Lagrange multipliers:

Joseph Louis Lagrange, invented a strategy for finding the local maxima and minima of a function subject to equality constraint.

The method of Lagrange multipliers

Lagrange noticed that when we try to slove optimization problem of the form:

$$\underset{x}{\text{minimize}}\ f(x)$$

$$\text{subject to } g(x) = 0$$

the minimum of f is found when its gradient point in the same direction as the gradient of g. In other words:

$$\nabla f(x) = \alpha \nabla g(x)$$

So if we want to find the minimum of f under the constraint g, we just need to slove for: $\nabla f(x) - \underbrace{\alpha}_{\text{Lagrange multiplier}} \nabla g(x) = 0$

To simplify the method:

$$\nabla L(x, \alpha) = \nabla f(x) - \alpha \nabla g(x)$$

follow these steps:

1. Construct the Lagraugian function L by introducing one multiplier per constraint
2. Get the gradient ∇L of the Lagragian
3. Slove for $\nabla L(x, \alpha) = 0$,,

$$\boxed{\nabla = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k}}$$

The SVM Lagrangian problem:

Recall:- our goal is to maximize $f(w) = \frac{2}{||w||}$

instead let's minimize $\frac{1}{f(w)} = \frac{||w||}{2}$ bcz

we can use lagrangian to minimize it right!

New goal:- $\underset{w, d}{\text{Minimize}} \; \frac{||w||^2}{2}$ why?

subject to $y_i[(w \cdot x_i) + d] - 1 \geq 0$

Squaring the norm has the advantage of removing the square root $(\because ||w||^2 = \langle w, w \rangle)$

The SVM Lagrangian problem:-

$$f_{new}(w) = \frac{1}{2}\|w\|^2$$

$$g_i(w,d) = y_i(w \cdot x_i + d) - 1, \quad i = 1, \ldots, m$$

$$L(w,d,\alpha) = f(w) - \sum_{i=1}^{m} \alpha_i g_i(w,d)$$

$$L(w,d,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[y_i(w \cdot x_i + d) - 1\right]$$

<u>Note</u>: We introduced one Lagrange multiplier α_i for each constraint function.

<u>Imp</u>: you may have noticed that the method of Lagrange multipliers is used for solving problems with <u>equality constraints</u>, and here we are using them with equality constraints. This is because the method still works for inequality constraints, provided KKT conditions are met

Karush-Kuhn-Tucker (KKT) conditions

Because we are dealing with inequality $cond^n$, there is an additional requirement: the Solution must also satisfy the <u>KKT</u> conditions

Lets understand with Example

$$\text{Min } f(X_1, X_2) = -x_1^2 - x_2^2 - x_3^2 + 4x_1 + 6x_2$$

STC $x_1 + x_2 \leq 2$ — h_1

$2x_1 + 3x_2 \leq 12$ — h_2

$x_1, x_2 \geq 0$

KKT conditions:

1. Convert to lagrange function and do partial differentation wrt variables and Equate to zero
2. $\lambda_i h_i = 0$
3. $h_i \leq 0$
4. $\lambda_i \geq 0$

condition-1:

$$L(x_1, x_2, x_3, \lambda_1, \lambda_2) = -x_1^2 - x_2^2 - x_3^2 + 4x_1 + 6x_2 - \lambda_1(x_1 + x_2 - 2) - \lambda_2(2x_1 + 3x_2 - 12)$$

$$\frac{\partial L}{\partial x_1} = -2x_1 + 4 + \lambda_1 - 2\lambda_2 = 0 \quad -1a$$

$$\frac{\partial L}{\partial x_2} = -2x_2 + 6 + \lambda_1 - 3\lambda_2 = 0 \quad -1b$$

$$\frac{\partial L}{\partial x_3} = -2x_3 = 0 \Rightarrow \boxed{x_3 = 0}$$

condition-2:

$$\lambda_1(x_1 + x_2 - 2) = 0 \quad -2a \quad \& \quad \lambda_2(2x_1 + 3x_2 - 12) = 0 - 2b$$

condition-3

$x_1+x_2-2 \leq 0$ — 3a

$2x_1+3x_2-12 \leq 0$ — 3b

condition-4

$\lambda_1 \geq 0$ & $\lambda_2 \geq 0$

Case-I: $\lambda_1=0$ $\lambda_2=0$ - substitute in 1a & 1b

and slove we get $x_1=2$ & $x_2=3$

Now substitul x_1, x_2 in 3a & 3b

$3a \rightarrow x_1+x_2-2 \leq 0$

$2+3-2 \leq 0 \Rightarrow \boxed{3 \leq 0}$ ✗

$3b \rightarrow 2x_1+3x_2-12 \leq 0$

$1 \leq 0$ ✗

Case-2: $\lambda_1 \neq 0$ $\lambda_2 \neq 0$

$2a \rightarrow \lambda_1(x_1+x_2-2)=0$ — (i)

$\neq 0 \Rightarrow$ Mathematicians found that both

λ_1 & x_1+x_2-2 will not become $=0$ simaultaneously

$\Rightarrow x_1+x_2-2=0$ (i)

$||^{ly}$ from 2b $2x_1+3x_2-12=0$ (ii), on sloving i & ii

we get $x_2=8$ and $x_1=-6$

Now substitue them in 1a & 1b and slove for

λ_1 & λ_2 we get $\boxed{\lambda_2=-26}$ ✗

case-3 :- $\lambda_1 = 0 \quad \lambda_2 \neq 0$

Substitue in 1a & 1b

$-2x_1 + 4 - 2\lambda_2 = 0$

$-2x_2 + 6 - 3\lambda_2 = 0$

solving $x_1 = \frac{2}{3} x_2$

$\lambda_2 \neq 0$ so $2x_1 + 3x_2 - 12 = 0 \Rightarrow \boxed{x_2 = 3} \ \boxed{x_1 = 2}$

substitue in 3a & 3b

$x_1 + x_2 - 2 \leq 0 \Rightarrow 5 - 2 \leq 0 \Rightarrow 3 \leq 0$ X

case-4 : $\lambda_1 \neq 0 \quad \lambda_2 = 0$

similarly on solving we get

$\lambda_1 = 3 \quad \lambda_2 = 0$

$x_1 = 1/2 \quad x_2 = 3/2$

$$\boxed{\begin{array}{c} 0 \leq 0 \\ -13 \leq 0 \end{array}}$$ ✓

Now let's apply KKT on SVM

1. ~~Remer~~ Recall :- our lagrangian function:

$$L(w, d, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i [y_i (w \cdot x_i + d) - 1]$$

$y_i \in \{-1, 1\}$

condition - 1

Let's equate $\frac{\partial L}{\partial w} = 0$

$$\frac{\partial L}{\partial w} = \frac{2}{2}\|w\| - \sum_{i=1}^{m} \alpha_i y_i \frac{\partial}{\partial w}(w \cdot x_i) - \frac{\partial}{\partial w}\overset{\to 0}{\left[\sum \alpha_i y_i d\right]}$$

$$- \frac{\partial}{\partial w} \overset{\to 0}{\sum_{i=1}^{m} \alpha_i}$$

$$= \|w\| - \sum_{i=1}^{m} \alpha_i y_i x_i = 0$$

$$\Rightarrow \boxed{w = \sum_{i=1}^{m} \alpha_i y_i x_i}$$

$$\frac{\partial L}{\partial d} = \frac{\partial}{\partial d}\overset{\to 0}{\left(\frac{1}{2}\|w\|^2\right)} - \frac{d}{dd}\overset{\to 0}{\left[\sum \alpha_i y_i (w \cdot x_i)\right]}$$

$$- \frac{\partial}{\partial d}\left[\sum \alpha_i y_i d\right] - \frac{\partial}{\partial d}\overset{\to 0}{\sum_{i=1}^{m} \alpha_i}$$

$$= -\sum_{i=1}^{m} \alpha_i y_i = 0$$

$$\Rightarrow \boxed{\sum_{i=1}^{m} \alpha_i y_i = 0} \qquad \left(w \cdot w = |w||w|\cos(0) = |w|^2\right)$$

Let us substitute w by this value into L:

$$W(\alpha, d) = \frac{1}{2}\left(\sum_{i=1}^{m} \alpha_i y_i x_i\right) \cdot \left(\sum_{j=1}^{m} \alpha_j y_j x_j\right) -$$

$$\sum_{i=1}^{m} \alpha_i \left[y_i\left(\left(\sum_{j=1}^{m} \alpha_j y_j x_j\right) \cdot x_i + d\right) - 1\right]$$

$$= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$- d \overset{\to 0}{\sum_{i=1}^{m} \alpha_i y_i} + \sum_{i=1}^{m} \alpha_i$$

$$\Rightarrow W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad \text{— (A)}$$

condition-2

$\lambda_i \rightarrow \alpha_i \qquad h_i \rightarrow y_i(w \cdot x_i + d) - 1 \geq 0$

$\alpha_i [y_i (w \cdot x_i + d) - 1] = 0$

we see that either $\alpha_i = 0$ or $y_i (w \cdot x_i + d) - 1 = 0$

If $y_i (w \cdot x_i + d) - 1 = 0$ and using condition 4

we know $\boxed{\alpha_i \geq 0}$. which means support vectors are the datapoints that have a positive Lagrane multiplier.

condition-3

$h_i \leq 0$

$\Rightarrow -y_i (w \cdot x_i + d) + 1 \leq 0$

* To compute W also we need α_i, now how to find out these $\alpha_1, \alpha_2, \ldots \alpha_m$.

Imagine guys how tedious task it will be if we try to find $\alpha_1, \alpha_2 \ldots \alpha_m$ manually!

Think about it!