Machine Learning

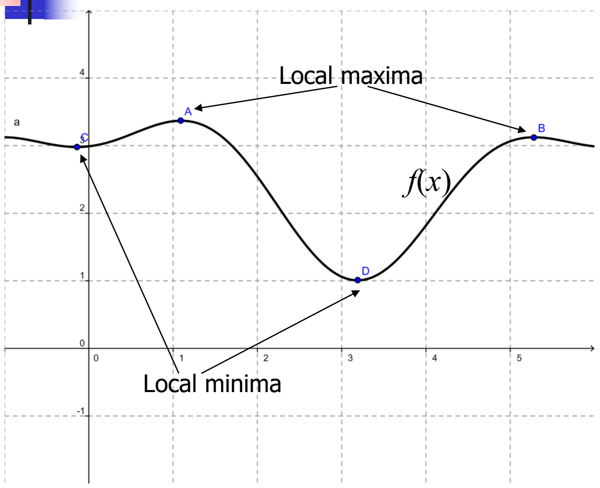
11. week

- Lagrange Optimization
- Support Vector Machine (SVM)
- Quadratic Programming

Optimization

- With the simplest use in mathematics, it is the methodology of finding the largest and smallest values of a function that can take.
- The most basic method to find the local minima and maxima is to take the first derivation of the function.

Local Minima-Maxima



$$f' = \frac{\partial f}{\partial x} = 0$$

Umut ORHAN, PhD.

3

3

Local Minima-Maxima

- At each local minimum and maximum point, the first derivation is zero, but each point where the derivation is zero may not be local minimum or maximum.
- For this, the second derivation must be examined:

  If $f''(x) < 0$ local maximum

  If $f''(x) > 0$ local minimum

Umut ORHAN, PhD.

4

4

Lagrange Optimization

- Lagrange Optimization is the most basic method applied when it is necessary to find the optimum (largest or smallest) value of a function depending on some constraints.
- The Lagrange function is constructed by adding a constraint g(x) at the rate of a α coefficient to the objective f(x).

$$L(x,\alpha) = f(x) + \alpha\, g(x)$$

Umut ORHAN, PhD. 5

5

Lagrange Optimization

- In the case of finding the local min-max values, the solution is searched by equating the derivations of the Lagrange function to zero for each dependent variable.

$$\frac{\partial L(x,\alpha)}{\partial x} = 0$$

$$\frac{\partial L(x,\alpha)}{\partial \alpha} = 0$$

Umut ORHAN, PhD. 6

6

Example

Find the optimum value of the equation

f(x, y) = 25-x^2-y^2 satisfying the constraint 4-x-y = 0

$$L(x,y,\alpha) = 25 - x^2 - y^2 + \alpha(4 - x - y)$$

$$\frac{\partial L(x,y,\alpha)}{\partial x} = -2x - \alpha = 0$$

$$\frac{\partial L(x,y,\alpha)}{\partial y} = -2y - \alpha = 0$$

$$\frac{\partial L(x,y,\alpha)}{\partial \alpha} = 4 - x - y = 0$$

Example

Find the optimum value of the equation

f (x, y) = 25-x^2-y^2 satisfying the condition 4-x-y = 0

$$x = -0.5\alpha \qquad \alpha = -4$$

$$y = -0.5\alpha \qquad x = 2$$

$$x + y = 4 \qquad y = 2$$

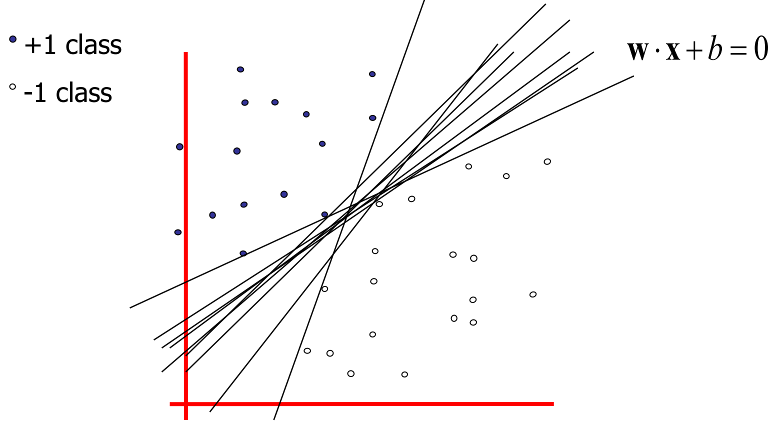
$f(x=2, y=2) = 25 - 2^2 - 2^2 = 17$ is found.

Support Vector Machine(SVM)

- The support vector machine (SVM) aims to maximize the perpendicular distances of the closest sample of a class to the separating surface.
- The separating surface may have many different alternatives without changing the performance on the dataset.
- Through the SVM, the separator surface is at the same distance (maximum distance) from both classes.

Support Vector Machine (SVM)

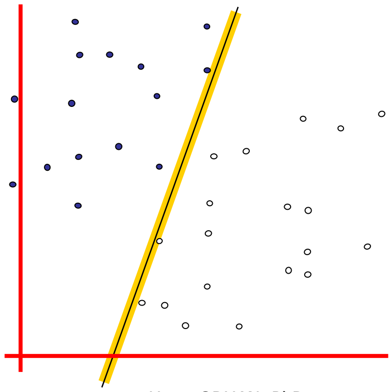
Other classifiers may find many alternatives for the separating surface.



Support Vector Machine (SVM)

Some of them can not maximize the distance between classes.

- +1 class
- -1 class



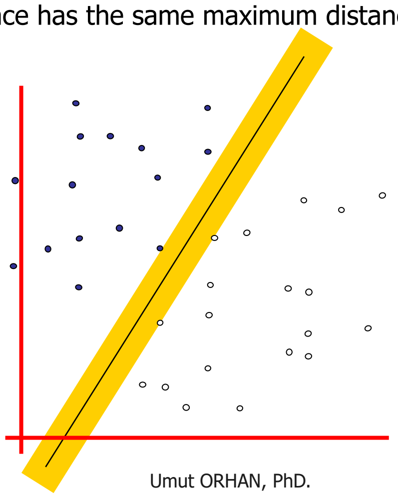
$$\mathbf{w}\cdot\mathbf{x}+b=0$$

Umut ORHAN, PhD.

Support Vector Machine (SVM)

The SVM surface has the same maximum distance from both classes.

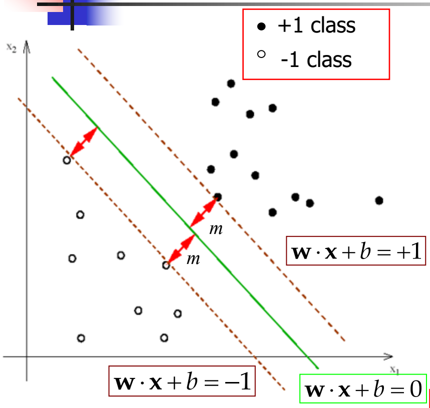
- +1 class
- -1 class



$$\mathbf{w}\cdot\mathbf{x}+b=0$$

Umut ORHAN, PhD.

Support Vector Machine(SVM)



$$w \cdot x_i + b \geq +1 \text{ for } d_i = +1$$
$$w \cdot x_i + b \leq -1 \text{ for } d_i = -1$$

$$m = \frac{2}{\sqrt{w \cdot w}} \rightarrow f_{\min}(w) = \frac{w \cdot w}{2}$$
$$d_i(w \cdot x_i + b) \geq 1, \quad 1 \leq i \leq N$$

$$L(w,b,\alpha) = \frac{w.w}{2} - \sum_i \alpha_i \left(d_i(w \cdot x_i + b) - 1\right)$$

Support Vector Machine (SVM)

$$\frac{\partial L(w,b,\alpha)}{\partial w} = 0 \qquad \frac{\partial L(w,b,\alpha)}{\partial b} = 0 \qquad \frac{\partial L(w,b,\alpha)}{\partial \alpha} = 0$$

- According to the traditional practice of the Lagrange equation, the above three equations need to be considered in a common account.
- By using the relation among the three variables, the dependency can be reduced to one variable.

Support Vector Machine (SVM)

We have to make the Lagrange equation dependent on the single variable by applying the classical solution.

$$\frac{\partial L(w,b,\alpha)}{\partial w}=0\rightarrow \qquad w=\sum_i \alpha_i d_i x_i$$

$$\frac{\partial L(w,b,\alpha)}{\partial b}=0\rightarrow \qquad \sum_i \alpha_i d_i = 0$$

Is found. These expressions are replaced by the Lagrange equation.

Support Vector Machine (SVM)

The Lagrange equation is made dependent on this variable by a single variable.

$$L(w,b,\alpha)=\frac{w.w}{2}-\sum_i \alpha_i\left(d_i\left(w\cdot x_i+b\right)-1\right)$$

$$L(\alpha)=\sum_i \alpha_i-\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j d_i d_j x_i x_j$$

Mapping

The linear SVM solution can be applied after non-linear data sets are mapped to a top space with a selected kernel function.

$$L(\alpha)=\sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j d_i d_j K(x_i, x_j)$$

Mapping

The most commonly used kernel functions are linear, polynomial and RBF functions.

$$K(x_i,x_j)=\begin{cases} x_i x_j & linear \\ (1+x_i x_j)^p & polynomial \\ \exp\left(-\frac{1}{2}\left\|x_i - x_j\right\|^2\right) & RBF \end{cases}$$

Quadratic Programming

- In order to solve the Lagrange equation which is connected to one variable, it is possible to use some methods such as backprobagation, simulated annealing and EM (expectation–maximization) .
- Nevertheless, the most preferred solution is the Quadratic Programming.

$$\min_{u}\left(\frac{u^T R u}{2} + d^T u + c\right)$$

Quadratic term

Umut ORHAN, PhD.

19

19

Quadratic Programming

Quadratic Programming is basically an iterative approach, similar to back-propagation. It is tried to find optimum u matrix by continuously updating the random initial values. Through the constraints, the initial values dependencies of back-propagation has been removed.

$$\min_{u}\left(\frac{u^T R u}{2} + d^T u + c\right)$$

Umut ORHAN, PhD.

20

20

Quadratic Programming

If the quadratic programming optimization technique is adapted to the univariate Lagrange equation;

$$L(\alpha)=\sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j d_i d_j x_i x_j$$

$$R \approx -D^T D\, X^T X \qquad and \qquad u \approx \alpha_i$$

Umut ORHAN, PhD. 21

21

Presentation Task

By using back-propagation technique instead of quadratic programming, we find α_i values that minimize the univariate Lagrange equation on any data set.

Umut ORHAN, PhD. 22

22