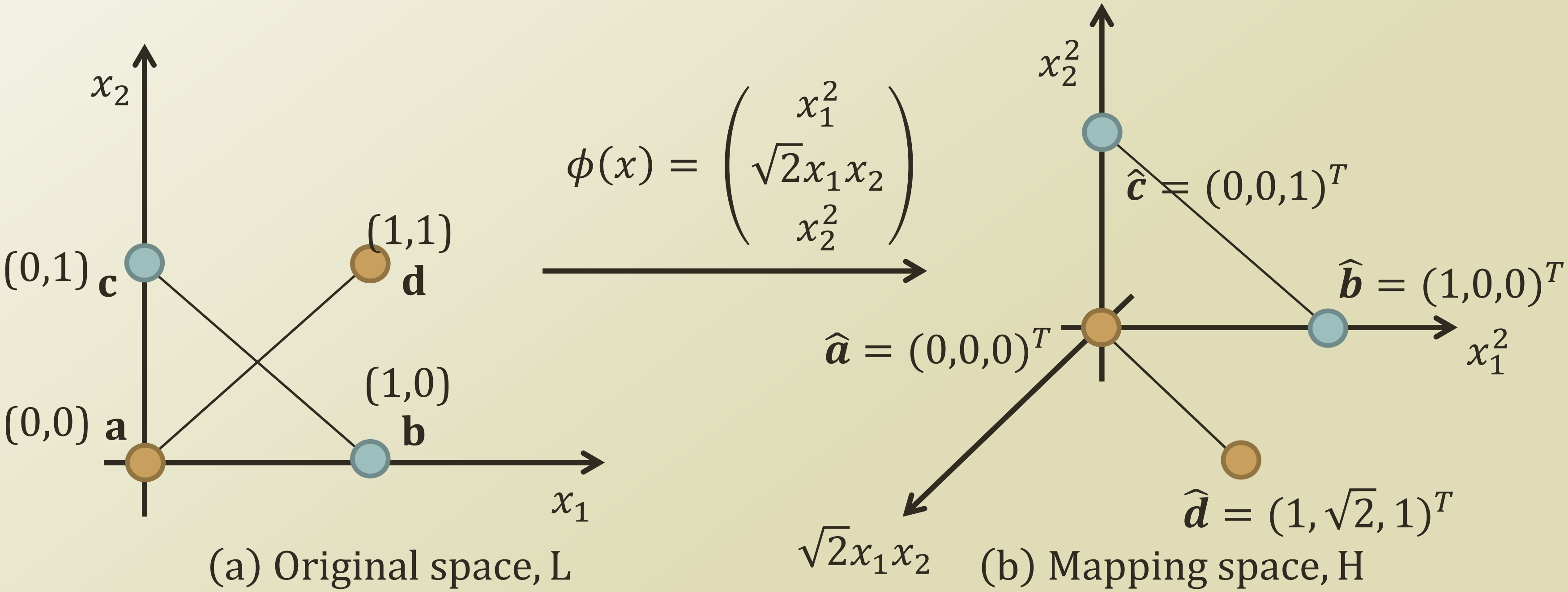
Support Vector Machine

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Mapping Functions

- Suppose that there are non-linearly separable data sets…
- The non-linear separable case can be linearly separable when we increase the basis space
  - Standard basis: $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \ldots, \boldsymbol{e_n}$ → Linearly independent and generate $\boldsymbol{R^n}$
- Expanding the Basis through Space mapping function $\phi : L \rightarrow H$
  - Or, transformation function, etc…
- Any problem????
  - Feature space becomes bigger and bigger….



(a) Original space, L (b) Mapping space, H

Kernel Function

- The kernel calculates the inner product of two vectors in a different space (preferably without explicitly representing the two vectors in the different space)
  - $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \varphi(\boldsymbol{x_i}) \cdot \varphi(\boldsymbol{x_j})$

- Some common kernels are following :
  - Polynomial(homogeneous)
    - $k(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} \cdot \boldsymbol{x_j})^d$
  - Polynomial(inhomogeneous)
    - $k(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} \cdot \boldsymbol{x_j} + 1)^d$
  - Gaussian kernel function, a.k.a. Radial Basis Function
    - $k(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-\gamma \|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2\right)$
      - For $\gamma > 0$. Sometimes parameterized using $\gamma = \frac{1}{2\sigma^2}$
  - Hyperbolic tangent, a.k.a. Sigmoid Function
    - $k(\boldsymbol{x_i}, \boldsymbol{x_j}) = \tanh(\kappa \boldsymbol{x_i} \cdot \boldsymbol{x_j} + c)$
    - For some(not every) $\kappa > 0$ and $c < 0$

Polynomial Kernel Function

- Imagine we have
  - $\mathbf{x}=<x_1, x_2>$ and $\mathbf{z}=<z_1, z_2>$
  - Polynomial Kernel Function of degree 1
    - $K(<x_1, x_2>, <z_1, z_2>) = <x_1, x_2> \cdot <z_1, z_2> = x_1 z_1 + x_2 z_2 = \mathbf{x} \cdot \mathbf{z}$
  - Polynomial Kernel Function of degree 2
    - $K(<x_1, x_2>, <z_1, z_2>) = <x_1^2, \sqrt{2} x_1 x_2, x_2^2> \cdot <z_1^2, \sqrt{2} z_1 z_2, z_2^2>$
    - $= x_1^2 z_1^2 + 2 x_1 x_2 z_1 z_2 + x_2^2 z_2^2 = (x_1 z_1 + x_2 z_2)^2 = (\mathbf{x} \cdot \mathbf{z})^2$
  - Polynomial Kernel Function of degree 3
    - $K(<x_1, x_2>, <z_1, z_2>) = (\mathbf{x} \cdot \mathbf{z})^3$
  - Polynomial Kernel Function of degree n
    - $K(<x_1, x_2>, <z_1, z_2>) = (\mathbf{x} \cdot \mathbf{z})^n$
- Do we need to express and calculate the transformed coordinate values for ***x*** and ***z*** to know the polynomial kernel of ***K***?
  - Do we need to convert the feature spaces to exploit the linear separation in the high order?
  - **Condition: only the inner product is computable with this trick**