NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

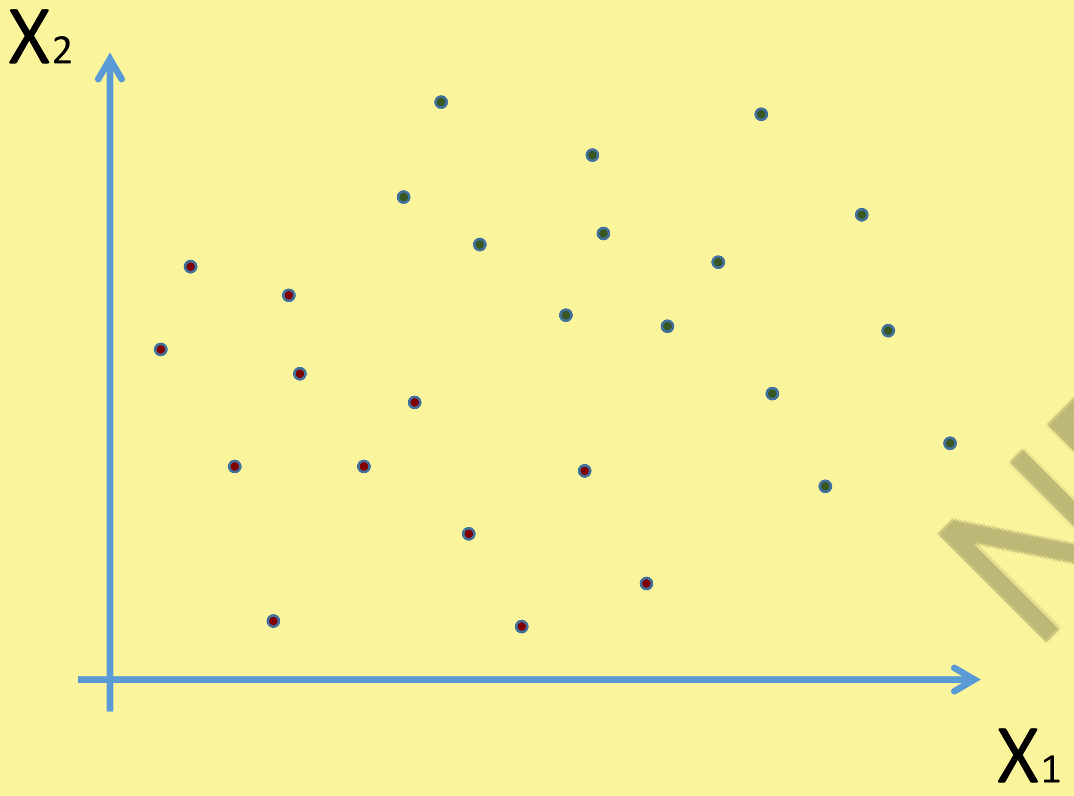
Lecture 11: Support Vector Machine

CONCEPTS COVERED

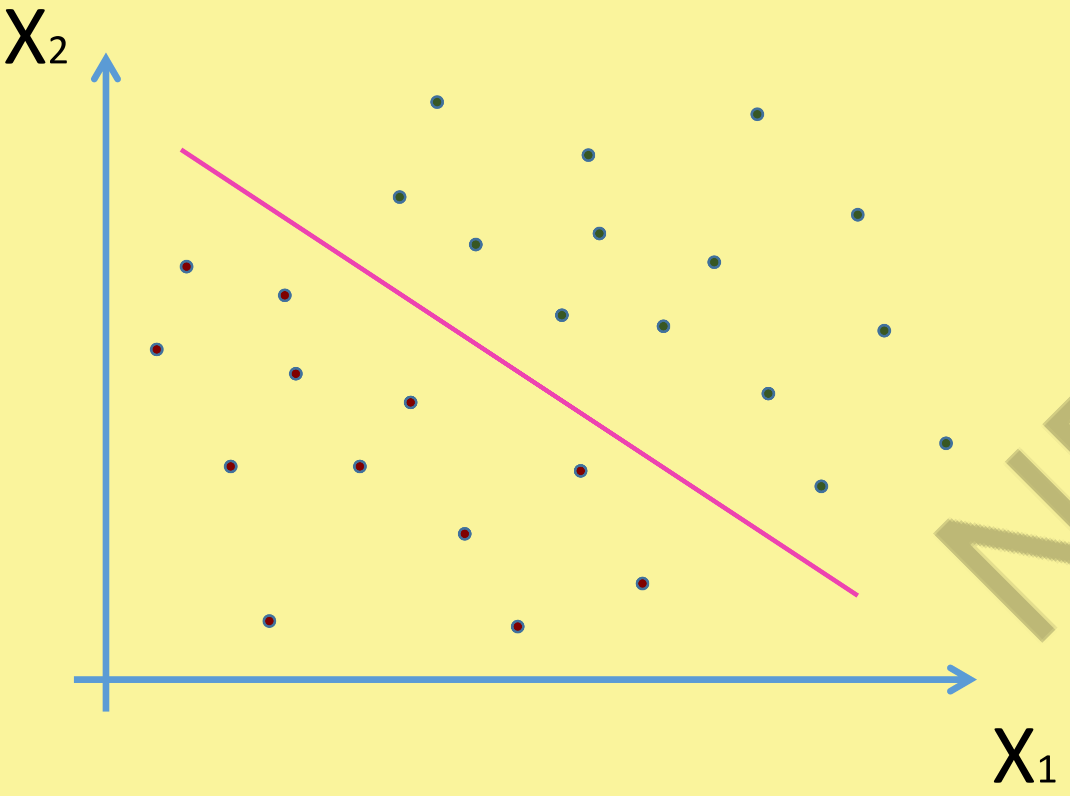
Concepts Covered:

- Linear Discriminator
- Perceptron Algorithm
- Support Vector Machine (SVM)

Linear Classifier – 2 Class Problem



Linear Classifier – 2 Class Problem



$$a^t X + b = 0$$

$$For\ X \in \omega_1 : \quad a^t X + b > 0$$

$$For\ X \in \omega_2 : \quad a^t X + b < 0$$

Linear Classifier – 2 Class Problem

$$a^t X + b = 0 \;\Rightarrow\; a^t X = 0$$

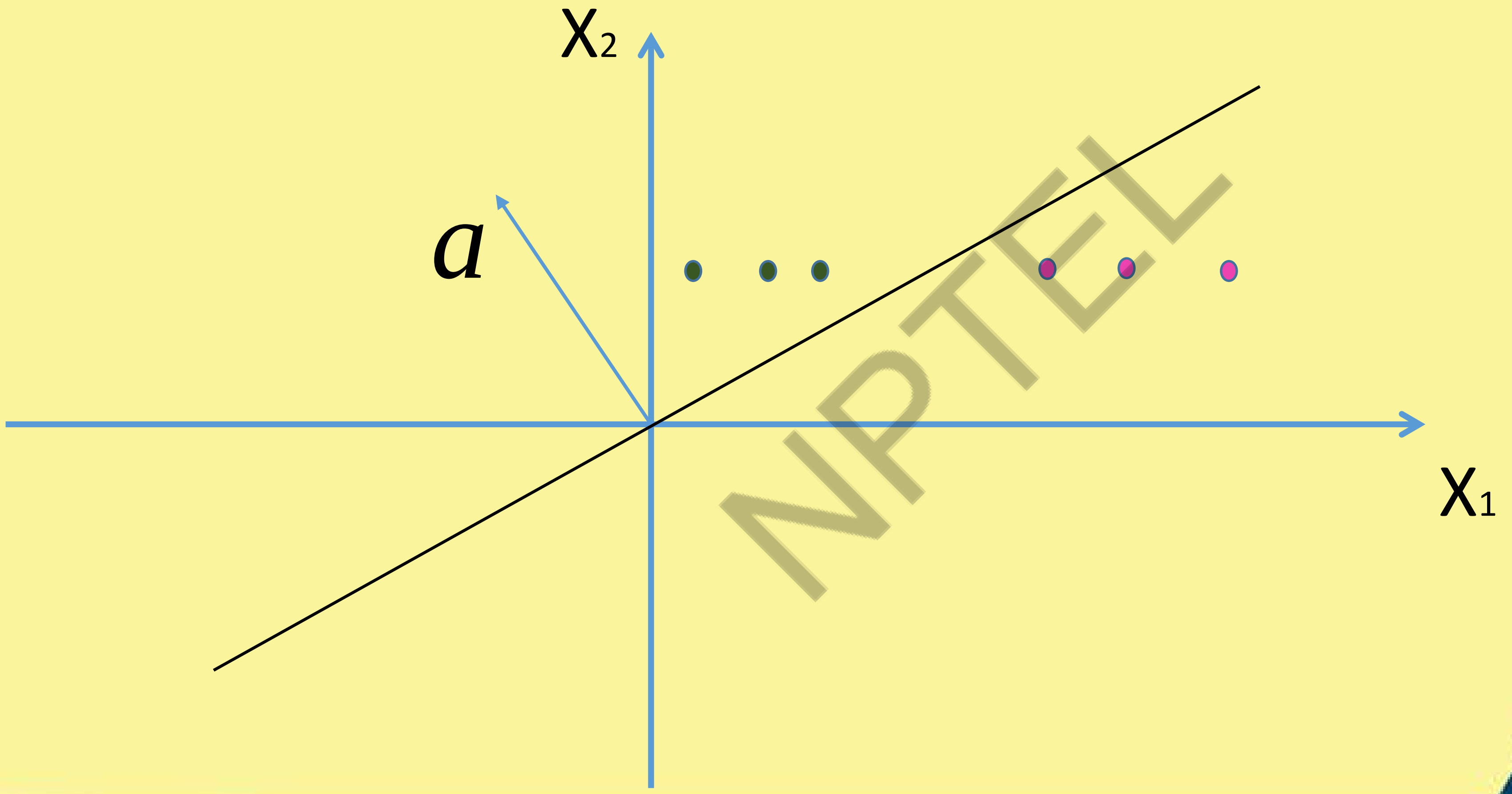
$$a = \begin{bmatrix} a_1 \\ a_2 \\ : \\ a_d \\ b \end{bmatrix} \qquad X = \begin{bmatrix} X_1 \\ X_2 \\ : \\ X_d \\ 1 \end{bmatrix}$$

Classification Rule

$$For Y \in \omega_1 :\; a^t X > 0$$

$$For Y \in \omega_2 :\; a^t X < 0$$

Linear Classifier – 2 Class Problem



Linear Classifier – 2 Class Problem

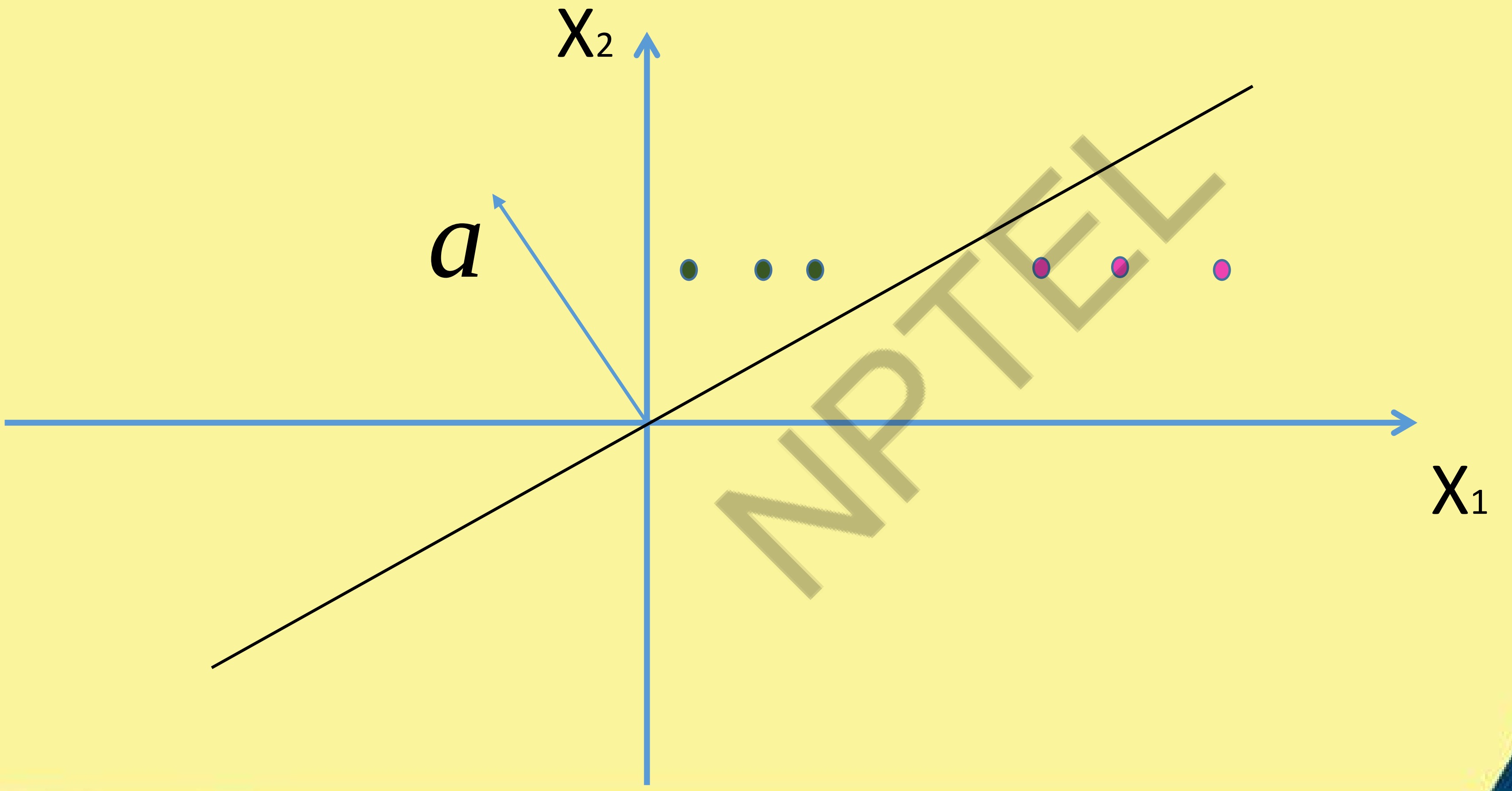
Negating all X from ω_2 : $X \leftarrow -X$

Classification Rule: $a^t X > 0$

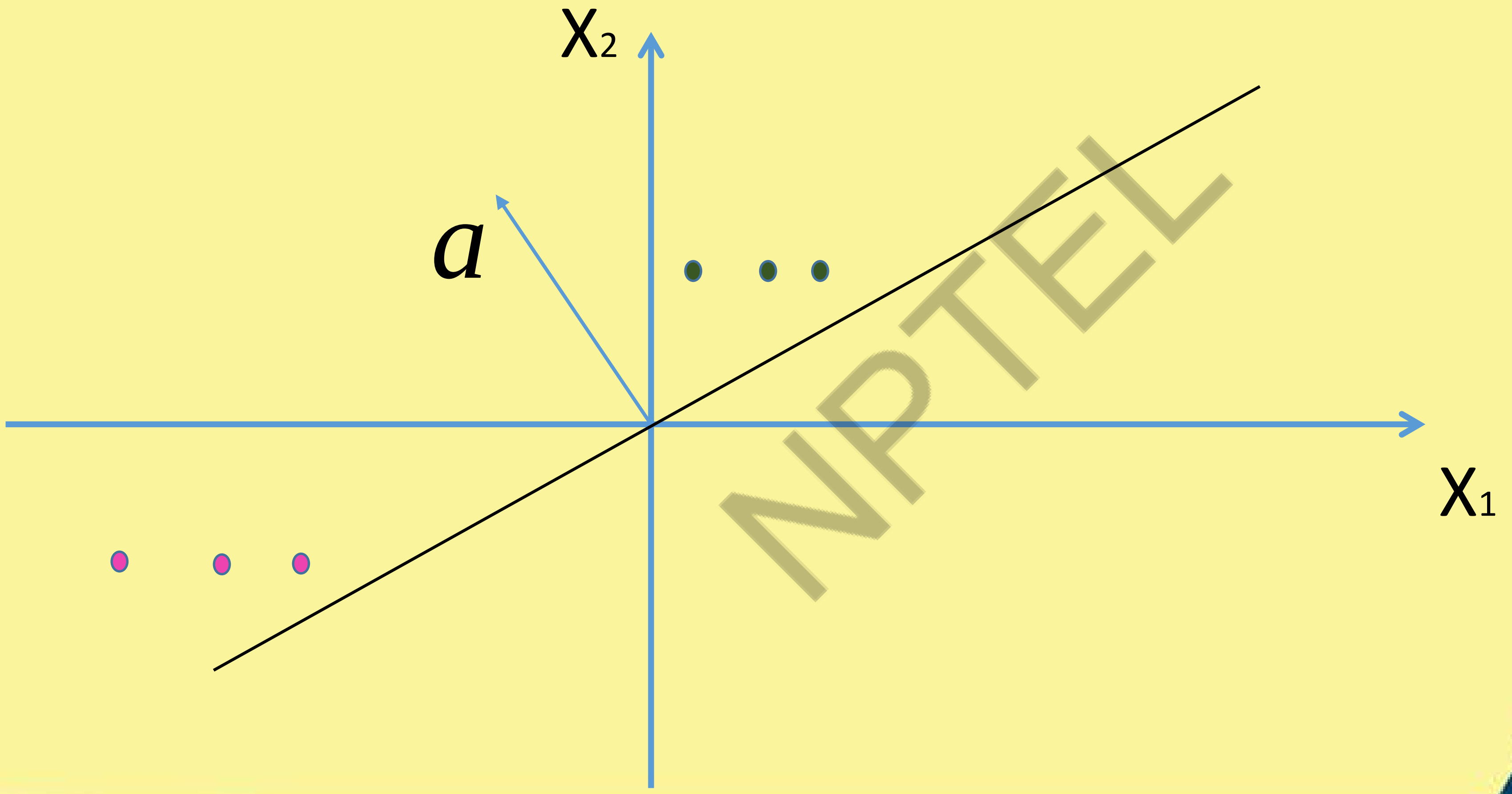
$If\ a^t X < 0;\ for\ any\ X\ irrespective\ of\ class$

$\Rightarrow "a"\ misclassifies\ that\ particular\ X$

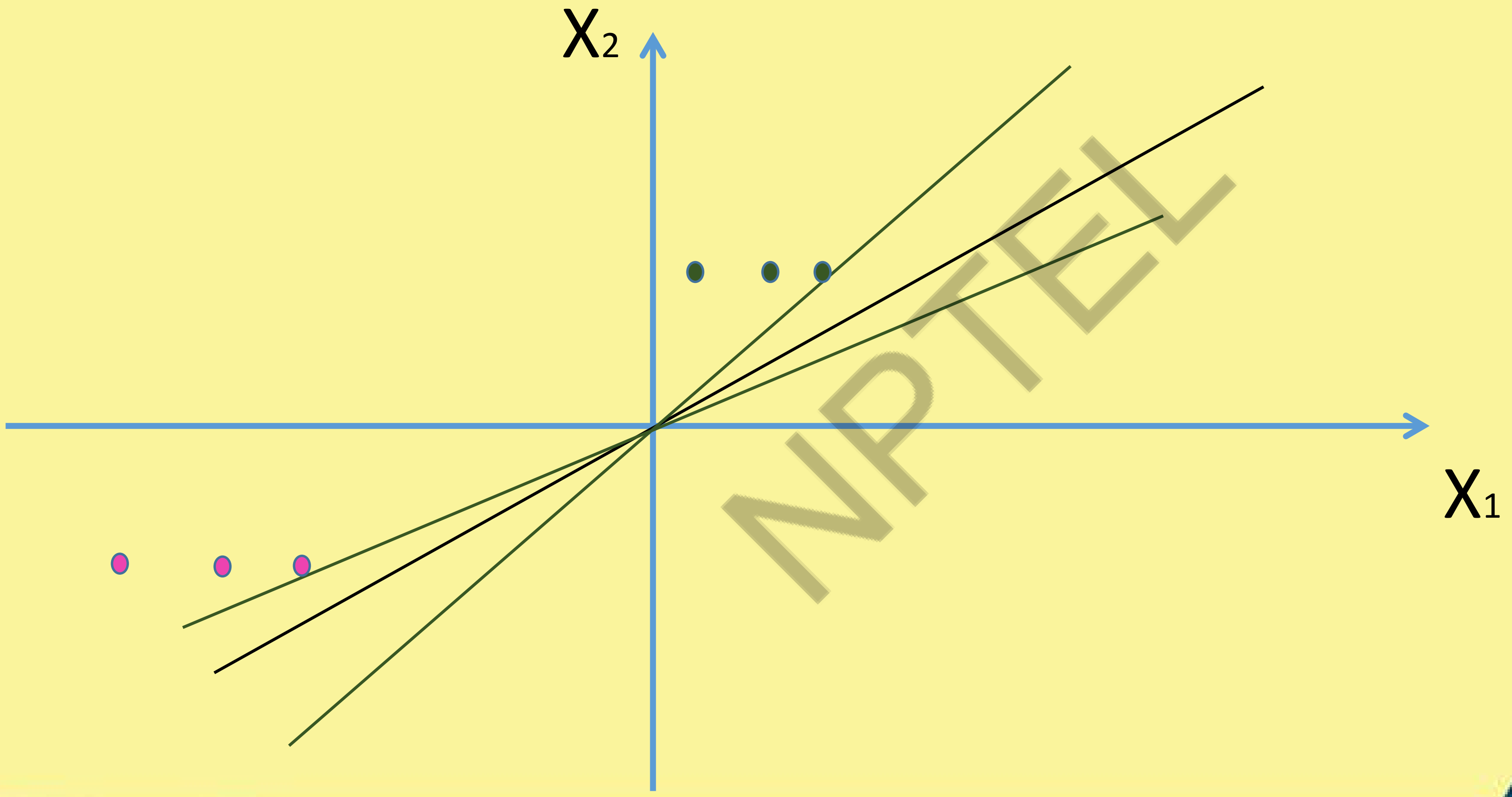
Linear Classifier – 2 Class Problem



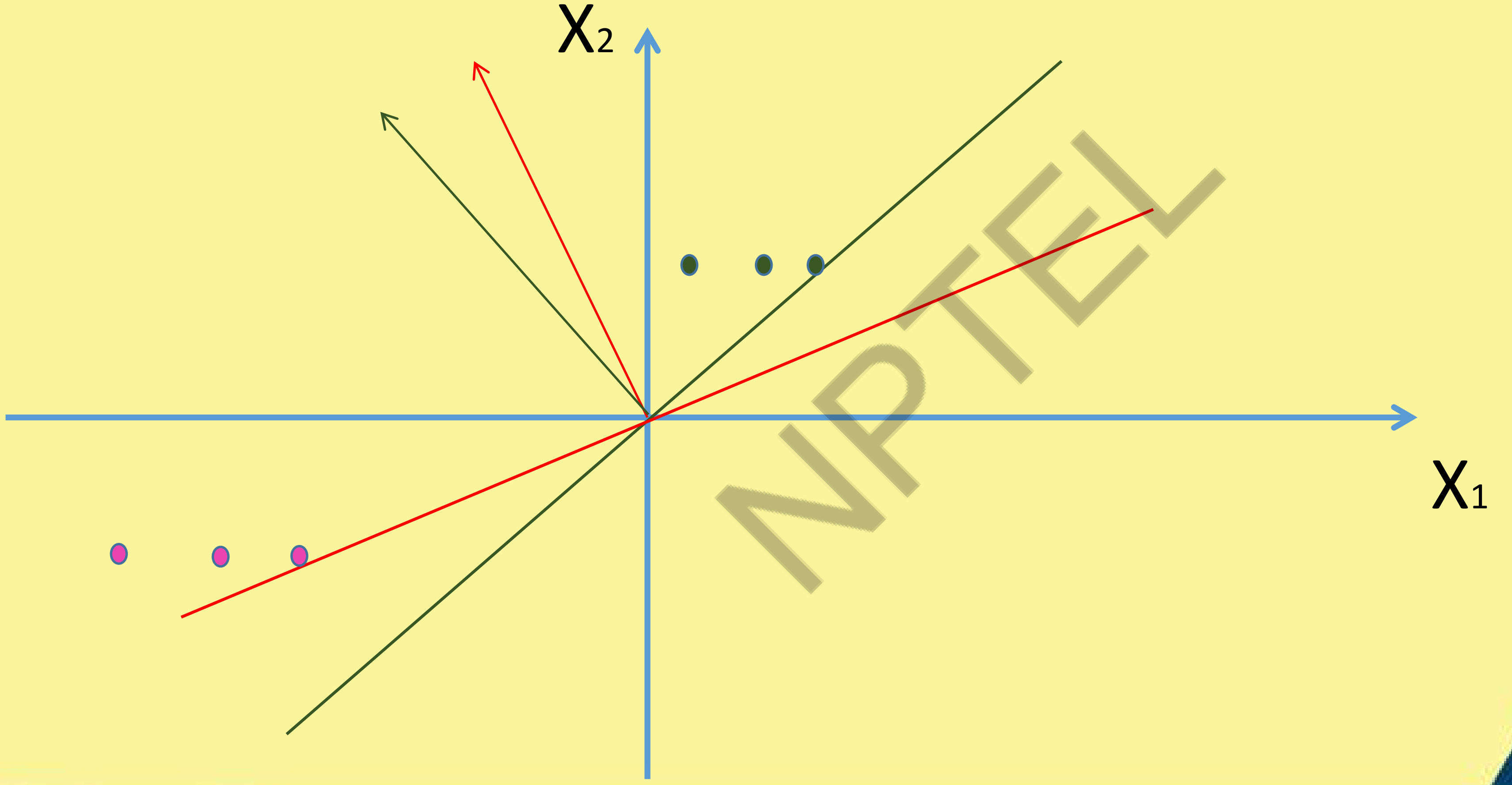
Linear Classifier – 2 Class Problem



Linear Classifier – 2 Class Problem



Linear Classifier – 2 Class Problem



Linear Classifier – Learning

$$Any\ "a"\ misclassifies\ X \Rightarrow a^t X < 0$$

$$This\ leads\ to\ an\ error:\quad J_P(a) = \sum_{\forall Y:\,misclassified} -a^t X$$

$$Follow\ Gradient\ Descent\ Algorithm$$

$$a \leftarrow a - \eta\ \nabla_a J_P(a)$$

Linear Classifier – Learning

Perceptron Criteria

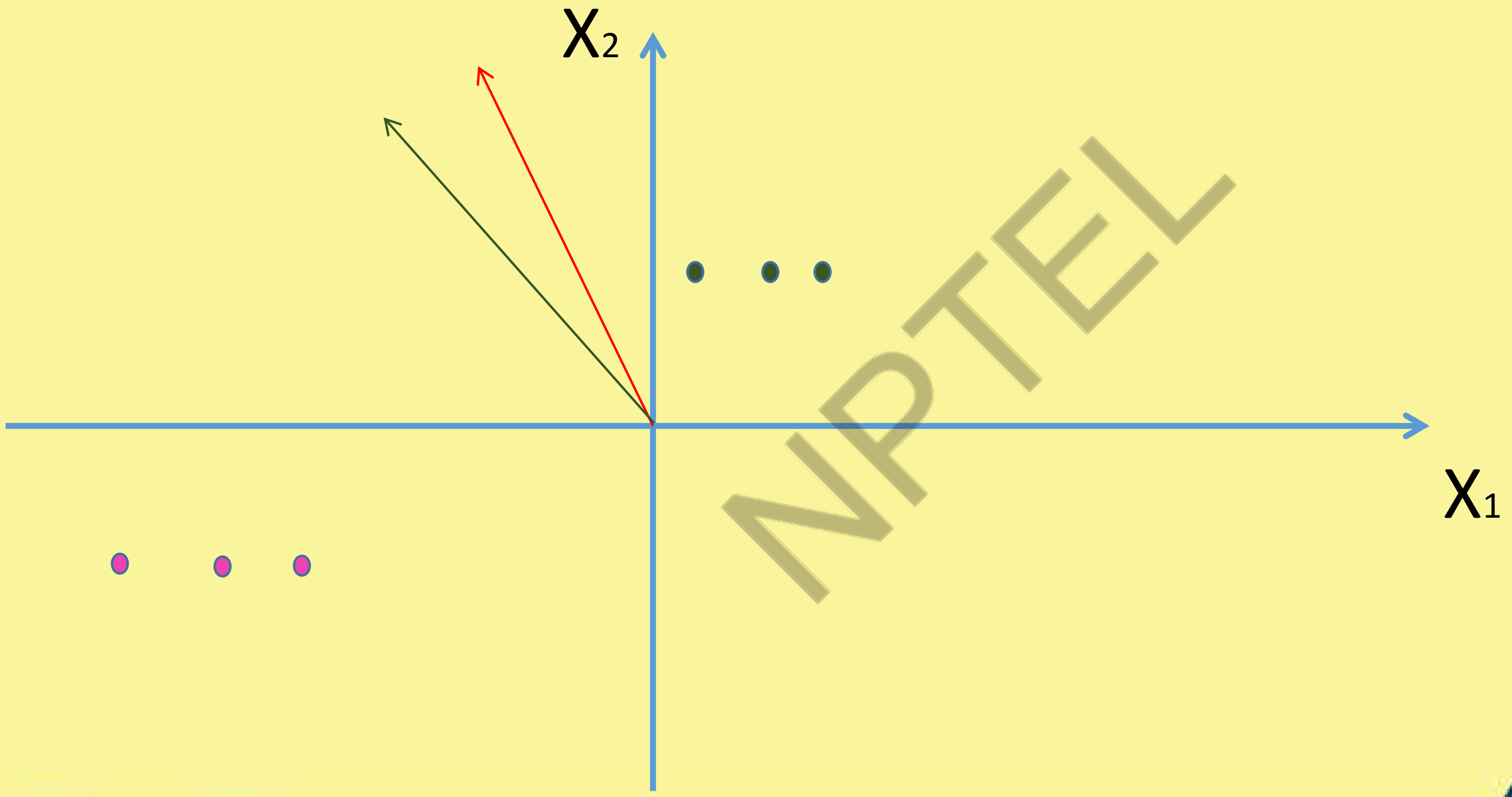
$$J_P(a) = \sum_{\forall X : misclassified} -a^t X \quad \Rightarrow \quad \nabla_a J_p(a) = -\sum_{\forall X : misclassified} X$$

Weight Updation Rule

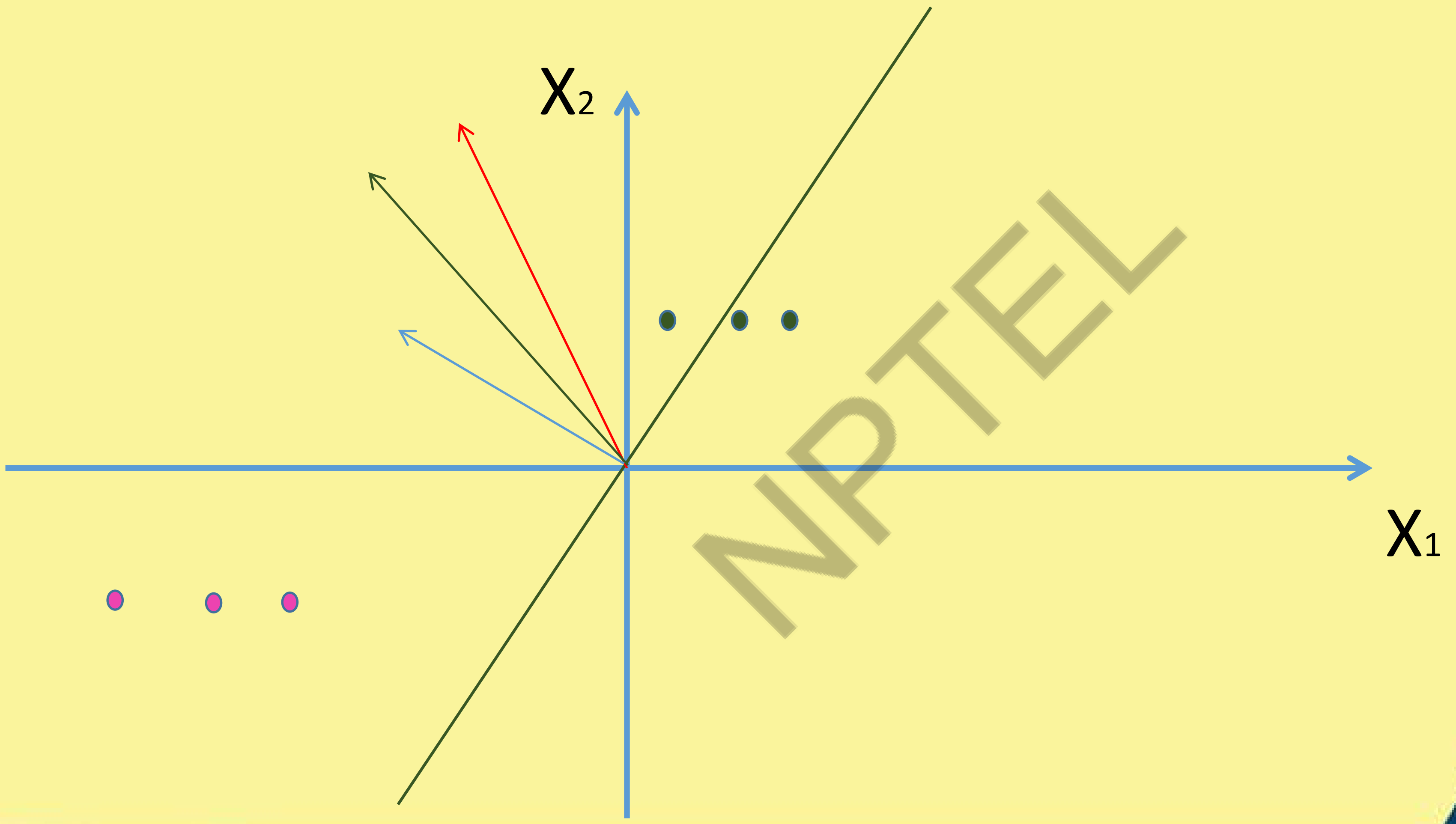
$$a(0) \leftarrow Random$$

$$a(k+1) \leftarrow a(k) + \eta \sum_{\forall X : Misclassified} X$$

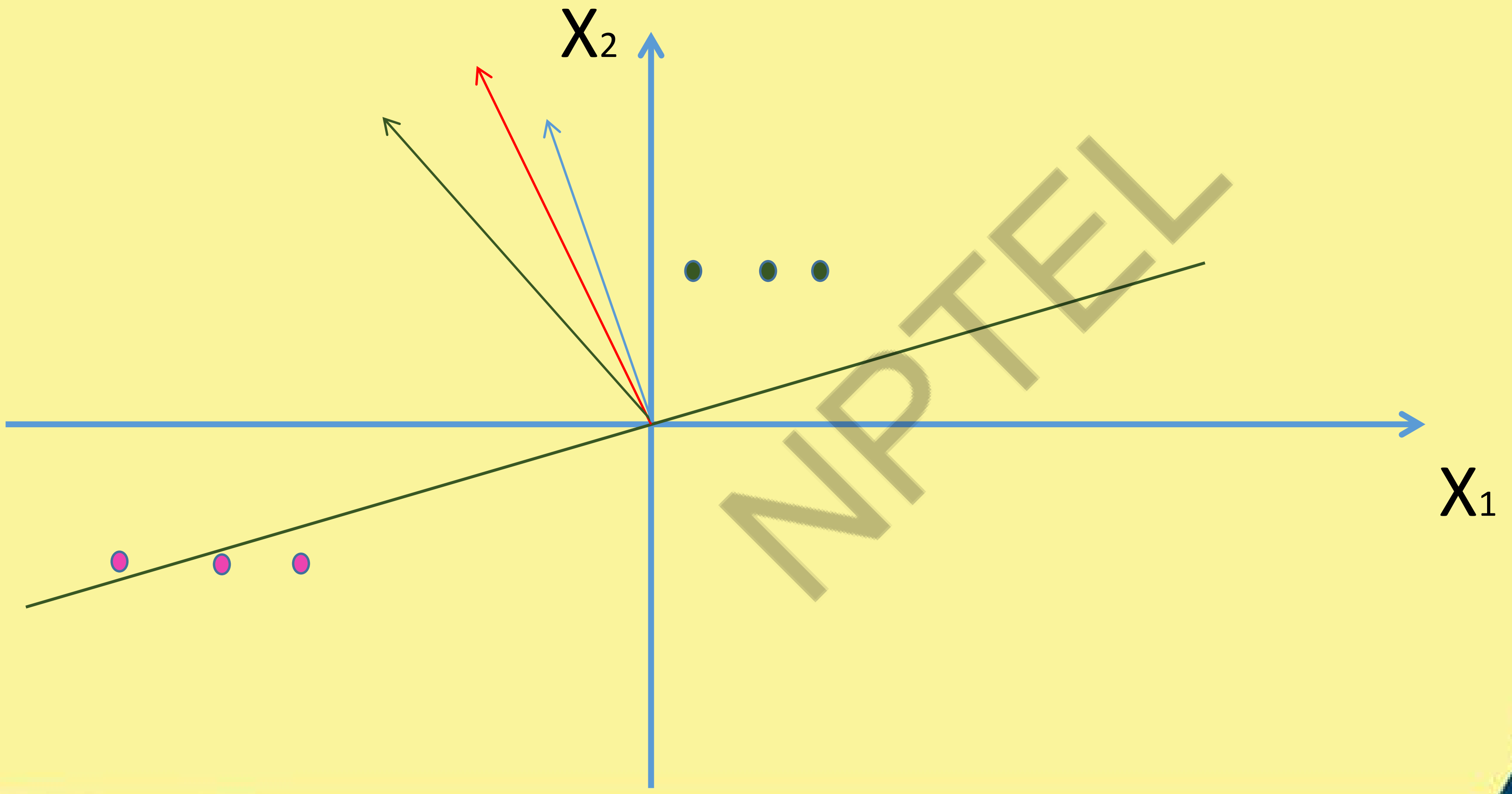
Linear Classifier – 2 Class Problem



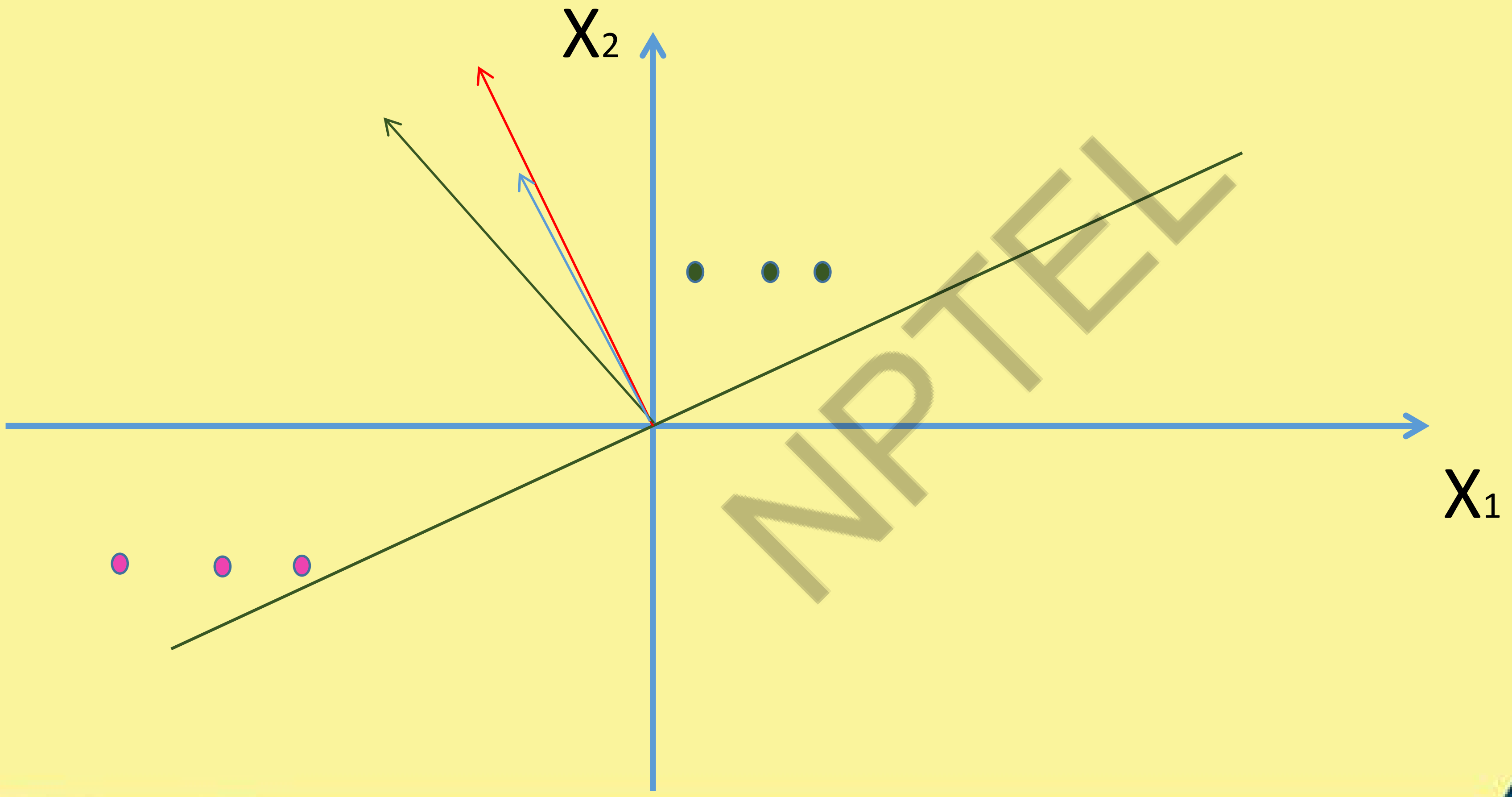
Linear Classifier – 2 Class Problem



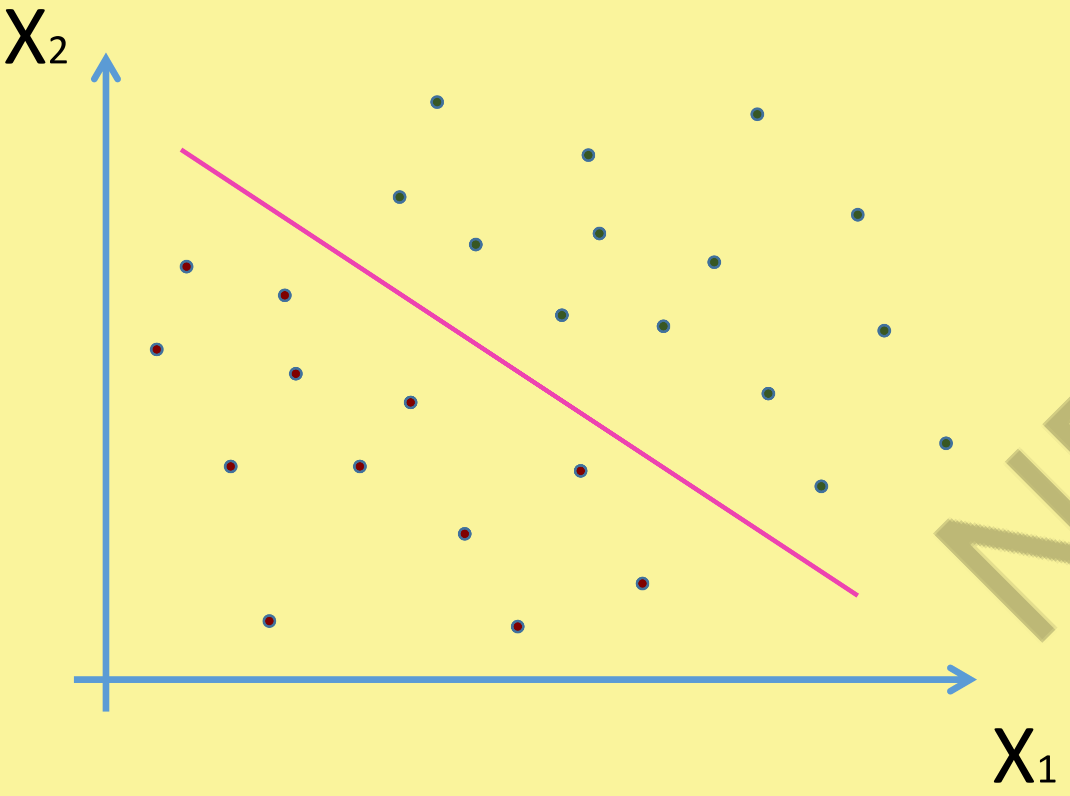
Linear Classifier – 2 Class Problem



Linear Classifier – 2 Class Problem



Support Vector Machine

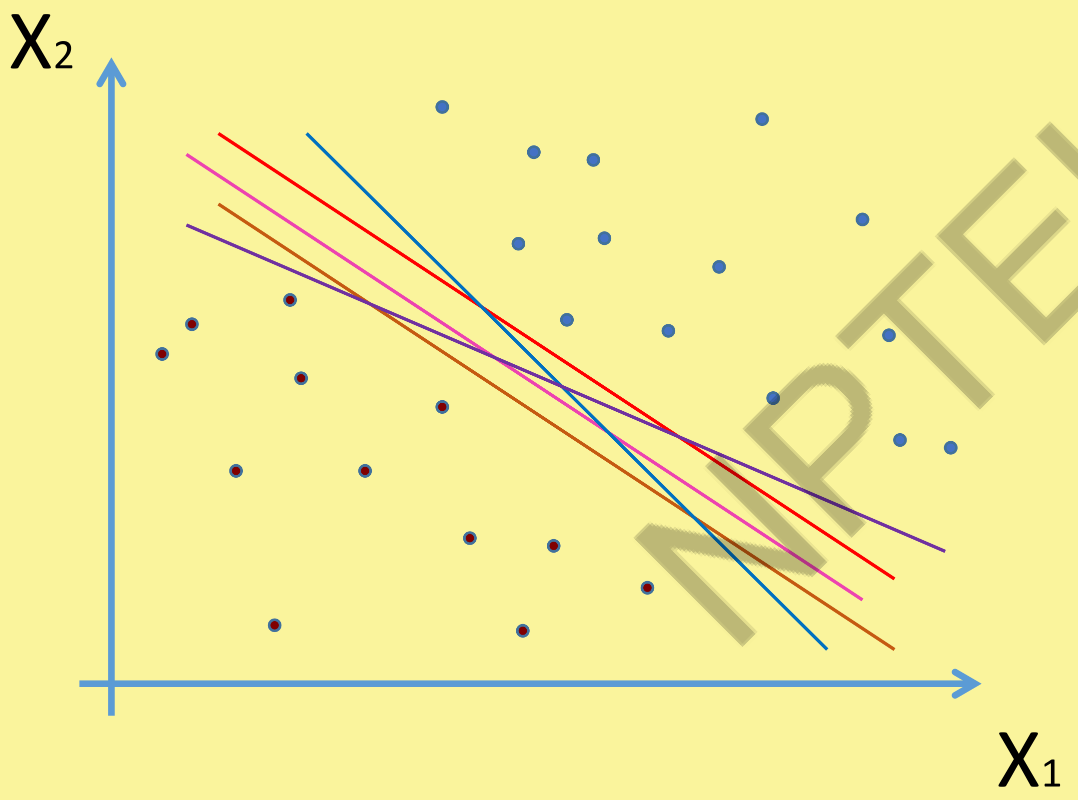


$$a^t X + b = 0$$

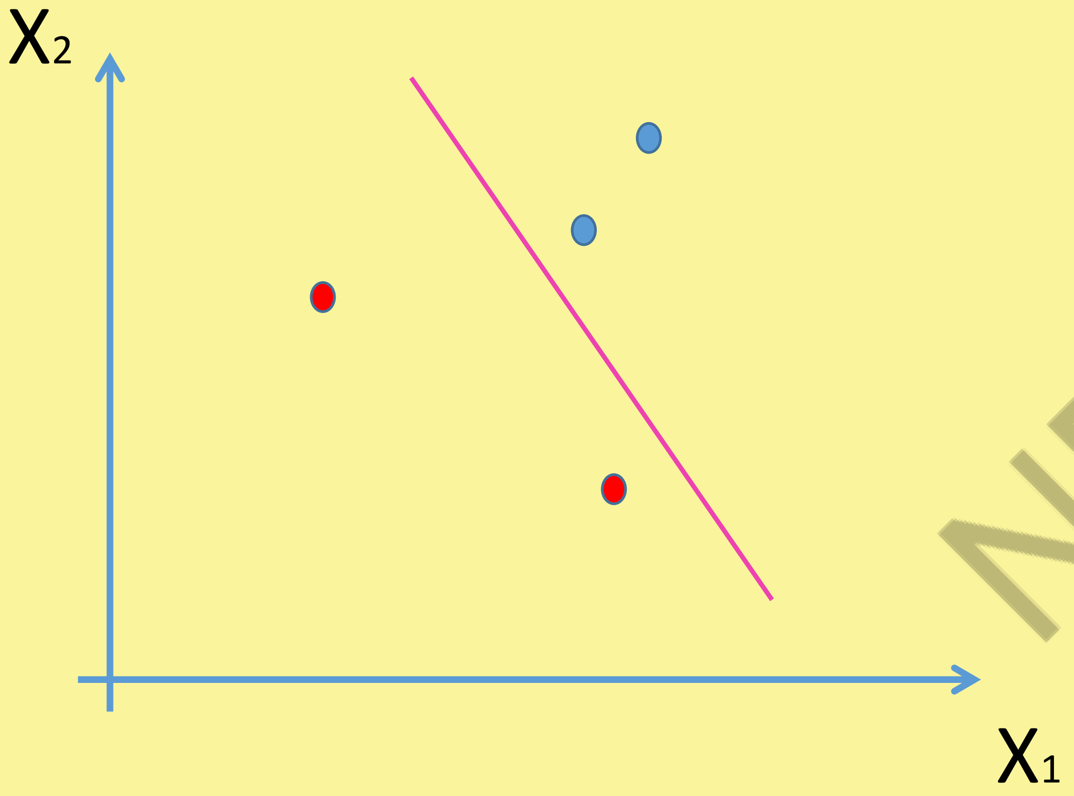
$$For\ X \in \omega_1 : \quad a^t X + b > 0$$

$$For\ X \in \omega_2 : \quad a^t X + b < 0$$

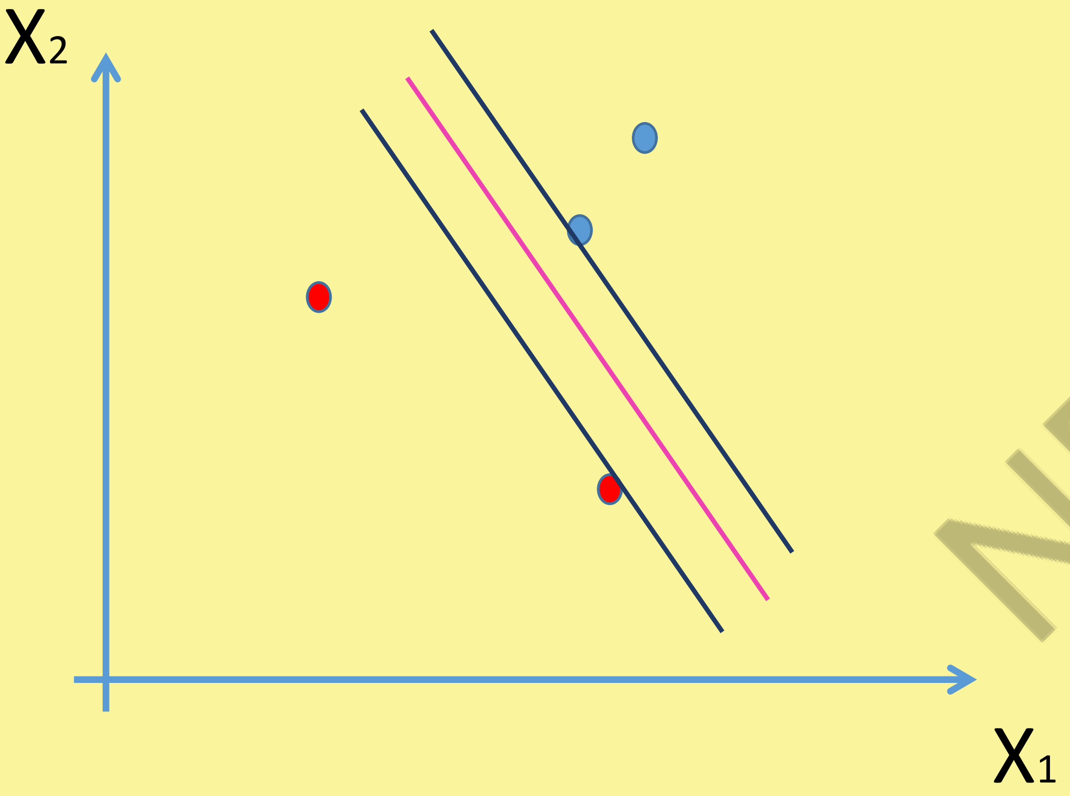
Support Vector Machine



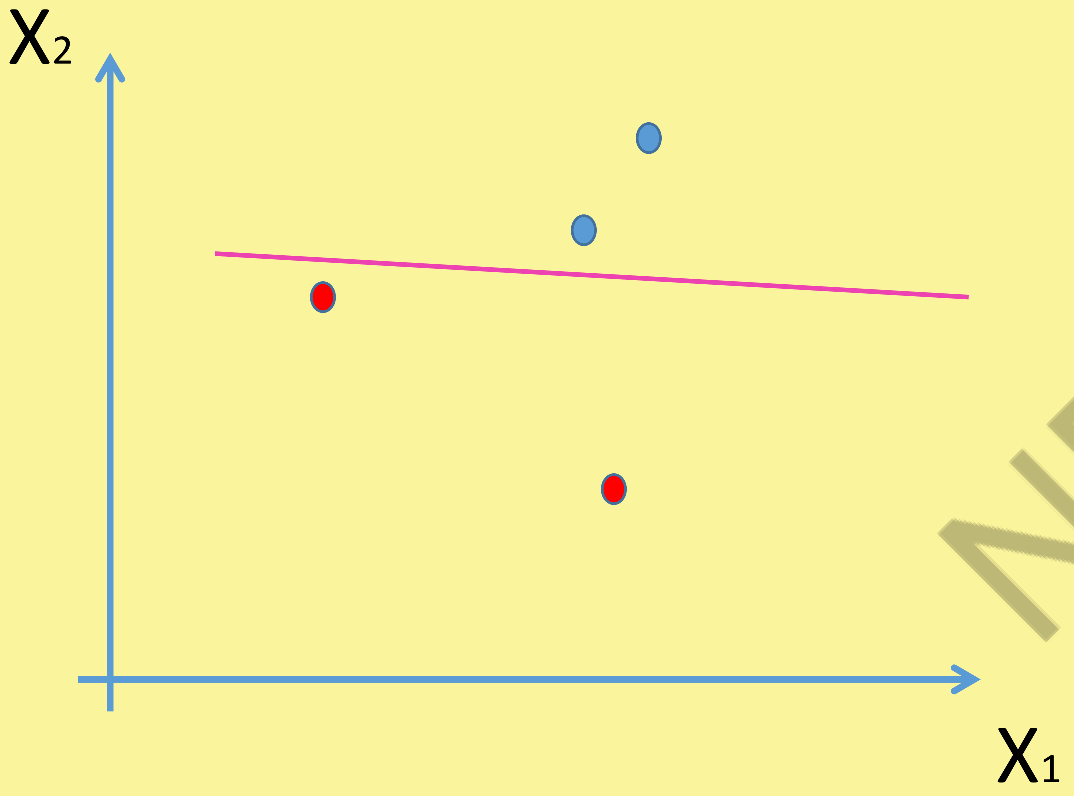
Support Vector Machine



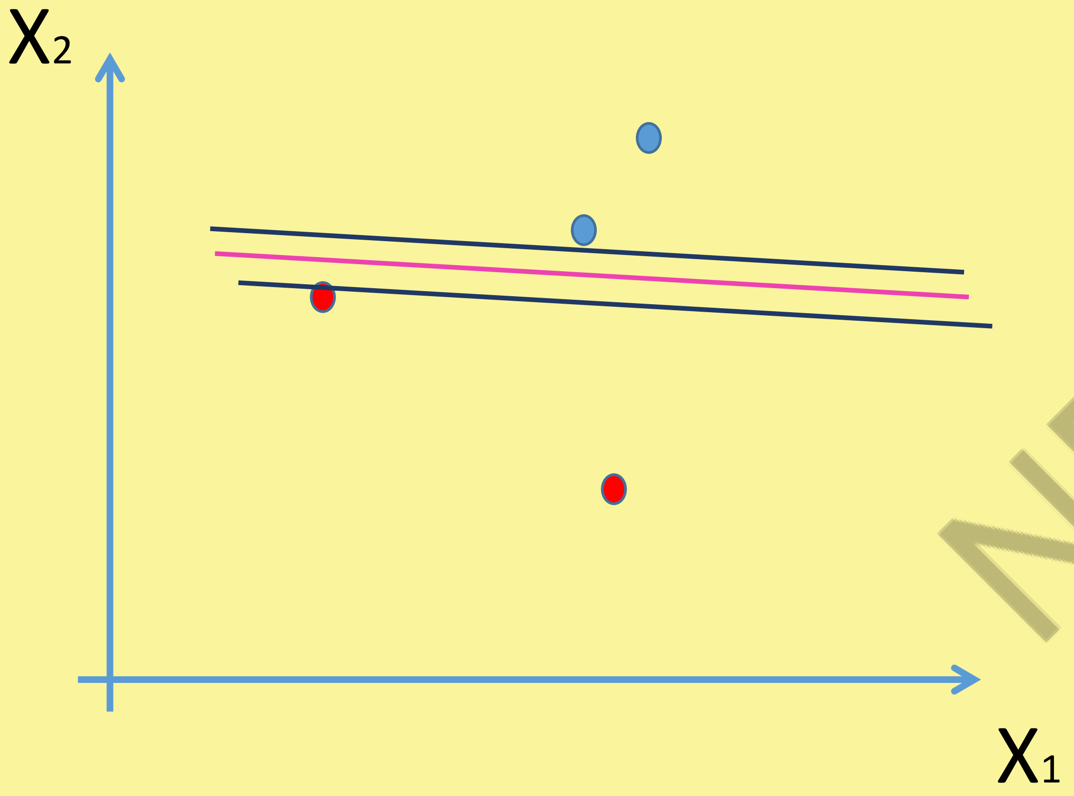
Support Vector Machine



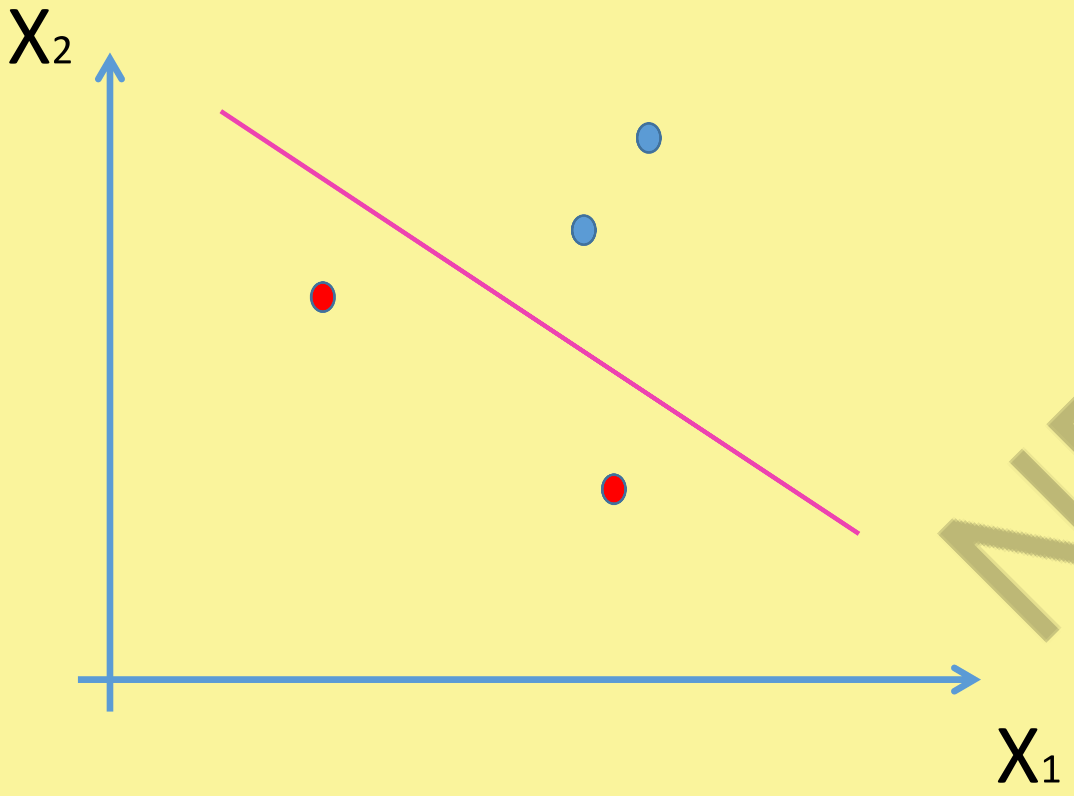
Support Vector Machine



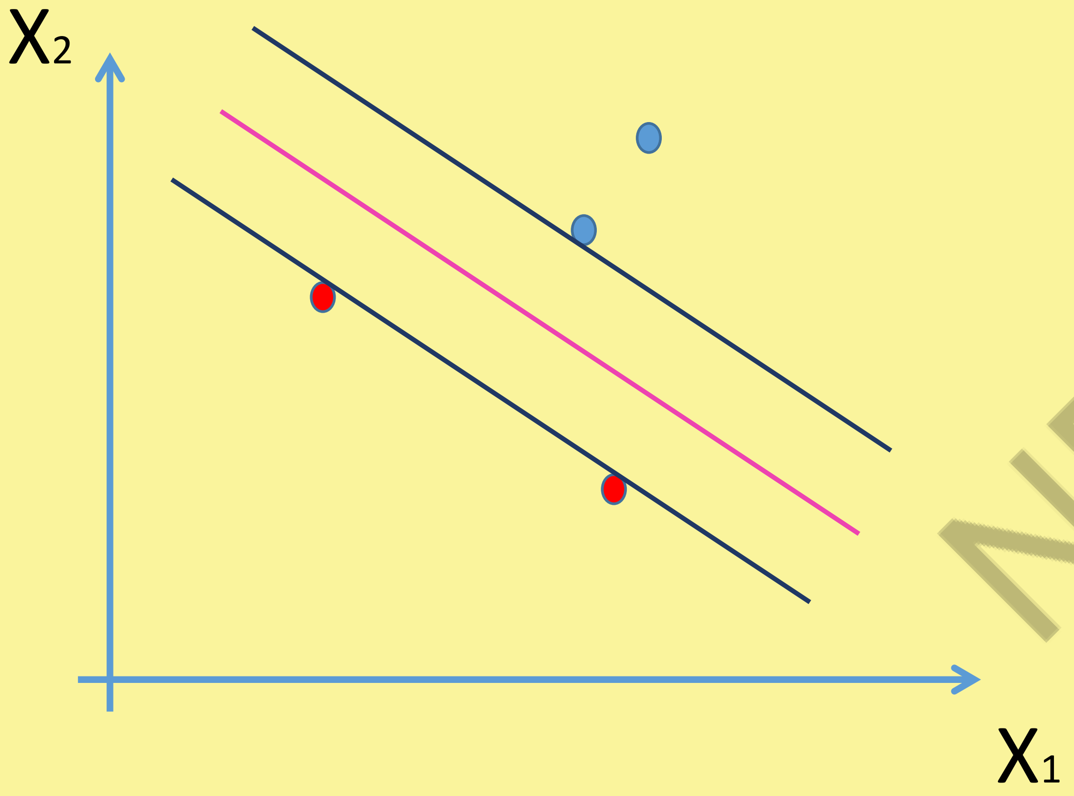
Support Vector Machine



Support Vector Machine



Support Vector Machine



NPTEL ONLINE CERTIFICATION COURSES

Thank you

NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas

Department : E & ECE, IIT Kharagpur

Topic

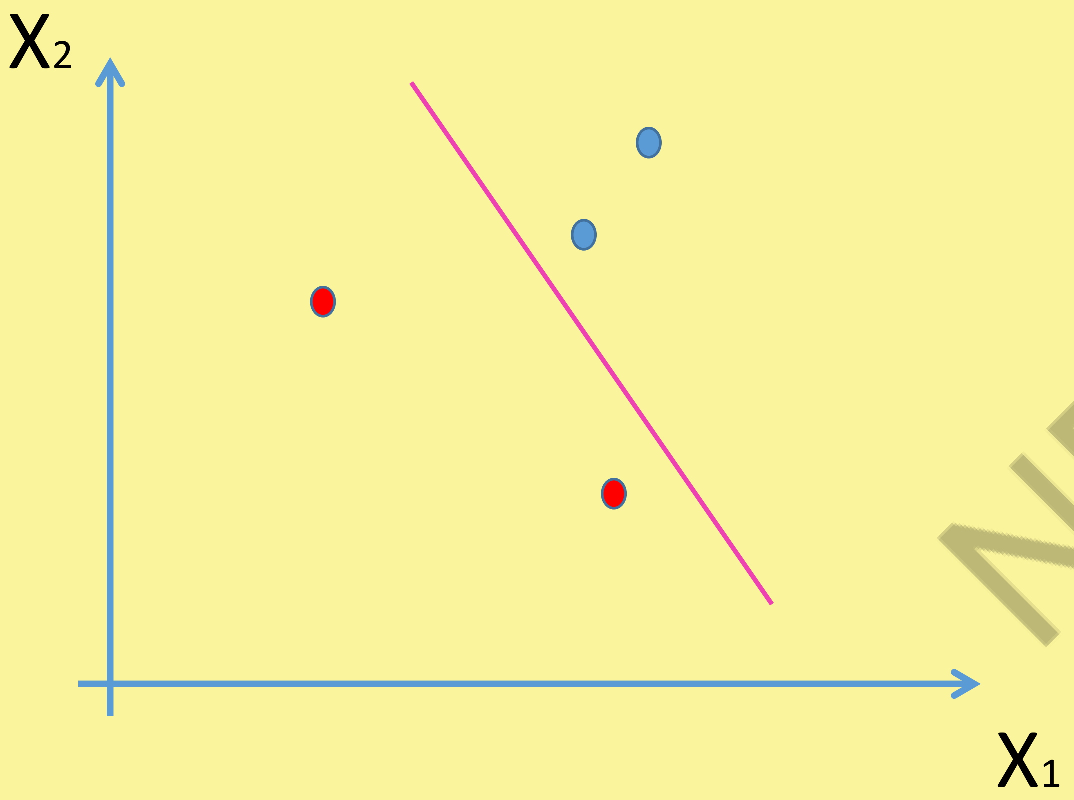
Lecture 12: Support Vector Machine II

CONCEPTS COVERED

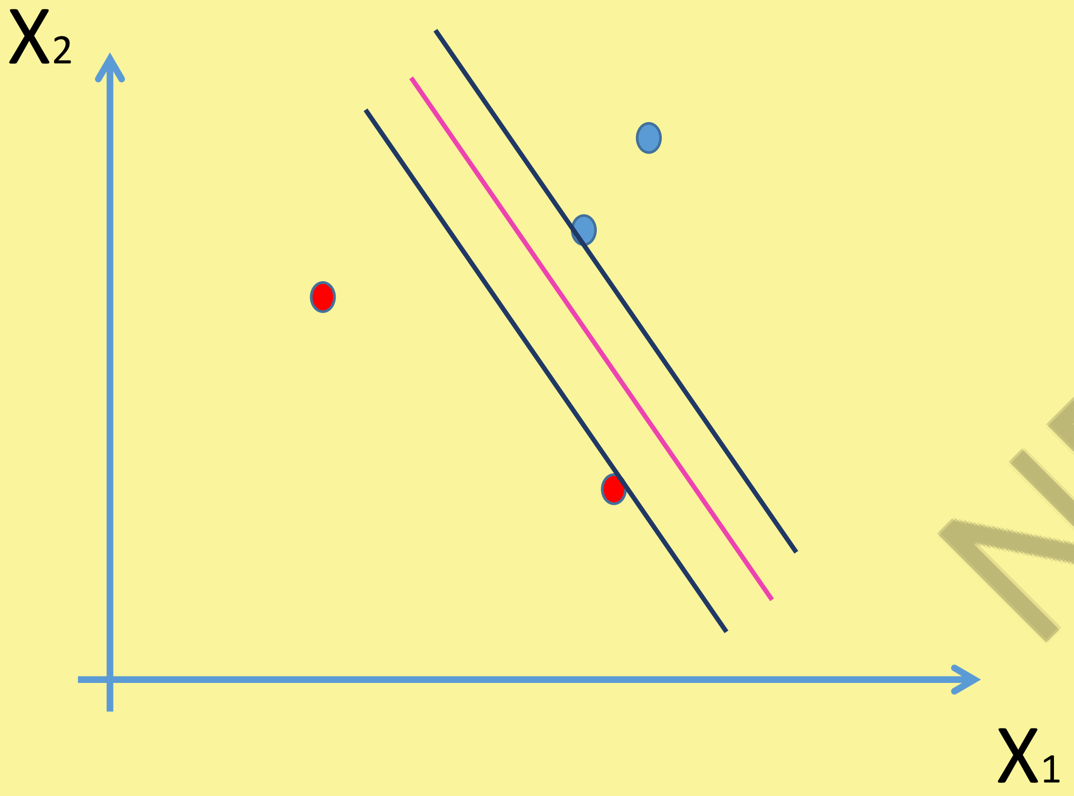
Concepts Covered:

- Support Vector Machine (SVM)
- Support Vector Machine Design

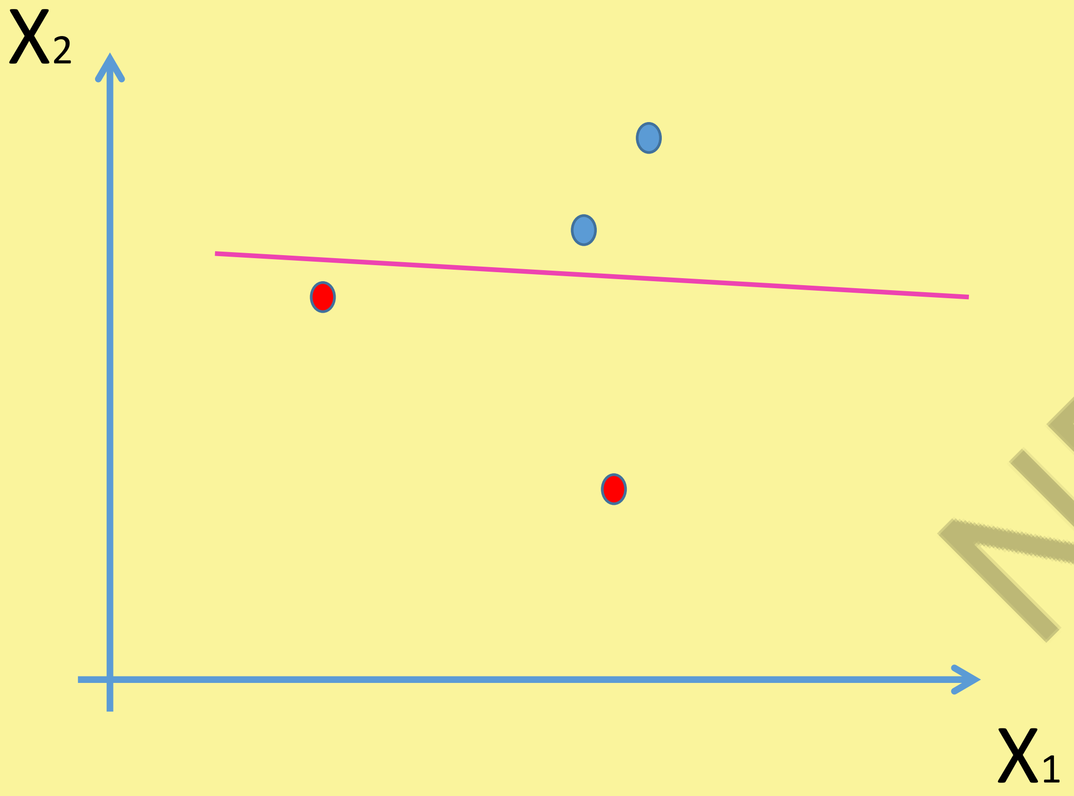
Support Vector Machine



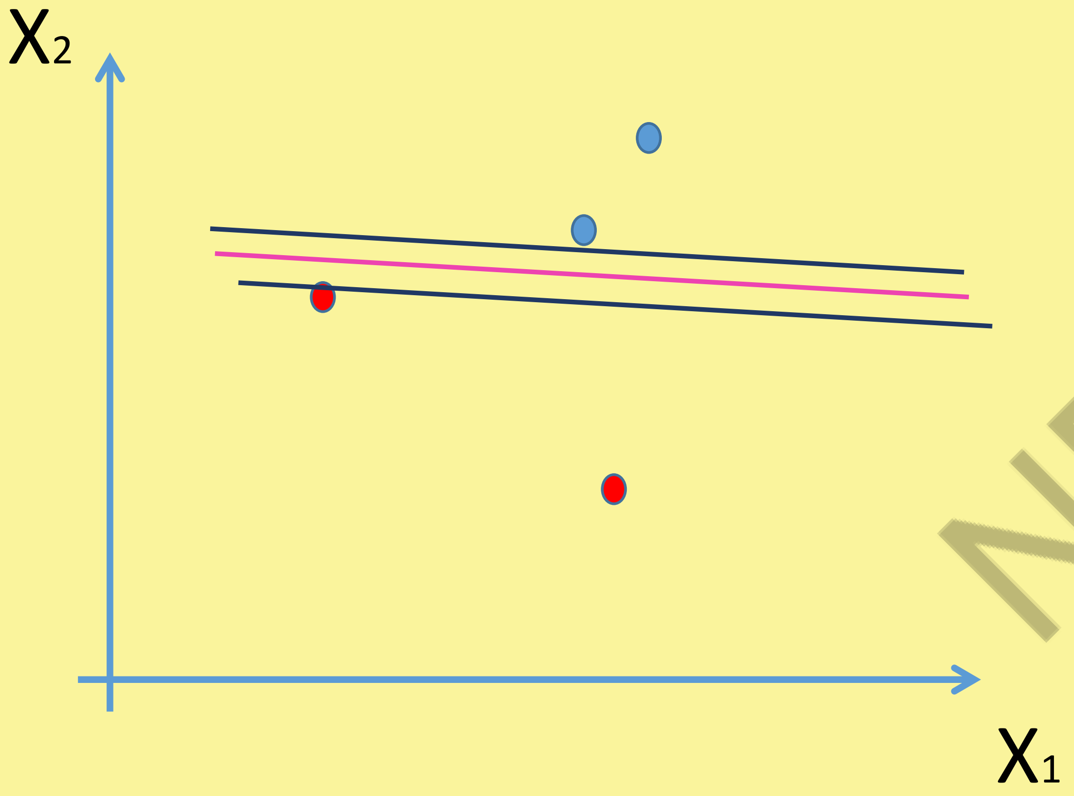
Support Vector Machine



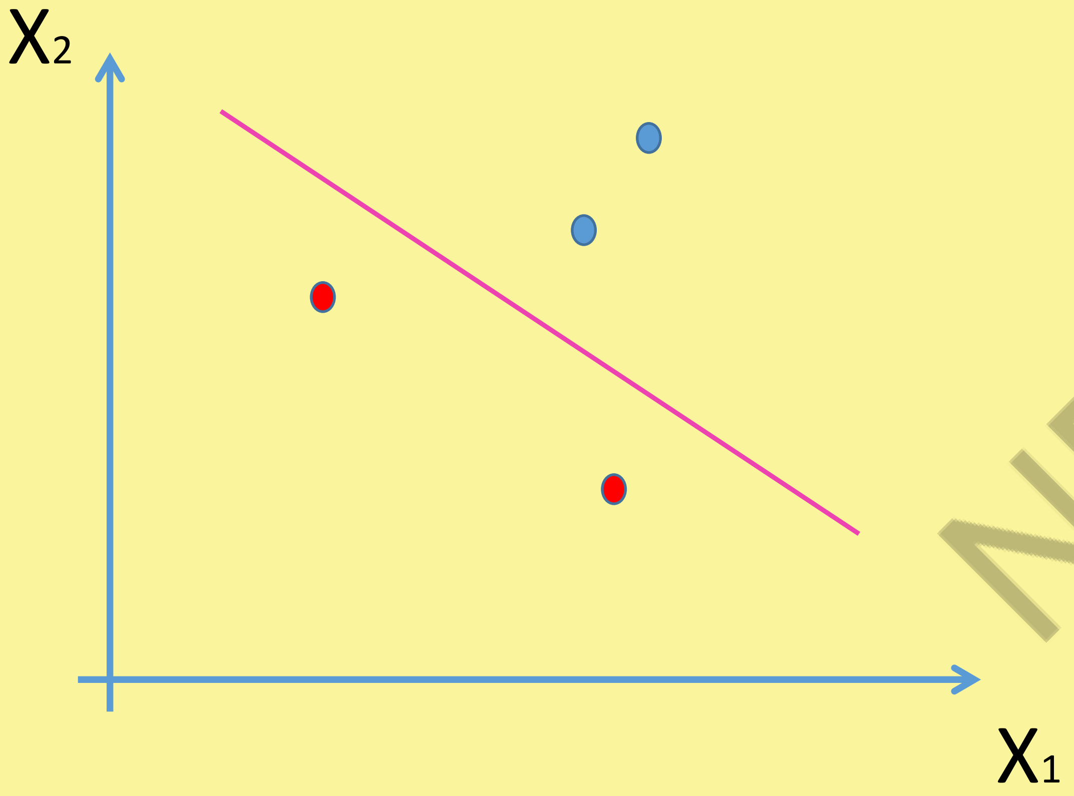
Support Vector Machine



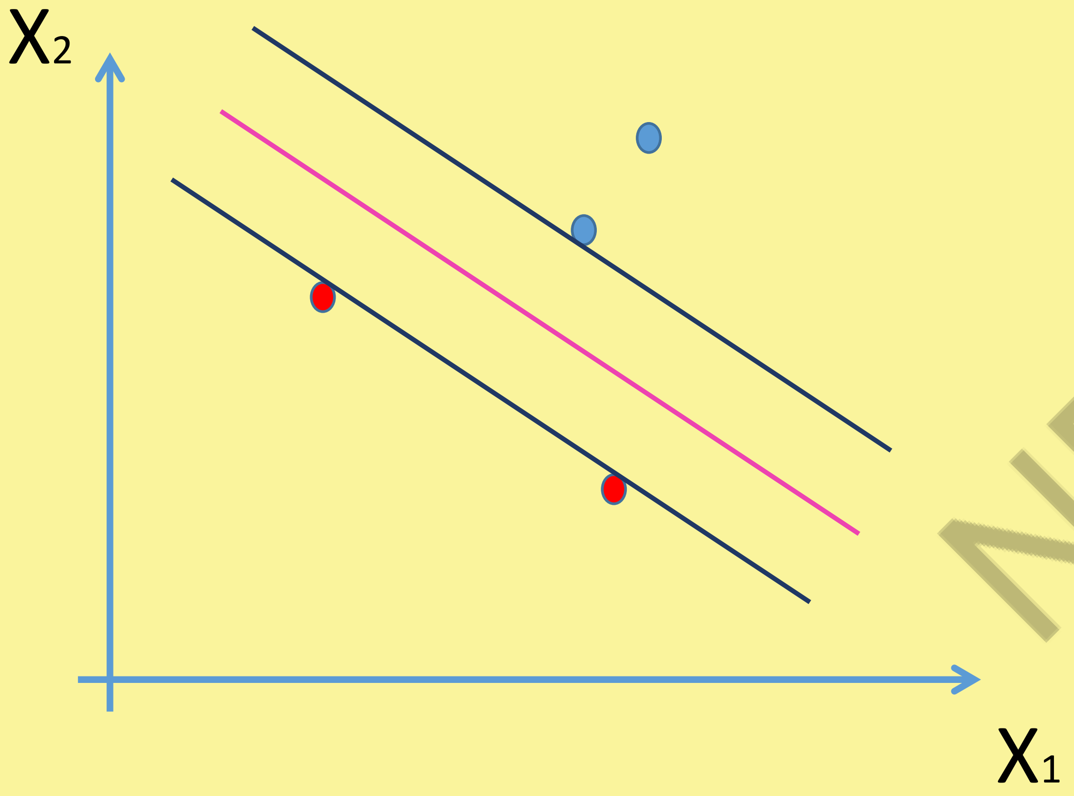
Support Vector Machine



Support Vector Machine



Support Vector Machine



NPTEL ONLINE CERTIFICATION COURSES

Thank you

NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 13: Linear Machine

CONCEPTS COVERED

Concepts Covered:

- Linear Classifier
- Support Vector Machine
- Linear Machine
- Multiclass Support Vector Machine

Multiclass Problem: Linear Machine

Linear Classifier & SVM

Multiclass Problem: Linear Machine

Multiclass Problem: Linear Machine

$$\begin{bmatrix} 0.2 & 0.6 & -1.0 & 0.8 \\ 1.5 & 0.9 & 3.1 & 0.1 \\ 0.5 & 1.1 & 0.7 & 0.0 \\ 2.1 & 0.3 & 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 45 \\ 110 \\ 21 \\ 16 \end{bmatrix} + \begin{bmatrix} 1.1 \\ 5.3 \\ -2.1 \\ 0.6 \end{bmatrix} \Rightarrow \begin{array}{cc} \begin{bmatrix} 67.9 \\ 238.5 \\ 156.1 \\ 140.3 \end{bmatrix} & \begin{array}{c} \textit{Cat score} \\ \textit{Bird score} \\ \textit{Dog score} \\ \textit{Car score} \end{array} \end{array}$$

$$f(X_i, W, b)$$

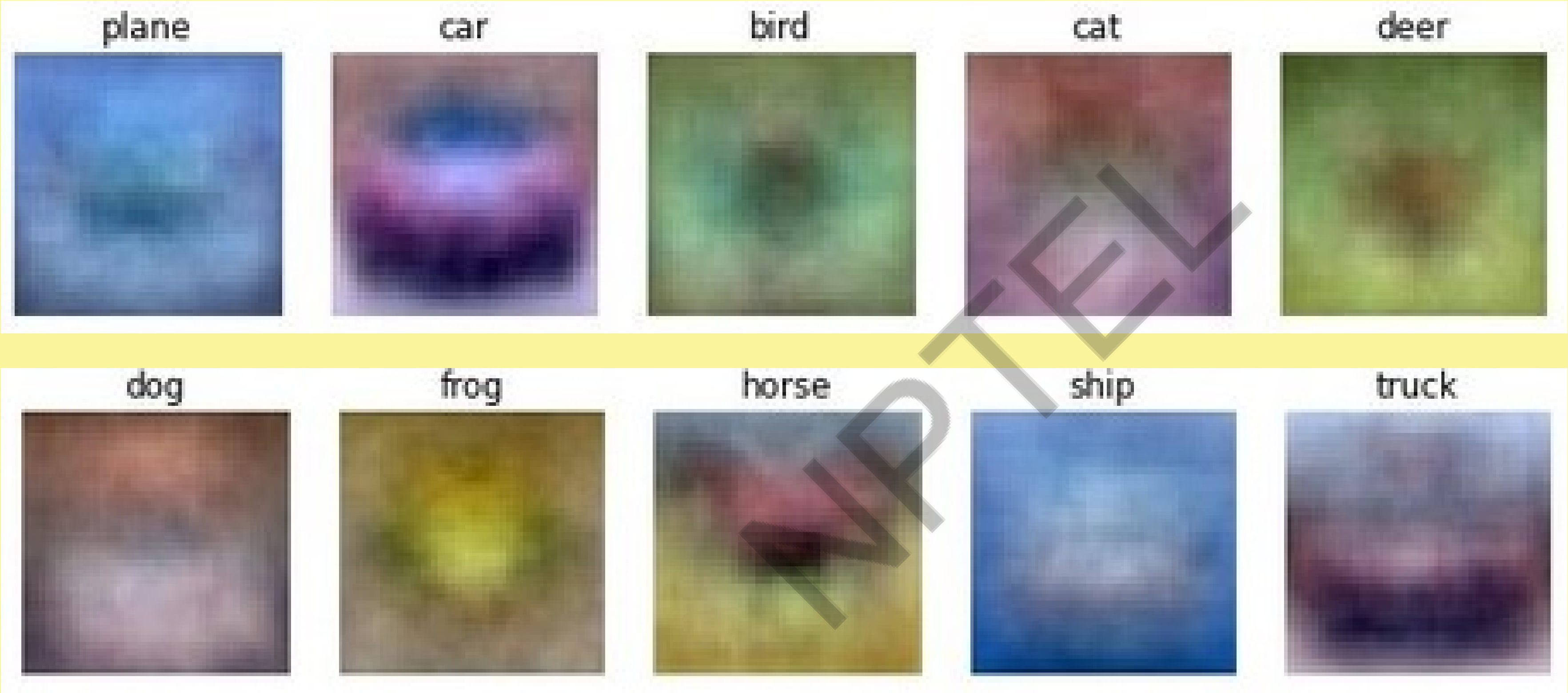
Interpretation

$$\begin{bmatrix} 0.2 & 0.6 & -1.0 & 0.8 \\ 1.5 & 0.9 & 3.1 & 0.1 \\ 0.5 & 1.1 & 0.7 & 0.0 \\ 2.1 & 0.3 & 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 45 \\ 110 \\ 21 \\ 16 \end{bmatrix} + \begin{bmatrix} 1.1 \\ 5.3 \\ -2.1 \\ 0.6 \end{bmatrix} \Rightarrow \begin{array}{cl} \begin{bmatrix} 67.9 \\ 238.5 \\ 156.1 \\ 140.3 \end{bmatrix} & \begin{array}{l} \textit{Cat score} \\ \textit{Bird score} \\ \textit{Dog score} \\ \textit{Car score} \end{array} \end{array}$$

$$f(X_i, W, b)$$

Interpretation

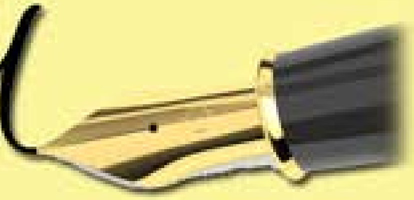
Multiclass Problem: Linear Machine



Source - http://cs231n.github.io

NPTEL ONLINE CERTIFICATION COURSES

Thank you



NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 14: Linear Machine and Multiclass SVM

CONCEPTS COVERED

Concepts Covered:

- Linear Classifier
- Support Vector Machine
- Linear Machine
- Multiclass Support Vector Machine

Multiclass Problem: Linear Machine

Multiclass Problem: Linear Machine

$$f : R^D \rightarrow R^K$$

$$f(X_i, W, b) = WX_i + b = s$$

$$\begin{bmatrix} W_{11} & W_{12} & W_{13} & .. & W_{1D} \\ W_{21} & W_{22} & W_{23} & .. & W_{2D} \\ .. & .. & .. & .. & .. \\ W_{K1} & W_{K2} & W_{K3} & .. & W_{KD} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ .. \\ X_D \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ .. \\ b_K \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ .. \\ s_K \end{bmatrix}$$

Multiclass Problem: Linear Machine

$$\begin{bmatrix} 0.2 & 0.6 & -1.0 & 0.8 \\ 1.5 & 0.9 & 3.1 & 0.1 \\ 0.5 & 1.1 & 0.7 & 0.0 \\ 2.1 & 0.3 & 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 45 \\ 110 \\ 21 \\ 16 \end{bmatrix} + \begin{bmatrix} 1.1 \\ 5.3 \\ -2.1 \\ 0.6 \end{bmatrix} \Rightarrow \begin{array}{cc} \begin{bmatrix} 67.9 \\ 238.5 \\ 156.1 \\ 140.3 \end{bmatrix} & \begin{array}{l} \textit{Cat score} \\ \textit{Bird score} \\ \textit{Dog score} \\ \textit{Car score} \end{array} \end{array}$$

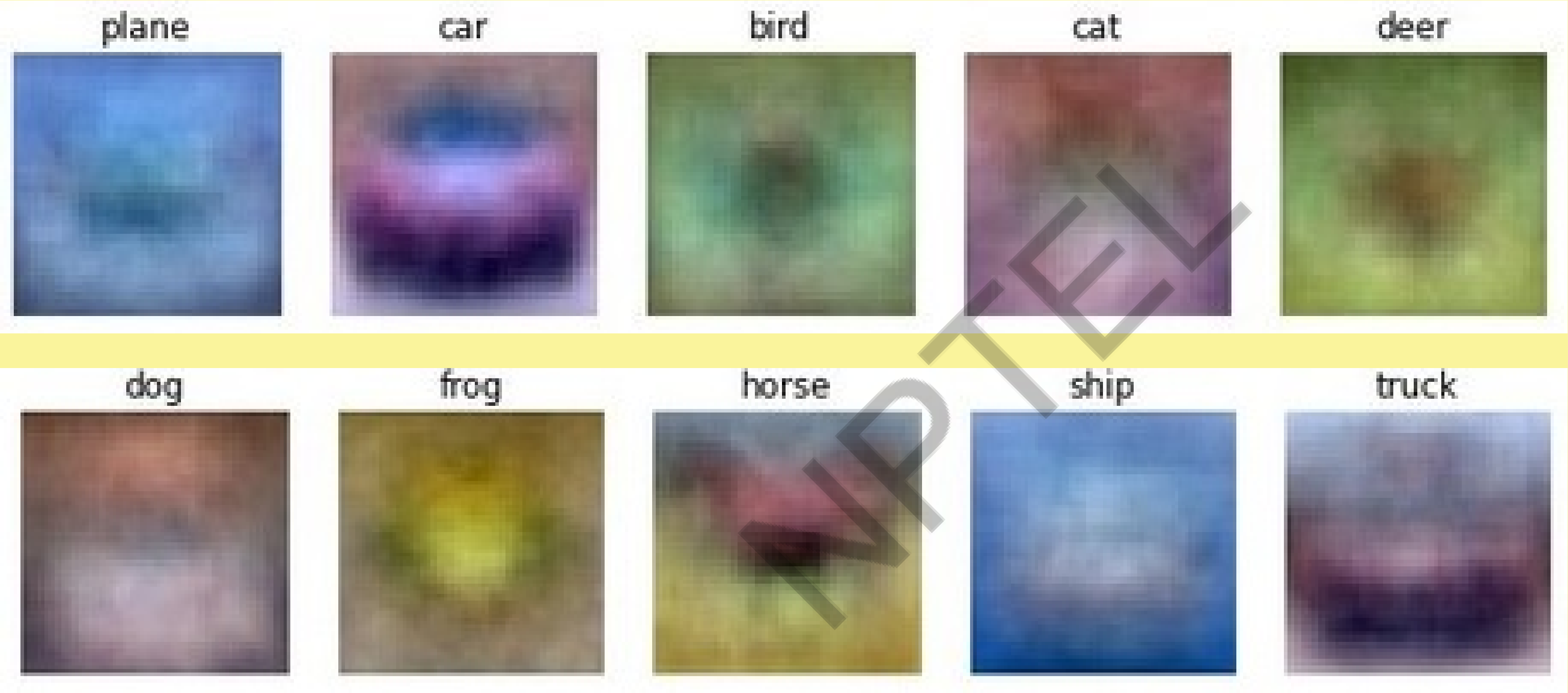
$$f(X_i, W, b)$$

Interpretation

$$\begin{bmatrix} 0.2 & 0.6 & -1.0 & 0.8 \\ 1.5 & 0.9 & 3.1 & 0.1 \\ 0.5 & 1.1 & 0.7 & 0.0 \\ 2.1 & 0.3 & 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 45 \\ 110 \\ 21 \\ 16 \end{bmatrix} + \begin{bmatrix} 1.1 \\ 5.3 \\ -2.1 \\ 0.6 \end{bmatrix} \Rightarrow \begin{array}{cl} \begin{bmatrix} 67.9 \\ 238.5 \\ 156.1 \\ 140.3 \end{bmatrix} & \begin{array}{l} \textit{Cat score} \\ \textit{Bird score} \\ \textit{Dog score} \\ \textit{Car score} \end{array} \end{array}$$

$$f(X_i, W, b)$$

Multiclass Problem: Linear Machine



Source - http://cs231n.github.io

Bias Trick

$$\begin{bmatrix} 0.2 & 0.6 & -1.0 & 0.8 & 1.1 \\ 1.5 & 0.9 & 3.1 & 0.1 & 5.3 \\ 0.5 & 1.1 & 0.7 & 0.0 & -2.1 \\ 2.1 & 0.3 & 0.2 & 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} 45 \\ 110 \\ 21 \\ 16 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 67.9 \\ 238.5 \\ 156.1 \\ 140.3 \end{bmatrix} \begin{array}{l} \textit{Cat score} \\ \textit{Bird score} \\ \textit{Dog score} \\ \textit{Car score} \end{array}$$

$$f(X_i, W)$$

Multiclass SVM

$$\left.\begin{aligned} s_j &= f(X_i, W)_j \\ &= WX_i \end{aligned}\right\} \rightarrow \textit{Score for } j^{th} \textit{ Class of } i^{th} \textit{ Vector } (X_i, y_i)$$

$$s_{y_i} = f(X_i, W)_{y_i} \quad \rightarrow \textit{should be} \max \textit{imum}$$

$$s_{y_i} - s_j \geq \Delta$$

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \Delta)$$

Loss Function: An Example

$$For\ some\ (X_i, y_i)\ where\ y_i = 2$$

$$s = (10\ \ 30\ \ -20\ \ 25)^t \qquad \Delta = 10$$

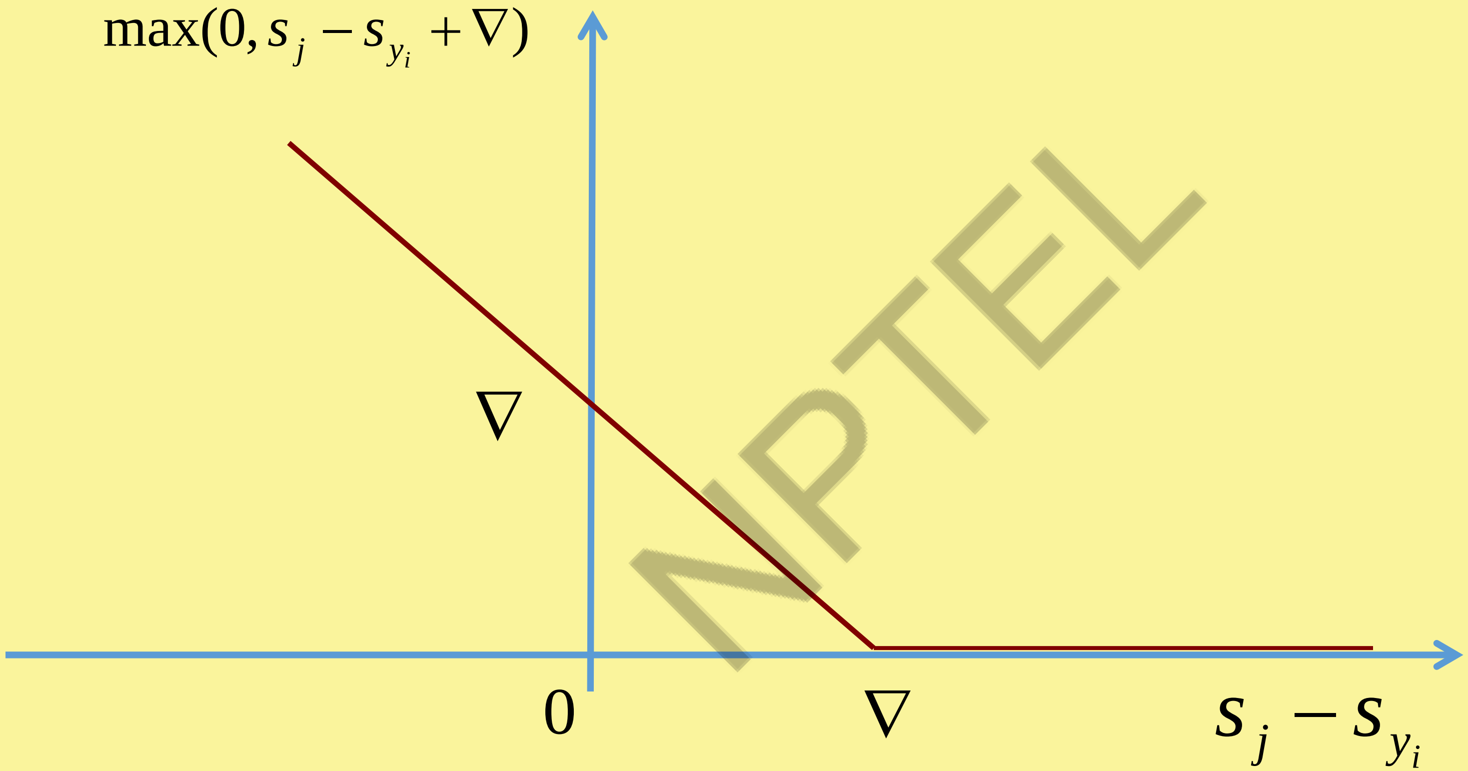
$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \Delta)$$

$$= \max(0,10-30+10) + \max(0,-20-30+10) + \max(0,25-30+10)$$

$$= 0 + 0 + 15$$

$$= 15$$

Hinge Loss



Regularization

$$s_j - s_{y_i} = W_j^t X_i - W_{y_i}^t X_i$$

$$\text{Scaling } W \text{ by } \lambda : W \leftarrow \lambda W$$

$$\Downarrow$$

$$s_j - s_{y_i} \leftarrow \lambda\left(s_j - s_{y_i}\right)$$

Regularization

Include a regularization term $R(W)$

$$R(W) = \lambda \sum_k \sum_l W_{kl}^2$$

$$L = \frac{1}{N} \sum_i L_i + \lambda R(W)$$

$$L = \frac{1}{N} \sum_i \sum_{j \neq y_i} [\max(0, f(X_i, W)_j - f(X_i, W)_{y_i} \quad + \nabla) + \lambda \sum_k \sum_l W_{kl}^2$$

NPTEL ONLINE CERTIFICATION COURSES

Thank you

NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 15: Multiclass SVM Loss Function

CONCEPTS COVERED

Concepts Covered:

- Linear Machine
- Multiclass Support Vector Machine
- Multiclass SVM Loss Function
- Optimization

Multiclass Problem: Linear Machine

$$f : R^D \rightarrow R^K$$

$$f(X_i, W, b) = WX_i + b = s$$

$$\begin{bmatrix} W_{11} & W_{12} & W_{13} & .. & W_{1D} \\ W_{21} & W_{22} & W_{23} & .. & W_{2D} \\ .. & .. & .. & .. & .. \\ W_{K1} & W_{K2} & W_{K3} & .. & W_{KD} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ .. \\ X_D \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ .. \\ b_K \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ .. \\ s_K \end{bmatrix}$$

Multiclass SVM

$$\left.\begin{aligned} s_j &= f(X_i, W)_j \\ &= WX_i \end{aligned}\right\} \rightarrow Score\ \ for\ \ j^{th}\ Class\ of\ \ i^{th}\ Vector\,(X_i, y_i)$$

$$s_{y_i} = f(X_i, W)_{y_i} \quad \rightarrow \text{should be maximum}$$

$$s_{y_i} - s_j \geq \Delta$$

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \Delta)$$

Loss Function: An Example

$$For\ some\ (X_i, y_i)\ where\ y_i = 2$$

$$s = (10\ \ 30\ \ -20\ \ 25)^t \qquad \Delta = 10$$

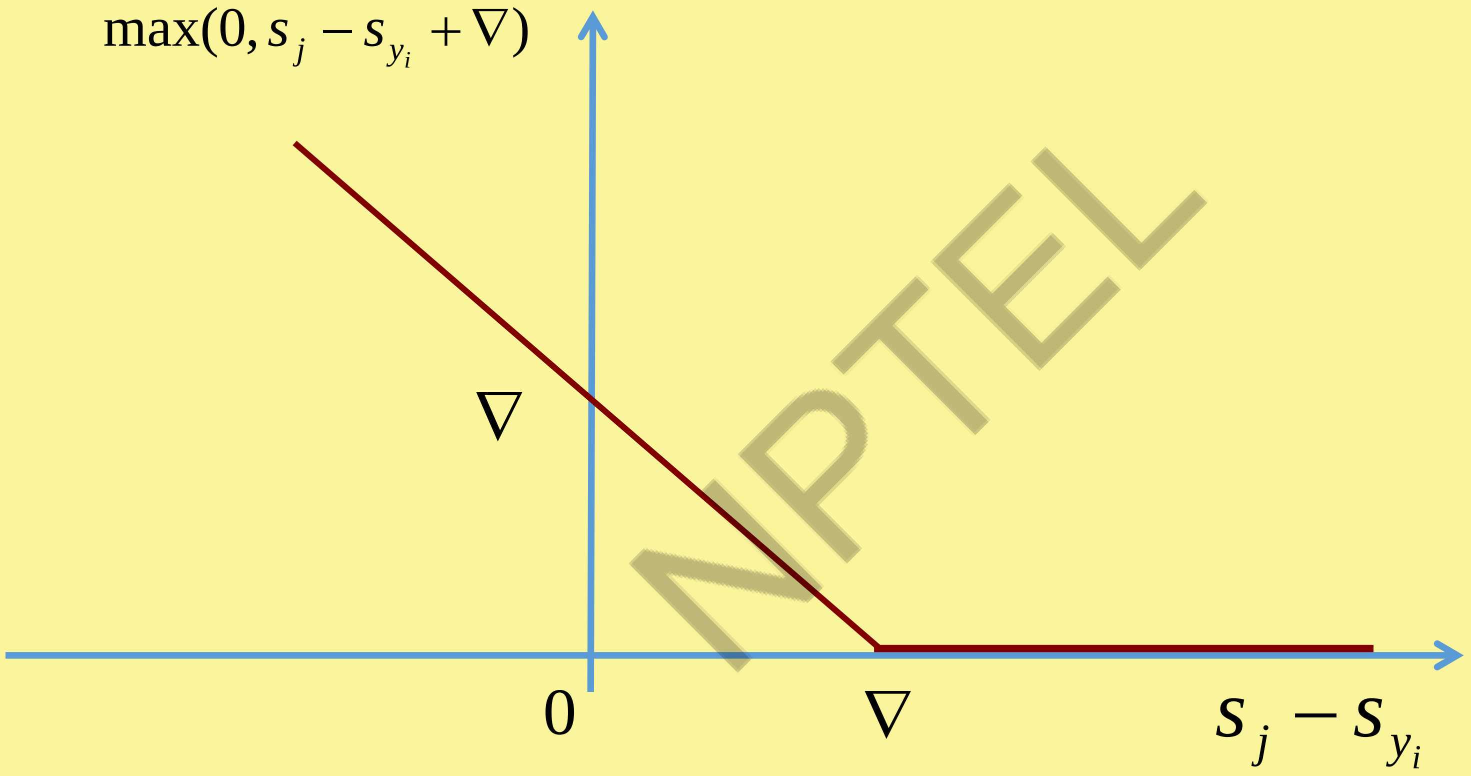
$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \Delta)$$

$$= \max(0, 10 - 30 + 10) + \max(0, -20 - 30 + 10) + \max(0, 25 - 30 + 10)$$

$$= 0 + 0 + 15$$

$$= 15$$

Hinge Loss



Regularization

$$s_j - s_{y_i} = W_j^t X_i - W_{y_i}^t X_i$$

$$\text{Scaling } W \text{ by } \lambda : W \leftarrow \lambda W$$

$$\Downarrow$$

$$s_j - s_{y_i} \leftarrow \lambda\left(s_j - s_{y_i}\right)$$

Regularization

Include a regularization term *R(W)*

$$R(W) = \lambda \sum_k \sum_l W_{kl}^2$$

$$L = \frac{1}{N} \sum_i L_i + \lambda R(W)$$

$$L = \frac{1}{N} \sum_i \sum_{j \neq y_i} [\max(0, f(X_i, W)_j - f(X_i, W)_{y_i} + \nabla) + \lambda \sum_k \sum_l W_{kl}^2$$

Choice of Hyper Parameter

$$L = \frac{1}{N}\sum_i \sum_{j \neq y_i} [\max(0,\ f(X_i, W)_j - f(X_i, W)_{y_i} \ \ + \nabla) + \lambda \sum_k \sum_l W_{kl}^2$$

∇ and λ control the same tradeoff $\quad \Rightarrow \nabla = 1$

$$L = \frac{1}{N}\sum_i \sum_{j \neq y_i} [\max(0,\ f(X_i, W)_j - f(X_i, W)_{y_i} + 1) + \lambda \sum_k \sum_l W_{kl}^2$$

Binary SVM $\Rightarrow L_i = C \max(0, 1 - y_i W^t X_i) + R(W)$

Loss Function Visualization

Visualizing Loss Function

$$\text{Consider 3 Classes} \Rightarrow W = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix}$$

$$3 \;\; 1\text{–dimensional points} \Rightarrow (X_1, 1), (X_2, 2) \text{ and } (X_3, 3)$$

Visualizing Loss Function

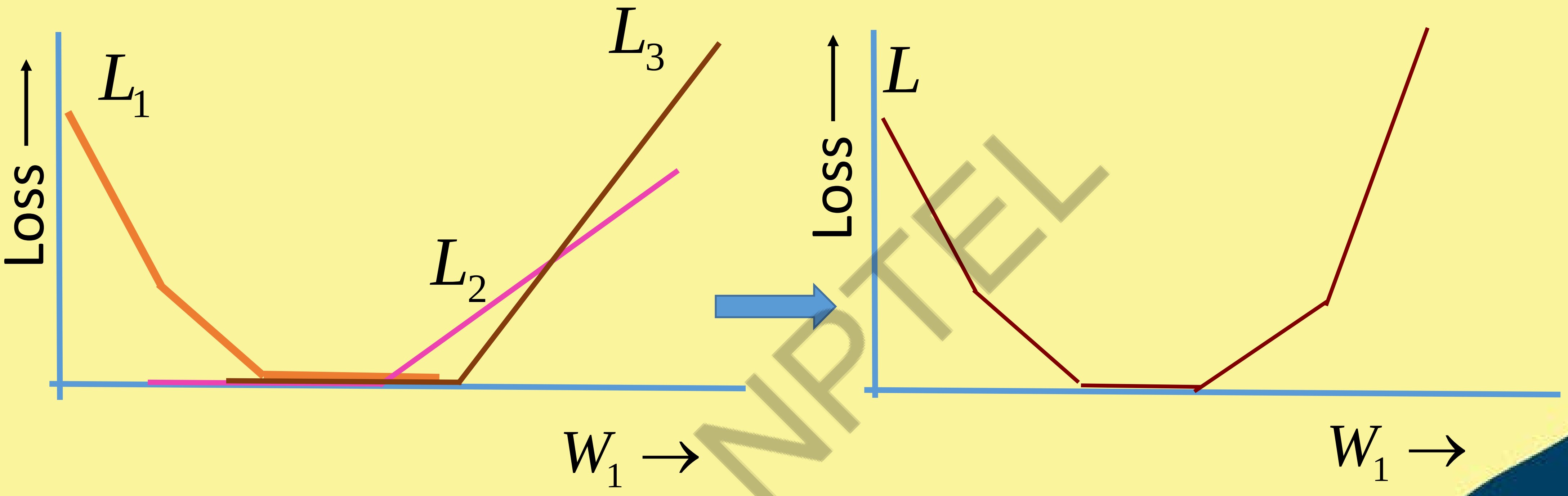
$$L_1 = \max(0, W_2^t X_1 - W_1^t X_1 + 1) + \max(0, W_3^t X_1 - W_1^t X_1 + 1)$$

$$L_2 = \max(0, W_1^t X_2 - W_2^t X_2 + 1) + \max(0, W_3^t X_2 - W_2^t X_2 + 1)$$

$$L_3 = \max(0, W_1^t X_3 - W_3^t X_3 + 1) + \max(0, W_2^t X_3 - W_3^t X_3 + 1)$$

$$L = \frac{1}{3}(L_1 + L_2 + L_3)$$

Visualizing Loss Function



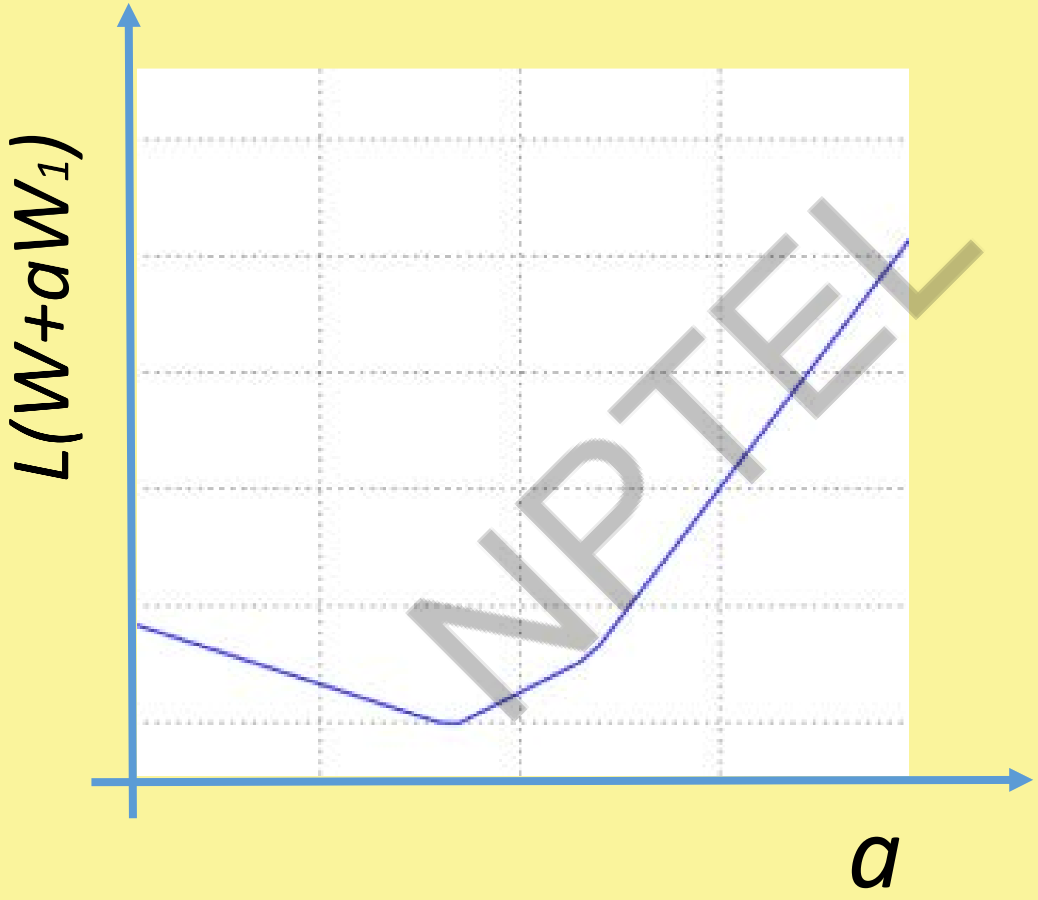
Visualizing Loss Function

- Take a random W (a single point in space)
- Take a random direction W_1
- Record the Loss along W_1

$$\Rightarrow L(W + aW_1)$$

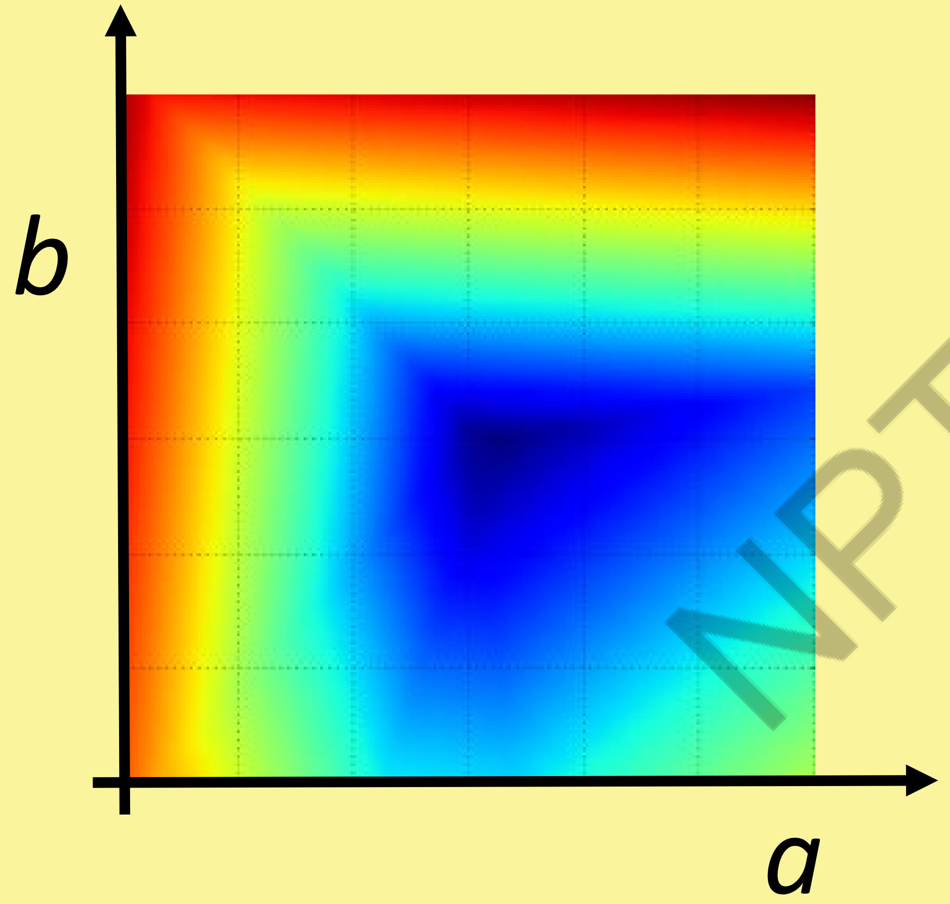
Source - http://cs231n.github.io

Visualizing Loss Function



Source - http://cs231n.github.io

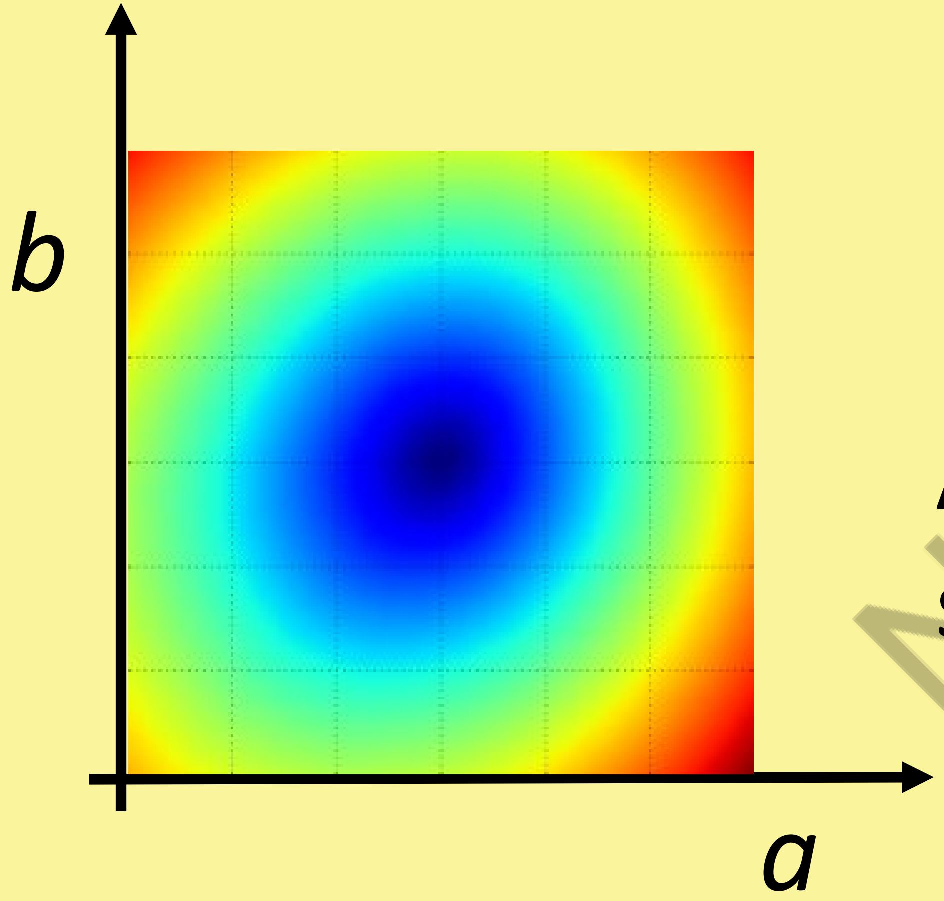
Visualizing Loss Function



$$L(W + aW_1 + bW_2)$$

Source - http://cs231n.github.io

Visualizing Loss Function



$$L(W + aW_1 + bW_2)$$

Averaged over multiple samples

Source - http://cs231n.github.io

Optimizing Loss Function

$$L = \frac{1}{N}\sum_{i}\sum_{j\neq y_i}\left[\max(0, W_j^t X_i - W_{y_i}^t X_i + \nabla\right] + \lambda\sum_{k}\sum_{l} W_{kl}^2$$

$$\nabla_{W_{y_i}} = -\frac{1}{N}\sum_{i}\sum_{j\neq y_i}[X_i \mid (W_j^t X_i - W_{y_i}^t X_i + \nabla > 0)] + \eta W_{y_i}$$

$$\nabla_{W_j} = \frac{1}{N}\sum_{i}\sum_{j\neq y_i}[X_i \mid (W_j^t X_i - W_{y_i}^t X_i + \nabla > 0)] + \xi W_j$$

Optimizing Loss Function

$$\nabla_{W_{y_i}} = -\frac{1}{N}\sum_i \sum_{j \neq y_i} [X_i \mid (W_j^t X_i - W_{y_i}^t X_i + \nabla > 0)] + \eta W_{y_i} \qquad \nabla_{W_j} = \frac{1}{N}\sum_i \sum_{j \neq y_i} [X_i \mid (W_j^t X_i - W_{y_i}^t X_i + \nabla > 0)] + \xi W_j$$

Gradient descent

$$W_{y_i}(k+1) \leftarrow (1-\eta) W_{y_i}(k) + \frac{1}{N}\sum_i \sum_{j \neq y_i} [X_i \mid (W_j^t X_i - W_{y_i}^t X_i + \nabla > 0)]$$

$$W_j(k+1) = (1-\xi) W_j(k) - \frac{1}{N}\sum_i \sum_{j \neq y_i} [X_i \mid (W_j^t X_i - W_{y_i}^t X_i + \nabla > 0)]$$

Source - http://cs231n.github.io

NPTEL ONLINE CERTIFICATION COURSES

Thank you