Fundamentals Machine Learning

Vector Calculus: Lagrange Multiplier

Constrained optimization

- **Constrained optimization is the process of optimizing an objective function with respect to some variables in the presence of constraints on those variables.**
- The **objective function** is either
  - a cost function or energy function which is to be minimized, or
  - a reward function or utility function, which is to be maximized.
- **Constraints** can be either
  - hard constraints which set conditions for the variables that are **required** to be satisfied, or
  - soft constraints which have some variable values that are penalized in the objective function if the conditions on the variables are not satisfied

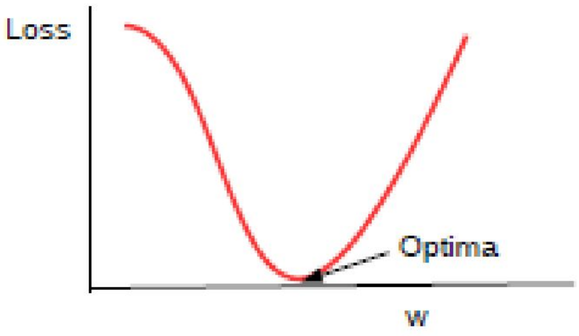
- A general constrained minimization problem:

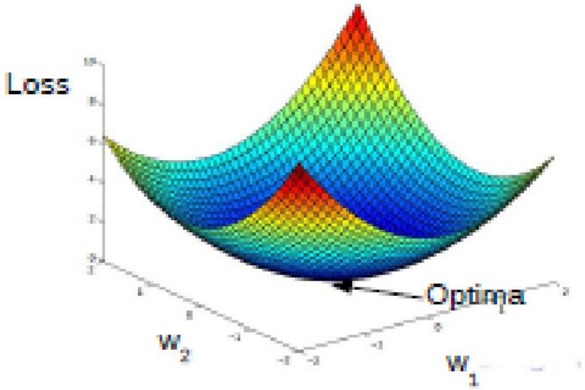
$$
\begin{aligned}
\min \quad & f(\mathbf{x}) \\
\text{subject to} \quad & g_i(\mathbf{x}) = c_i \text{ for } i = 1, \ldots, n \text{ (Equality constraints)} \\
& h_j(\mathbf{x}) \geq d_j \text{ for } j = 1, \ldots, m \text{ (Inquality constraints)}
\end{aligned}
\tag{1}
$$

where $g_i(\mathbf{x}) = c_i$ and $h_j(\mathbf{x}) \geq d_j$ are called *hard constraints*.

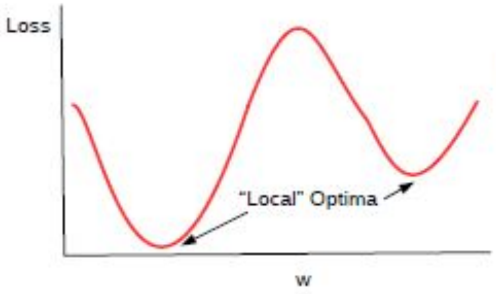
Optimization Problems in ML

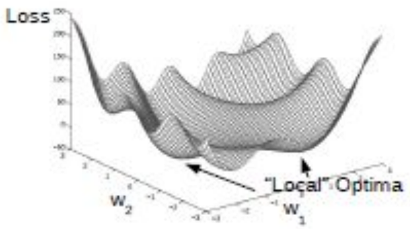
Wish to find the optima (minima) of an objective function, that can be seen as a curve/surface





In many cases, the functions may even look like this





Functions with unique minima: Convex;
Functions with many local minima: Non-convex

The Lagrange Multipliers

- Lagrange Multipliers are a mathematical method used to solve constrained optimization problems of differentiable functions

$$\text{Minimize } f(x_1, x_2)$$

subject to

$$g(x_1, x_2) = 0$$

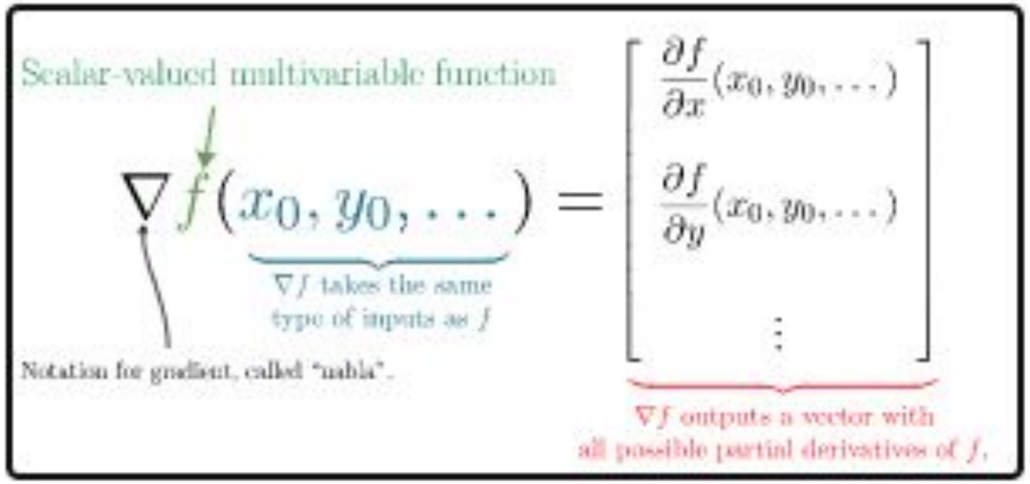
i.e., optimize f, while constraining f with g.

Lagrange Multipliers is the following equation:

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

The **gradient of f** is equal to some multiplier (**lagrange multiplier**) times the **gradient of g , where g(x)=0**

The gradient of a scalar-valued multivariable function $f(x,y,\ldots)$ denoted ∇f packages all its partial derivative information into a vector:



Combining Lagrange multiplier and constraint function g(x)=0, the Lagrange Function is defined as

$$L(x, \lambda) = f(x) - \lambda g(x)$$

Using the above equation, look for the points where:

$$\nabla L(x, \lambda) = 0$$

Example 1: One Equality Constraint

Problem: Given,

$$
\begin{aligned}
f(x, y) &= 2 - x^2 - 2y^2 \\
g(x, y) &= x + y - 1 = 0
\end{aligned}
$$

Find the extreme values

Solution: Step 1: we put the equations into the form of a Lagrangian:

$$
\begin{aligned}
L(x, y, \lambda) &= f(x, y) - \lambda g(x, y) \\
&= 2 - x^2 - 2y^2 - \lambda(x + y - 1)
\end{aligned}
$$

Step 2: solve for the gradient of the Lagrangian

$$\nabla L(x, y, \lambda) = \nabla f(x, y) - \lambda \nabla g(x, y) = 0$$

$$L(x, y, \lambda) = 2 - x^2 - 2y^2 - \lambda(x + y - 1)$$

$$\nabla L(x, y, \lambda) = \nabla f(x, y) - \lambda \nabla g(x, y) = 0$$

$$\frac{\partial}{\partial x} L(x, y, \lambda) = -2x - \lambda = 0 \quad (1)$$
$$\frac{\partial}{\partial y} L(x, y, \lambda) = -4y - \lambda = 0 \quad (2)$$
$$\frac{\partial}{\partial \lambda} L(x, y, \lambda) = x + y - 1 = 0 \quad (3)$$

From Equation (1) and (2) we have x = 2y. Substituting this into Equation (3) gives $x = \frac{2}{3}$ *and* $y = \frac{1}{3}$

These values give $\lambda = \frac{-4}{3}$ *and* $f = \frac{4}{3}$

Example 2: One Equality Constraint

Problem: Given,

$$\begin{aligned} f(x,y) &= x + 2y \\ g(x,y) &= y^2 + xy - 1 = 0 \end{aligned}$$

Find the extreme values.

Solution: First, we put the equations into the form of a Lagrangia

$$\begin{aligned} L(x,y,\lambda) &= f(x,y) - \lambda g(x,y) \\ &= x + 2y - \lambda(y^2 + xy - 1) \end{aligned}$$

and we solve for the gradient of the Lagrangian (Equation 4):

$$\nabla L(x,y,\lambda) = \nabla f(x,y) - \lambda \nabla g(x,y) = 0$$

which gives us:

$$\begin{aligned}
\frac{\partial}{\partial x}L(x,y,\lambda) &= 1-\lambda y=0\\
\frac{\partial}{\partial y}L(x,y,\lambda) &= 2-2\lambda y-\lambda x=0\\
\frac{\partial}{\partial \lambda}L(x,y,\lambda) &= y^2+xy-1=0
\end{aligned}$$

This gives $x=0$, $y=\pm 1$, $\lambda=\pm 1$ and $f=\pm 2$.

Multiple Constraints

Lagrangian function for multiple constraints:

$$L(x, \lambda) = f(x) - \sum_i \lambda_i g_i(x)$$

Here $g_i(x)$ and λ_i are the multiple constraints (denoted by i), and associated Lagrange Multipliers.
Note : that each constraint has its own multiplier.

Again, look for points where:

$$\nabla L(x, \lambda) = 0$$

Example 3: Two Equality

Problem: Given,

$$
\begin{aligned}
f(x,y) &= x^2 + y^2 \\
g_1(x,y) &= x + 1 = 0 \\
g_2(x,y) &= y + 1 = 0
\end{aligned}
$$

Find the extreme values.

Solution: Step 1: Put the equations into the form of a Lagrangian:

$$
\begin{aligned}
L(x,y,\lambda) &= f(x,y) - \lambda_1 g_1(x,y) - \lambda_2 g_2(x,y) \\
&= x^2 + y^2 - \lambda_1(x+1) - \lambda_2(y+1)
\end{aligned}
$$

Step 2: solve for the gradient of the Lagrangian

$$\nabla L(x,y,\lambda) = \nabla f(x,y) - \lambda_1 \nabla g_1(x,y) - \lambda_2 \nabla g_2(x,y) = 0$$

$$
\begin{aligned}
\frac{\partial}{\partial x} L(x,y,\lambda) &= 2x - \lambda_1 = 0 && (1) \\
\frac{\partial}{\partial y} L(x,y,\lambda) &= 2y - \lambda_2 = 0 && (2) \\
\frac{\partial}{\partial \lambda_1} L(x,y,\lambda) &= x + 1 = 0 && (3) \\
\frac{\partial}{\partial \lambda_2} L(x,y,\lambda) &= y + 1 = 0 && (4)
\end{aligned}
$$

Equation (3) gives x = −1. Equation (4) gives y = −1.
Substituting this into Equation (1) and (2) gives $\lambda_1 = -2$, $\lambda_2 = -2$ and **f = 2**.

Inequality Constraints

Lagrange Multipliers with Inequality Constraints (ie $g(x) \leq 0, g(x) \geq 0$).

The formulation of the Lagrangian

$$L(x, \lambda) = f(x) - \lambda g(x)$$

The rules on how the Lagrange Multipliers encode the inequality constraints:

$$g(x) \geq 0 \quad \Rightarrow \quad \lambda \geq 0 \qquad \text{(R1)}$$

$$g(x) \leq 0 \quad \Rightarrow \quad \lambda \leq 0 \qquad \text{(R2)}$$

$$g(x) = 0 \quad \Rightarrow \quad \lambda \text{ is unconstrained} \qquad \text{(R3)}$$

Example 4: One Inequality Constraint

Problem: Given,

$$
\begin{aligned}
f(x,y) &= x^3 + y^2 \\
g(x,y) &= x^2 - 1 \geq 0
\end{aligned}
$$

Find the extreme values.

Solution: Step 1: Put the equations into the form of a Lagrangian:

$$
\begin{aligned}
L(x,y,\lambda) &= f(x,y) - \lambda g(x,y) \\
&= x^3 + y^2 - \lambda(x^2 - 1)
\end{aligned}
$$

Step 2: solve for the gradient of the Lagrangian

$$\nabla L(x,y,\lambda) = \nabla f(x,y) - \lambda \nabla g(x,y) = 0$$

$$\frac{\partial}{\partial x}L(x,y,\lambda) = 3x^2 - 2\lambda x = 0 \qquad (1)$$

$$\frac{\partial}{\partial y}L(x,y,\lambda) = 2y = 0 \qquad (2)$$

$$\frac{\partial}{\partial \lambda}L(x,y,\lambda) = x^2 - 1 = 0 \qquad (3)$$

Furthermore,

$$\lambda \geq 0 \qquad (4)$$

From Equation (2), we have y = 0.
From Equation (3), we have x = ±1.
Substituting this into Equation (1) gives $\lambda = \pm\frac{3}{2}$
Since we require that $\lambda \geq 0$, $\lambda = \frac{3}{2}$
This gives x = 1, y = 0 and f = 1.

Equations (1) through (4) are called **the KKT conditions**.

Example 5: Two Inequality Constraints

Problem: Given,

$$\begin{aligned} f(x,y) &= x^3 + y^3 \\ g_1(x,y) &= x^2 - 1 \geq 0 \\ g_2(x,y) &= y^2 - 1 \geq 0 \end{aligned}$$

Find the extreme values.

Solution: Step 1: put the equations into the form of a Lagrangian:

$$\begin{aligned} L(x,y,\lambda) &= f(x,y) - \lambda_1 g_1(x,y) - \lambda_2 g_2(x,y) \\ &= x^3 + y^3 - \lambda_1(x^2 - 1) - \lambda_2(y^2 - 1) \end{aligned}$$

Step 2: solve for the gradient of the Lagrangian

$$\nabla L(x,y,\lambda) = \nabla f(x,y) - \lambda_1 \nabla g_1(x,y) - \lambda_2 \nabla g_2(x,y) = 0$$

$$\frac{\partial}{\partial x} L(x,y,\lambda) = 3x^2 - 2\lambda_1 x = 0 \quad (1)$$

$$\frac{\partial}{\partial y} L(x,y,\lambda) = 3y^2 - 2\lambda_2 y = 0 \quad (2)$$

$$\frac{\partial}{\partial \lambda_1} L(x,y,\lambda) = x^2 - 1 = 0 \quad (3)$$

$$\frac{\partial}{\partial \lambda_2} L(x,y,\lambda) = y^2 - 1 = 0 \quad (4)$$

Furthermore,

$$\lambda_1 \geq 0 \quad (5)$$

$$\lambda_2 \geq 0 \quad (6)$$

- From Equations (3) and (4), we have x = ± 1 and y = ± 1. Substituting x = ± 1 into Equation (1) gives $\lambda_1 = \pm \frac{3}{2}$.

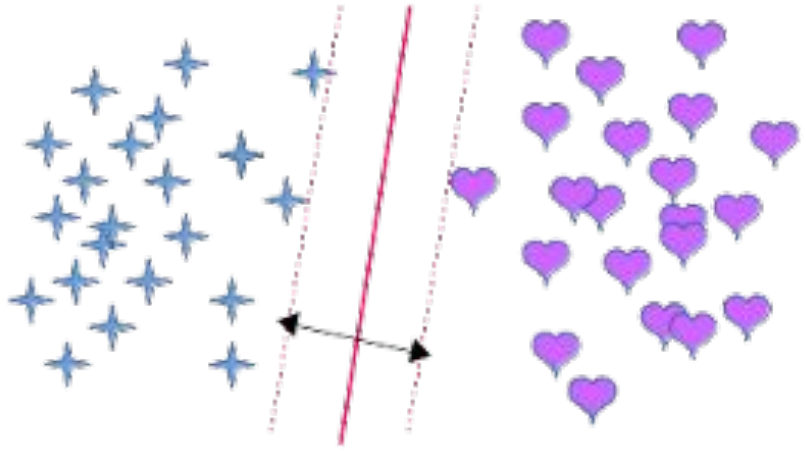
Furthermore $\lambda_1 \geq 0$, hence $\lambda_1 = +\frac{3}{2}$

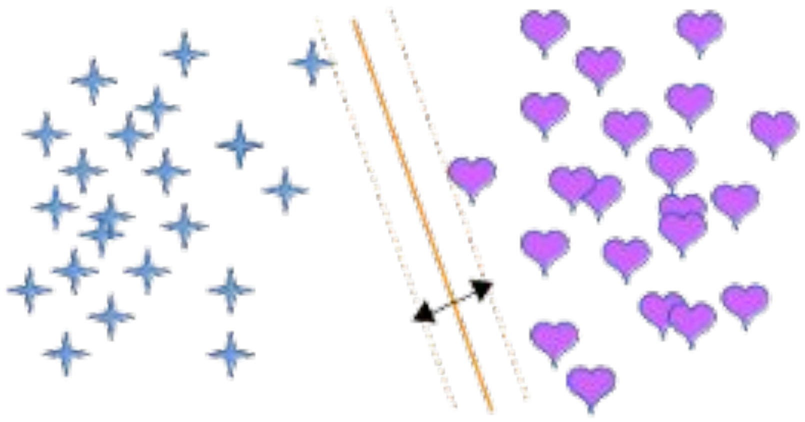
- substituting y = ± 1 into Equation (2) gives $\lambda_2 = \pm \frac{3}{2}$

Since we require that $\lambda_2 \geq 0$, then $\lambda_2 = +\frac{3}{2}$.

This gives x = 1, y = 1 and f = 2.

Application in Support Vector Machine



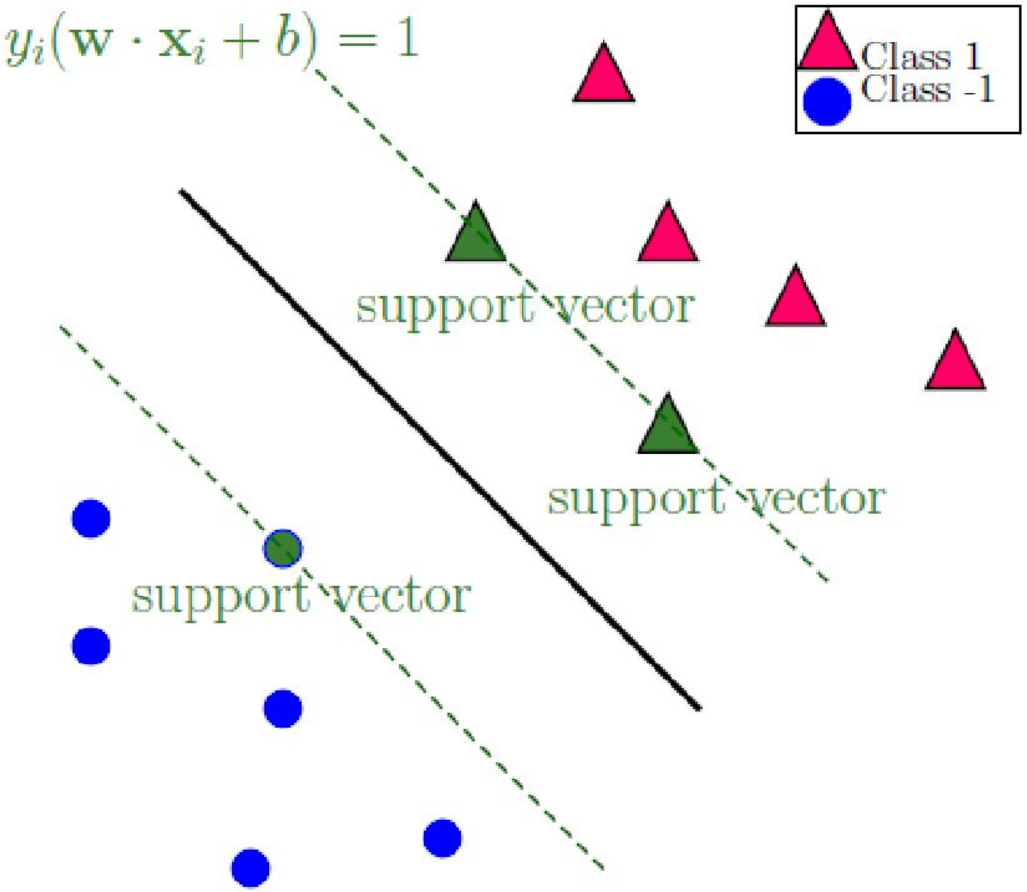


Which line creates the maximum margin or “widest road”?

SVM

- The samples on the edge of the boundary lines (dotted) lines, are known as '**Support Vectors'.**
- On the left side there are two such samples (blue stars), compared to the one on the right.
    - *Support Vectors are the samples that are most difficult to classify.*
    - *They directly affect the process to find the optimum location of the decision boundaries (dotted lines).*
- This is a **constrained optimization** problem.
- *Optimization* — because, we are to find the line from which the support vectors are maximally separated
- *Constrained* — because, the support vectors should be away from the road and not on the road.
- **Lagrange Multipliers are used to solve this problem in SVM**

SVM



$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
Class 1
Class -1
support vector
support vector
support vector

The binary SVM problem

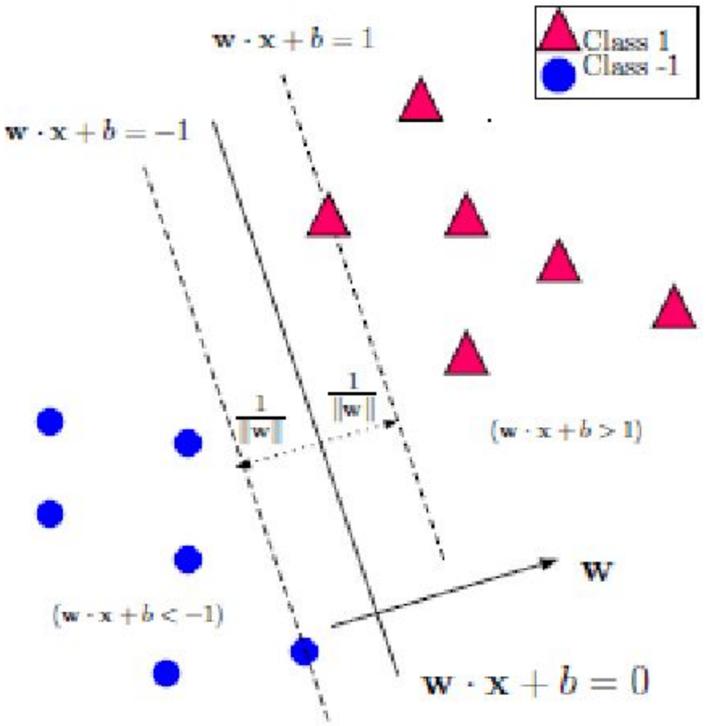
Problem. Given training data $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ with labels $y_i = \pm 1$, SVM finds the optimal separating hyperplane by maximizing the class margin.

Specifically, it tries to solve

$$\max_{\mathbf{w},b} \frac{2}{\|\mathbf{w}\|_2} \quad \text{subject to}$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1, \quad \text{if } y_i = +1;$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1, \quad \text{if } y_i = -1$$



Remark. The classification rule for new data $\mathbf{x}$ is $y = \mathrm{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$.

The previous problem is equivalent to

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2 \qquad \text{subject to} \quad y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 \text{ for all } 1 \leq i \leq n.$$

This is an optimization problem with linear, inequality constraints.

Lagrange method applied to binary SVM

- The Lagrange function is

$$L(\mathbf{w}, b, \lambda_1, \ldots, \lambda_n) = \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{i=1}^{n} \lambda_i (y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1)$$

- The KKT conditions are

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum \lambda_i y_i \mathbf{x}_i = 0, \quad \frac{\partial L}{\partial b} = \sum \lambda_i y_i = 0$$
$$\lambda_i (y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) = 0, \ \forall i$$
$$\lambda_i \geq 0, \ \forall i$$
$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \ \forall i$$

References

- https://www.engr.mun.ca/~baxter/Publications/LagrangeForSVMs.pdf
- https://www.kdnuggets.com/2019/08/lagrange-multipliers-visualizations-code.html
- https://medium.com/fintechexplained/calculus-multivariate-calculus-and-machine-learning-242b9efcb41c
- https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivative-and-gradient-articles/a/the-gradient#:~:text=In%20the%20case%20of%20scalar,partial%20derivatives%20into%20a%20vector.
- https://towardsdatascience.com/calculus-in-data-science-and-its-uses-3f3e1b5e5b35

https://www.engr.mun.ca/~baxter/Publications/LagrangeForSVMs.pdf