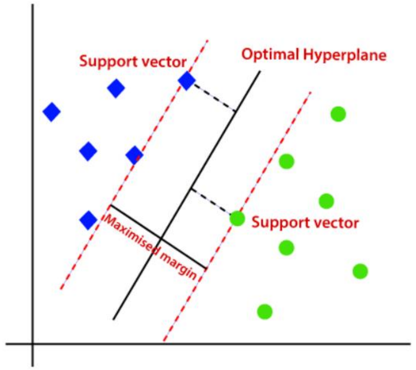
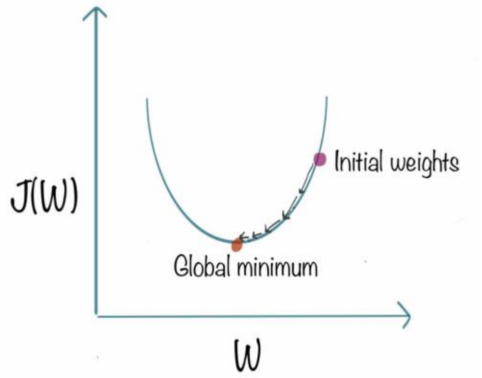
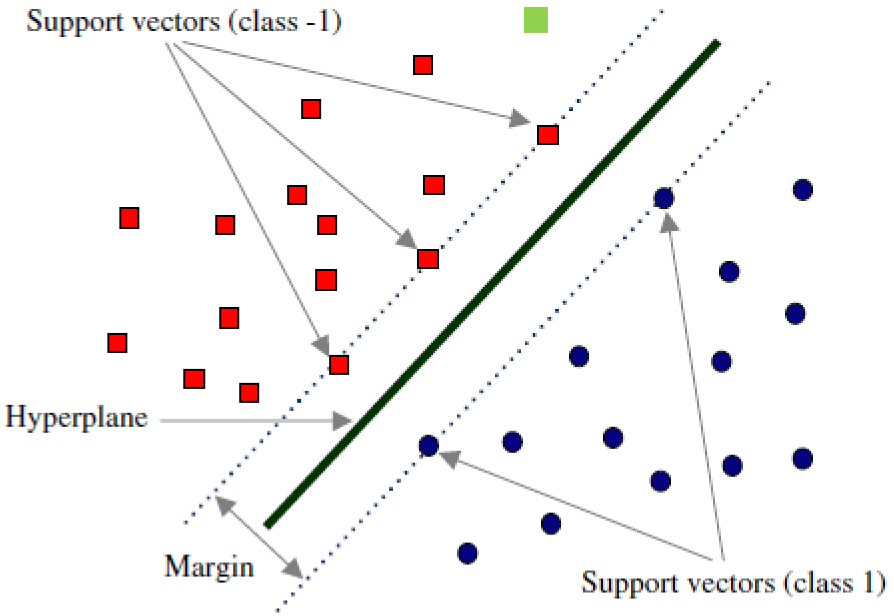
Siddhardhan

Gradient Descent for Support Vector Machine Classifier



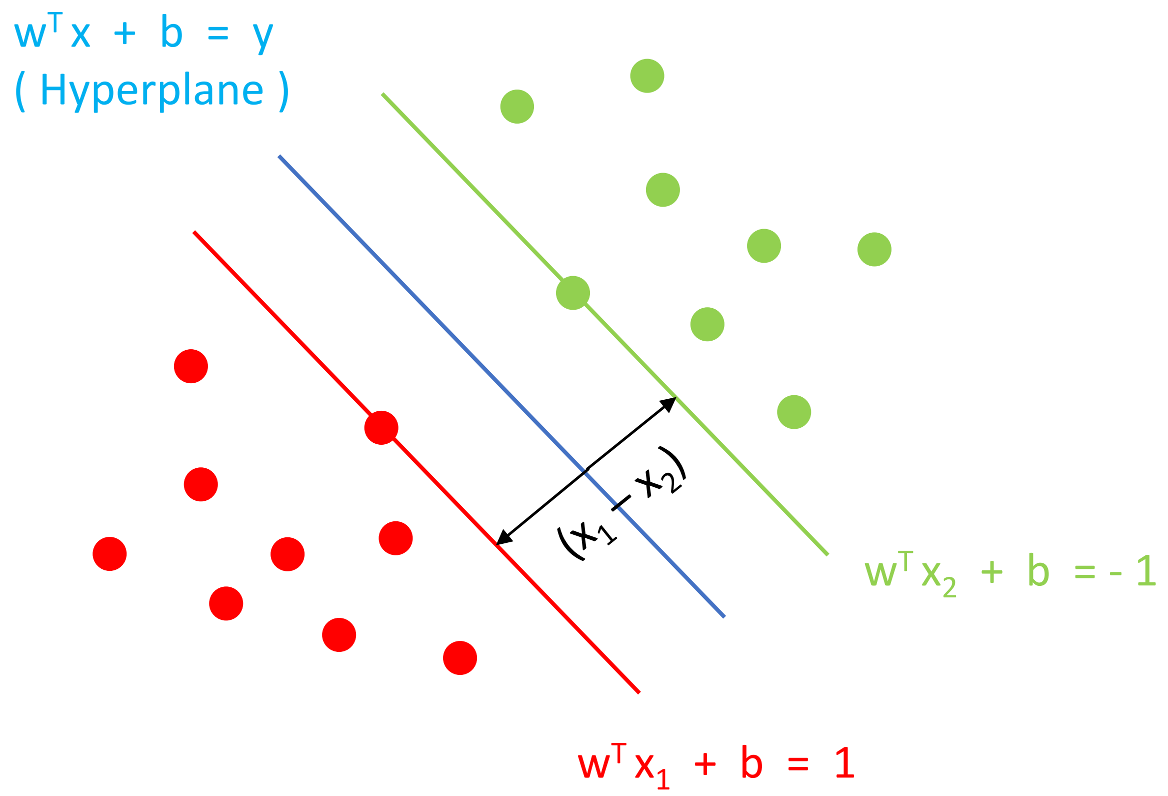


Support Vector Machine Classifier



- Hyperplane
- Support Vectors
- Margin

Support Vector Machine Classifier



$$\max \left(\frac{2}{\| w \|} \right) \quad \text{(margin)}$$

$$\hat{y}_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \\ 1, & w^T x_1 + b \geq 1 \end{cases}$$

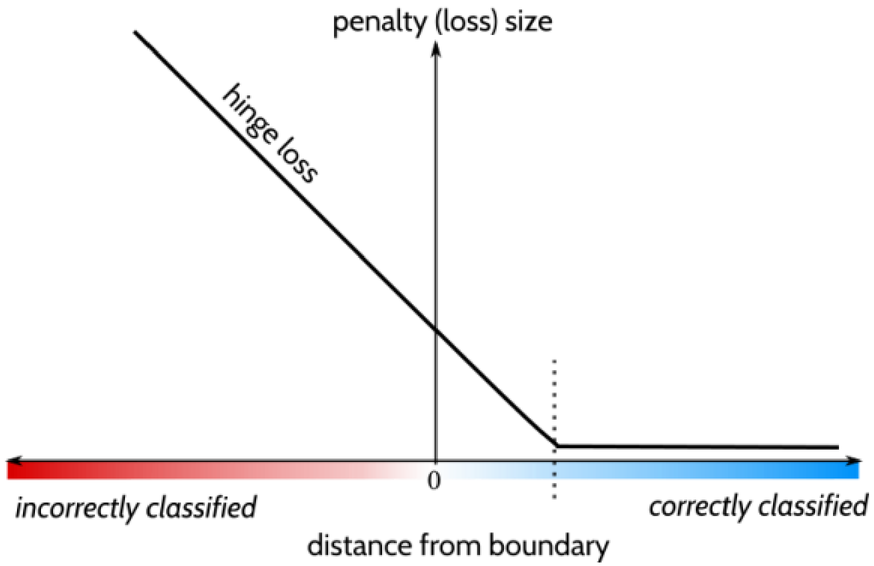
Hinge Loss

Hinge Loss is one of the types of Loss Function, mainly used for maximum margin classification models.

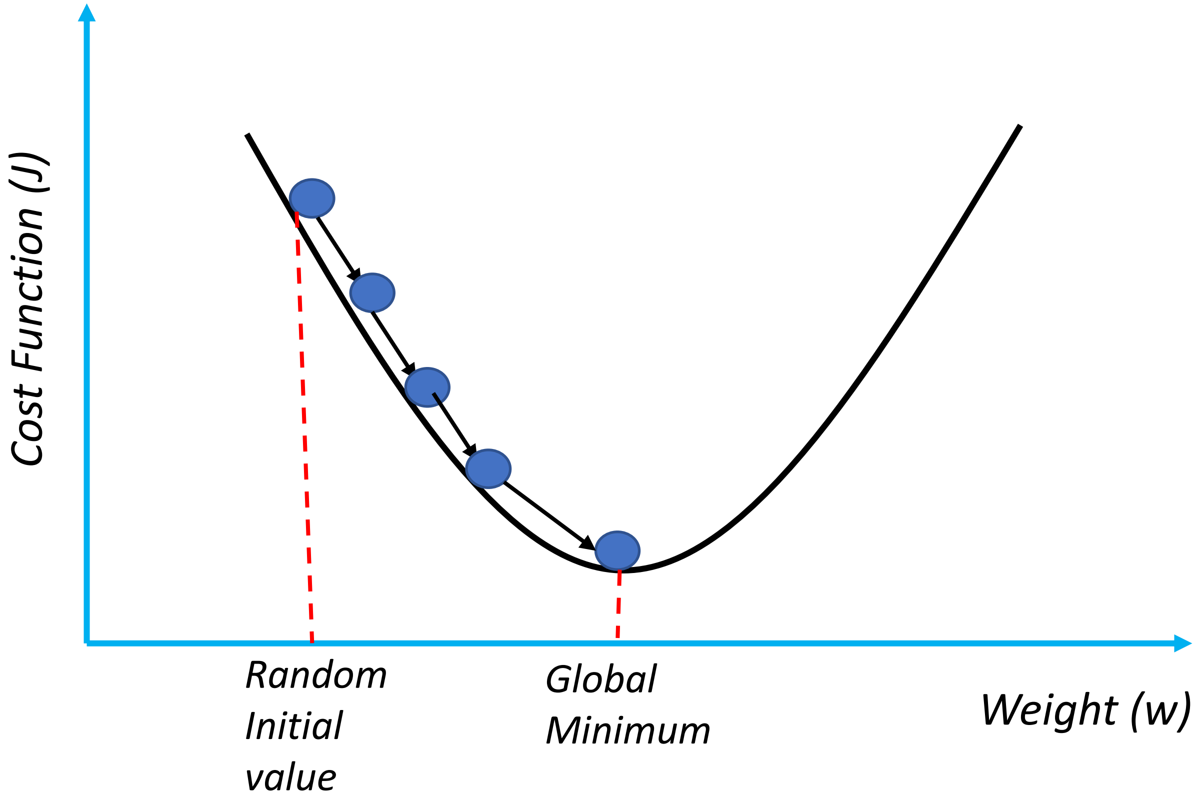
Hinge Loss incorporates a margin or distance from the classification boundary into the loss calculation. Even if new observations are classified correctly, they can incur a penalty if the margin from the decision boundary is not large enough.

$$L = \max(0, 1 - y_i(w^T x_i + b))$$

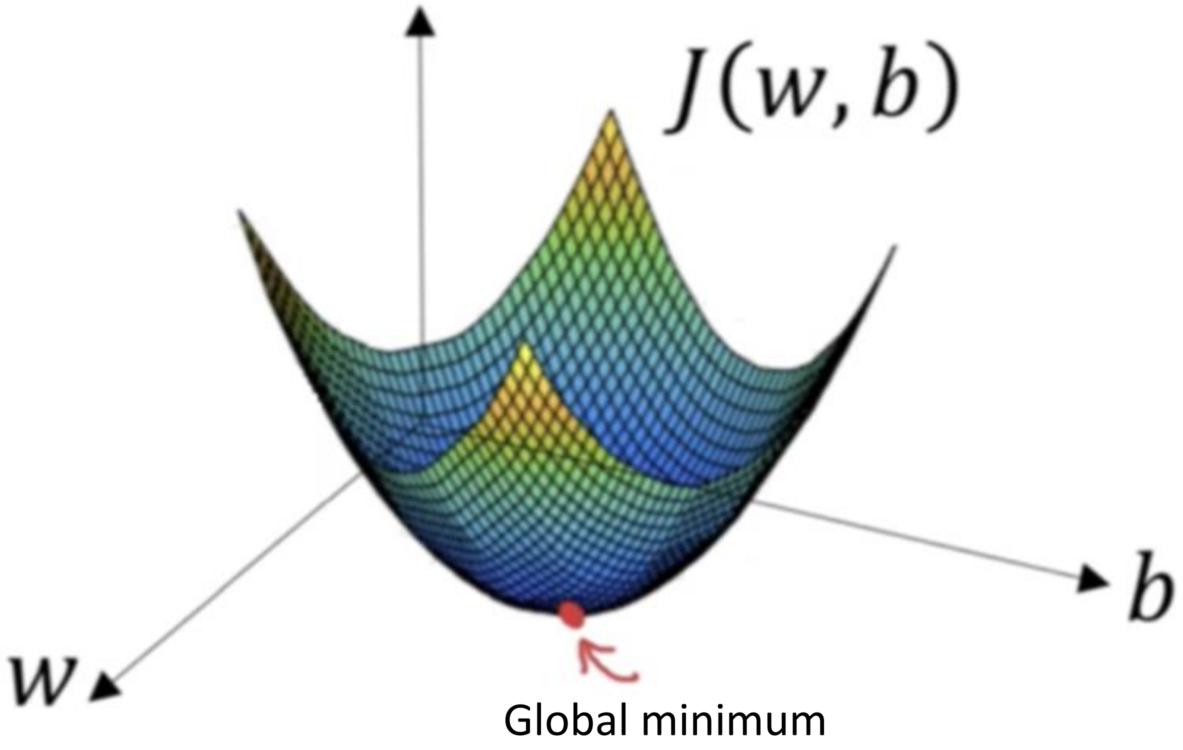
0 - for correct classification
1 - for wrong classification



Gradient Descent



Gradient Descent in 3 Dimension



Gradient Descent

Gradient Descent is an optimization algorithm used for minimizing the cost function in various machine learning algorithms. It is used for updating the parameters of the learning model.

$$w_2 = w_1 - L * \frac{dJ}{dw}$$

$$b_2 = b_1 - L * \frac{dJ}{db}$$

w --> weight
b --> bias
L --> Learning Rate

$\frac{dJ}{dw}$ --> Partial Derivative of cost function with respect to w

$\frac{dJ}{db}$ --> Partial Derivative of cost function with respect to b

Gradients for SVM Classifier

if $(\,y_i\,.\,(w.x + b)\ \geq\ 1)$ **:**

$$\frac{dJ}{dw} = 2\lambda w$$

$$\frac{dJ}{db} = 0$$

else $(\,y_i\,.\,(w.x + b)\ <\ 1)$ **:**

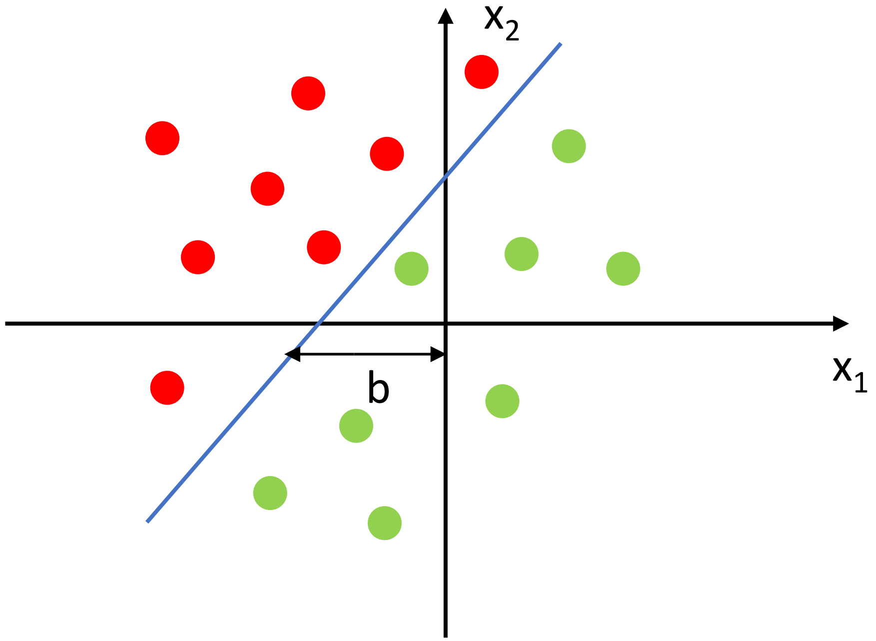
$$\frac{dJ}{dw} = 2\lambda w - y_i\,.\,x_i$$

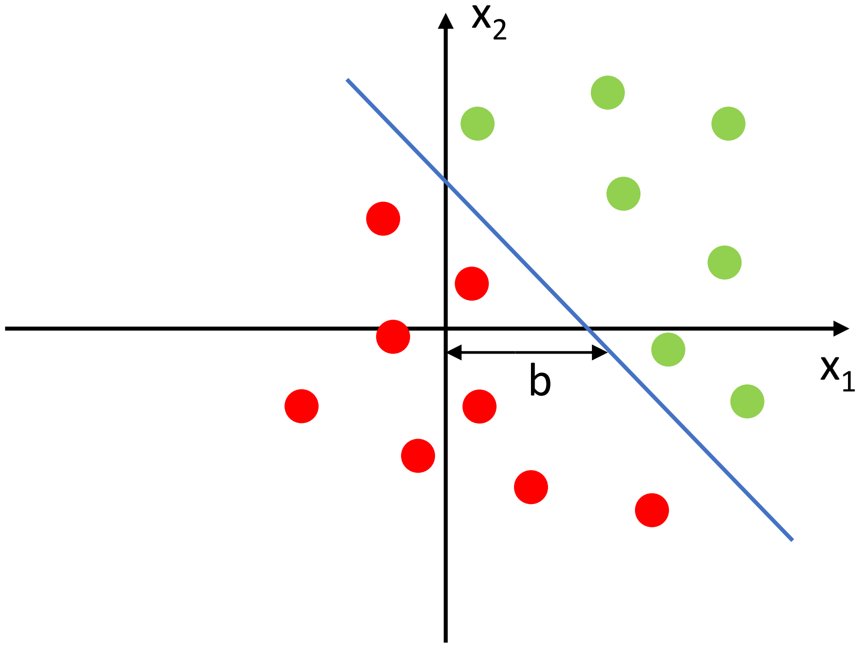
$$\frac{dJ}{db} = yi$$

$$w_2 = w_1 - L*\frac{dJ}{dw}$$

$$b_2 = b_1 - L*\frac{dJ}{db}$$

Support Vector Machine Classifier





Gradients for SVM Classifier

if $(\, y_i \,.\, (w.x_i - b) \;\geq\; 1) :$

$$\frac{dJ}{dw} = 2\lambda w$$

$$\frac{dJ}{db} = 0$$

else $(\, y_i \,.\, (w.x_i - b) \;<\; 1) :$

$$\frac{dJ}{dw} = 2\lambda w - y_i \,.\, x_i$$

$$\frac{dJ}{db} = yi$$

$$w_2 = w_1 - L * \frac{dJ}{dw}$$

$$b_2 = b_1 - L * \frac{dJ}{db}$$