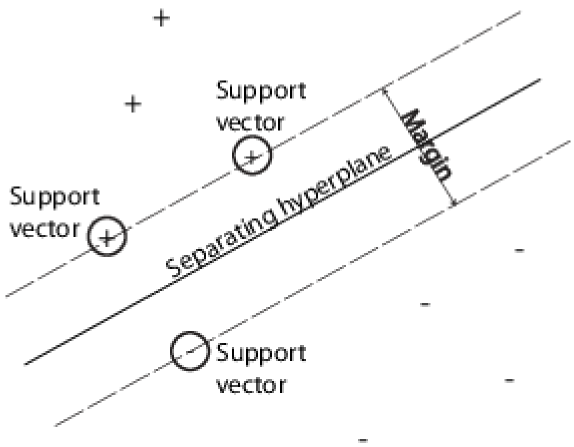
1.Derivation of optimal w for a binary SVM

Support machine learning (SVM) is one of a simple method for classification that represent as a point in n-dimensional space. It is represented by drawing a hyperplane that is a line in 2D or 3D. SVM tries to find the best solution between the two categories, it maximizes the distance points in either category which is margin and the category that fall in the line is support vectors.



$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n = 0, \quad \frac{\partial L}{\partial b} = -\sum_{n=1}^{N} \alpha_n y_n = 0$$

2.Derivation of optimal w's for a3-layer MLP

This is a solution to move the multi-layer of neurons in one layer and have multi-layers to have a better predictive value. There is a set of input data of the multi-layers or nodes and the layers in the middle between input and output. To get appropriate one, they need the gradients by derivative of multiple variables for each layer as the equation below:

$$(df_p)(v) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) & \cdots & \frac{\partial f}{\partial x_n}(p) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) v_i = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \nabla f(p) \cdot v$$

3.Define a loss-function for SVM

Loss function is a function to calculate how effective of machine learning model and find the optimal point and minimum loss. SVM is the model that support vectors which have effective impact on the model training and remove non-support vectors out, but it will not affect the value as it has no cost on the model.