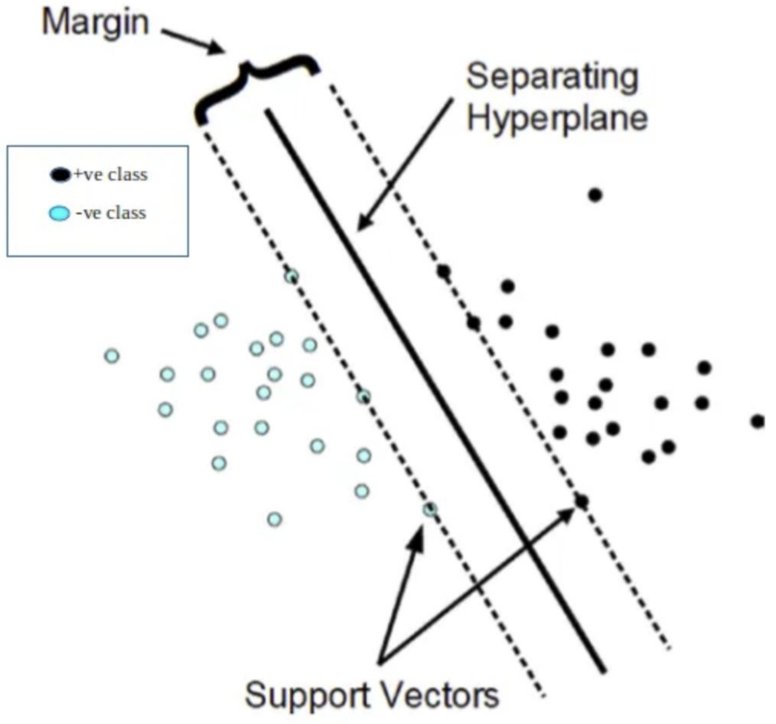
Advanced Machine Learning

Likhit Nayak

SVMs - Maximizing Margin



$$\max_{\mathbf{w},b} \gamma(\mathbf{w}, b)$$

$$\forall i \; y_i(\mathbf{w}^T x_i + b) \geq 0$$

$$\gamma(\mathbf{w}, b) = \min_{\mathbf{x} \in D} \frac{|\mathbf{w}^T\mathbf{x}+b|}{\|\mathbf{w}\|_2}$$

SVMs - Maximizing Margin

The margin is **scale invariant**

$$\gamma(\beta \mathbf{w}, \beta b) = \gamma(\mathbf{w}, b), \forall \beta \neq 0$$

Let’s fix the scale such that:

$$\left|\mathbf{w}^T \mathbf{x}_i + b\right| \quad = 1$$

where $\boldsymbol{x_i}$ is the closest point to a given hyperplane ***(w, b)***

SVMs - Maximizing Margin

$$\min_{\mathbf{w},b} \mathbf{w}^T\mathbf{w}$$

$$\text{s.t.} \quad \forall i \; y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$$

Lagrange Multipliers

Given the optimization problem:

$$\min_{w} f(w)$$

$$h_i(w) = 0, \forall 1 \leq i \leq l$$

The Lagrangian is given by:

$$\mathcal{L}(w, \boldsymbol{\beta}) = f(w) + \sum_{i} \beta_i h_i(w)$$

Lagrange Multipliers

Given the optimization problem:

$$\min_{w} \ f(w)$$

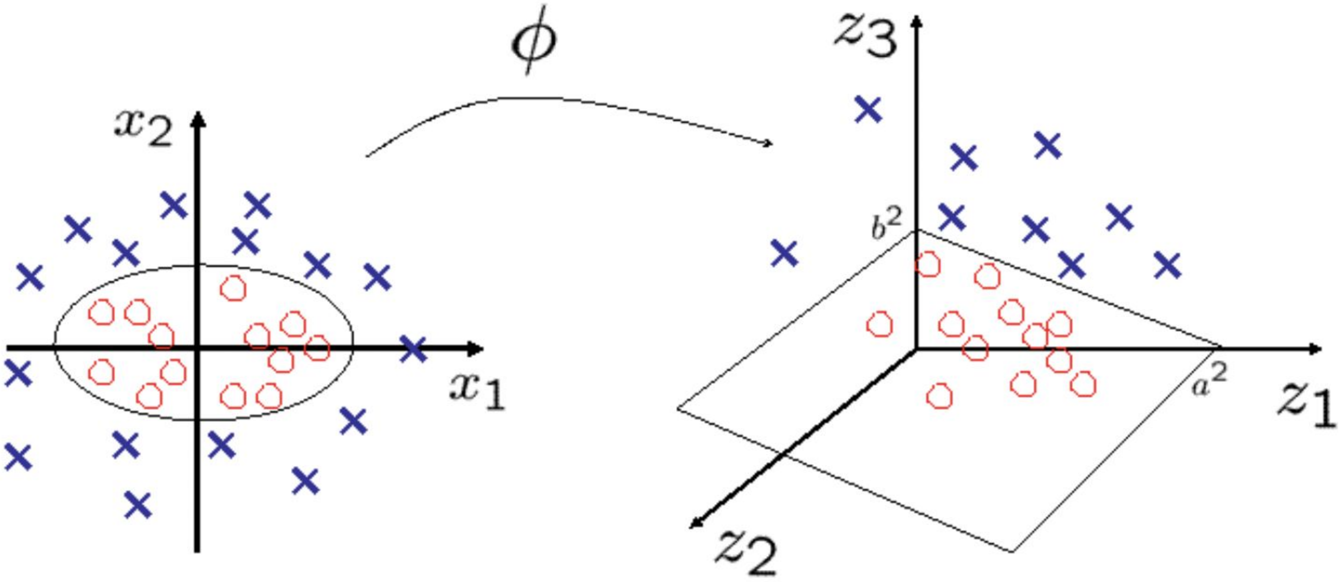
$$g_i(w) \leq 0, \forall 1 \leq i \leq k$$

$$h_j(w) = 0, \forall 1 \leq j \leq l$$

The Lagrangian is given by:

$$\mathcal{L}(w, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \ f(w) + \sum_i \alpha_i g_i(w) + \sum_j \beta_j h_j(w)$$

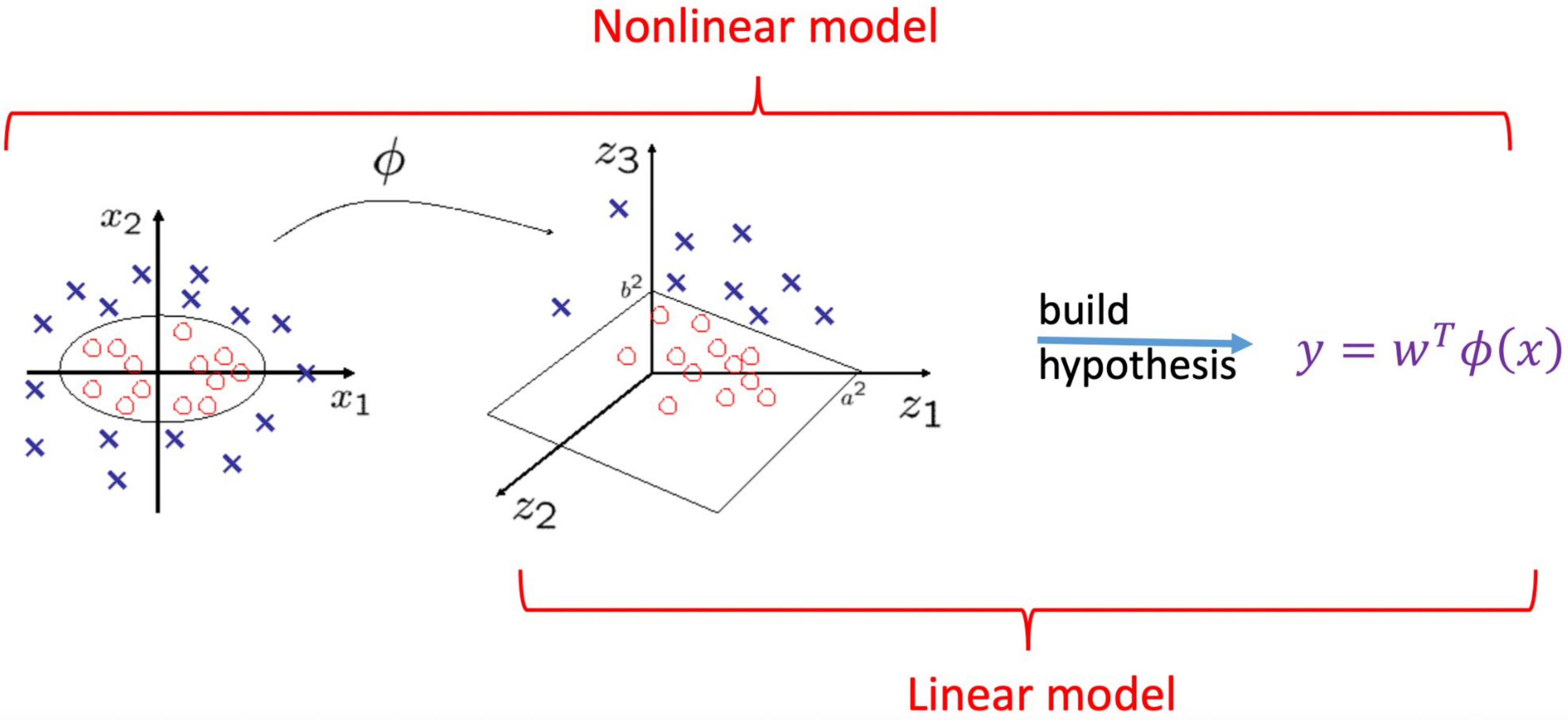
SVMs - Kernel



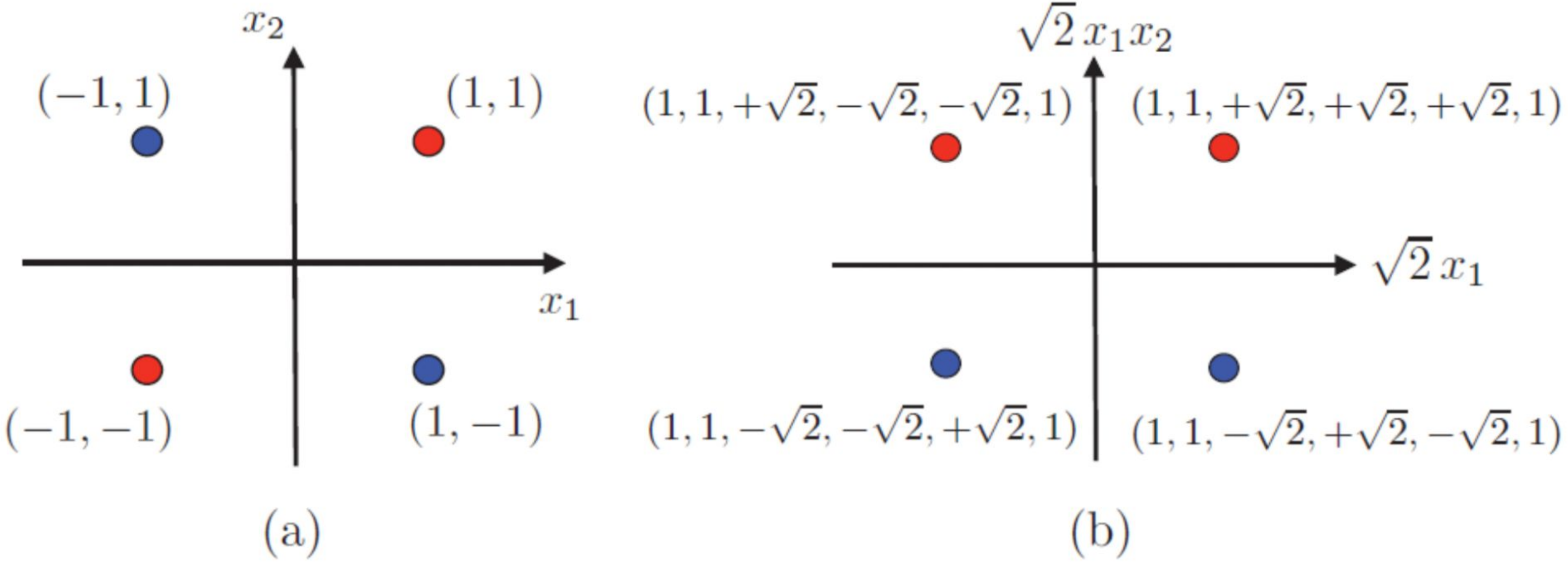
$$\phi : (x_1, x_2) \longrightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

SVMs to Neural Nets



SVMs - Kernel


x_2
$(-1, 1)$
$(1, 1)$
x_1
$(-1, -1)$
$(1, -1)$
(a)
$\sqrt{2}\,x_1x_2$
$(1, 1, +\sqrt{2}, -\sqrt{2}, -\sqrt{2}, 1)$
$(1, 1, +\sqrt{2}, +\sqrt{2}, +\sqrt{2}, 1)$
$\sqrt{2}\,x_1$
$(1, 1, -\sqrt{2}, -\sqrt{2}, +\sqrt{2}, 1)$
$(1, 1, -\sqrt{2}, +\sqrt{2}, -\sqrt{2}, 1)$
(b)