BITS-Pilani Jan-May 2021)

http://ktiwari.in/ml

Max and Argmax

x	$f(x)$
1	7
2	11
3	6
4	12
5	7
6	8
7	4
8	12
9	7
10	13
11	15
12	6
13	9

$$\begin{aligned} max(f(x)) &= 15 \\ \underset{x}{\text{argmax}}(f(x)) &= 11 \\ &= \underset{x}{\text{argmax}}(20 + f(x)) \\ &= \underset{x}{\text{argmax}}(20 \times f(x)) \\ &= \underset{x}{\text{argmax}}(\log f(x)) \end{aligned}$$

$$min(f(x)) = 4$$
$$\underset{x}{\text{argmax}}(-f(x)) = 7 = \underset{x}{\text{argmin}}(f(x))$$
$$\underset{x}{\text{argmax}}(20 - f(x)) = 7 = \underset{x}{\text{argmin}}(f(x))$$
$$\underset{x}{\text{argmin}}(-f(x)) = \underset{x}{\text{argmax}}(f(x)) = 11$$

Maximum Likelihood and Least Squared Error (LSE)

Under certain assumptions, a learning algorithm minimizing Least Squared Error, will output Maximum Likelihood hypothesis h_{ML}

- Assume noise is random and obeys normal distribution with zero mean and variance σ^2 which is independent for every data point

$$h_{ML} = \underset{h \in H}{\text{argmax}}\, P(D|h)$$

- $D =< (x_1, y_1), (x_2, y_2), ..., (x_m, y_m) >$ with $y_i = f(x_i) + e_i$. Assuming training data being mutually independent given h

$$h_{ML} = \underset{h \in H}{\text{argmax}}\, \Pi_{i=1}^{m} p(y_i|h)$$

- Data has variance σ^2 and mean around true target value $\mu = f(x_i)$. Therefore, $p(y_i|h)$ being a normal distribution would also have a variance σ^2 and mean μ

Maximum Likelihood and LSE (contd...)

$$h_{ML} = \underset{h \in H}{\text{argmax}}\, \Pi_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}$$

As we assume h to be correct description of f so $\mu = f(x_i) = h(x_i)$

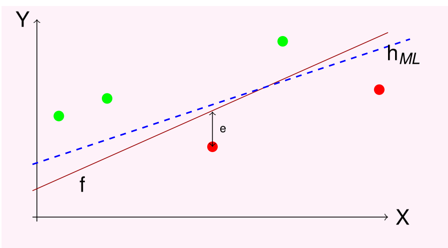
$$h_{ML} = \underset{h \in H}{\text{argmax}}\, \Pi_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-h(x_i))^2}$$

Being continuous function, we may choose its log to optimize

$$\begin{aligned} h_{ML} &= \underset{h \in H}{\text{argmax}} \sum_{i=1}^{m} \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(y_i - h(x_i))^2 \\ &= \underset{h \in H}{\text{argmax}} \sum_{i=1}^{m} -\frac{1}{2\sigma^2}(y_i - h(x_i))^2 = \underset{h \in H}{\text{argmin}} \sum_{i=1}^{m} (y_i - h(x_i))^2 \end{aligned}$$

As the first term being constant.

SSD gives Maximum likelihood hypothesis



$$h_{ML} = \underset{h \in H}{\text{argmin}} \sum_{i=1}^{m} (y_i - h(x_i))^2$$

Multivariate Optimization

Lagrange Multiplier: Used to find local optima (maxima or minima) of some objective function f, subject to equality constraint on g. *Example:* maximize $f(x, y)$ subject to $g(x, y) = c$

- Essentially we want a point where gradient of f and g are scaler multiple $\triangledown f = \lambda \triangledown g$ that is $\triangledown f - \lambda \triangledown g = 0$

Example-01: Let cost of producing x amount of product A and y amount of product B is $6x^2 + 12y^2$ and we want at least 90 items to be produced. Find optimal value of x and y

Lagrange Equation for the problem is $F = 6x^2 + 12y^2 - \lambda(x + y - 90)$

$$\frac{\partial}{\partial x}(F) = 12x - \lambda \quad \frac{\partial}{\partial y}(F) = 24y - \lambda \quad \frac{\partial}{\partial \lambda}(F) = x + y - 90$$

Solve above equations by equating to zero & get $\boxed{x = 60}$ and $\boxed{y = 30}$

Example: Lagrange Multiplier

Example-02

Optimize $f(x, y, z) = x^2 + x + 2y^2 + 3z^2$ on constraint $x^2 + y^2 + z^2 = 1$

Lagrange Equation: $F = x^2 + x + 2y^2 + 3z^2 - \lambda(x^2 + y^2 + z^2 - 1)$

$\frac{\partial F}{\partial x} = 2x + 1 - 2\lambda x$ $\frac{\partial F}{\partial y} = 4y - 2\lambda y$ $\frac{\partial F}{\partial z} = 6z - 2\lambda z$ $\frac{\partial F}{\partial \lambda} = x^2 + y^2 + z^2 - 1$

$4y - 2\lambda y = 0$ gives $\rightarrow$ either $y = 0$ or $\lambda = 2$

$6z - 2\lambda z = 0$ gives $\rightarrow$ either $z = 0$ or $\lambda = 3$

- Case-1: $y = z = 0 \rightarrow x = \pm 1$
- Case-2: $y = 0, \lambda = 3 \rightarrow x = 1/4, z = \pm\frac{\sqrt{15}}{4}$
- Case-3: $\lambda = 2, z = 0 \rightarrow x = 1/2, y = \pm\frac{\sqrt{3}}{2}$ note[1]

Therefore, max is 25/8 and min is 0 at $(\frac{1}{4}, 0, \frac{\sqrt{15}}{4})$ and $(-1, 0, 0)$

[1] Case-1 gets rise of two points $p1 = (1, 0, 0)$ and $p2 = (-1, 0, 0)$. Case-2 to $p3 = (1/4, 0, \frac{\sqrt{15}}{4})$ and $p4 = (1/4, 0, -\frac{\sqrt{15}}{4})$. Case-3 for $p5 = (1/2, \frac{\sqrt{3}}{2}, 0)$ and $p6 = (1/2, -\frac{\sqrt{3}}{2}, 0)$. We can evaluate $f(p1) = 2$, $f(p2) = 0$, $f(p3) = 25/8$, $f(p4) = 25/8$, $f(p5) = 9/4$, $f(p6) = 9/4$.

Thank You!

Thank you very much for your attention!

Queries ?