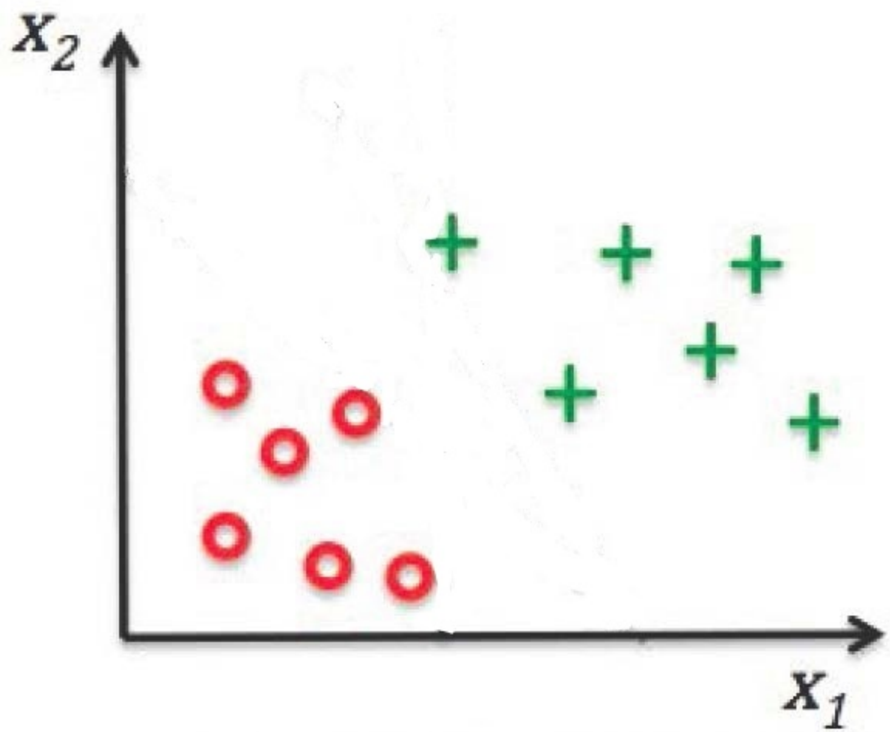
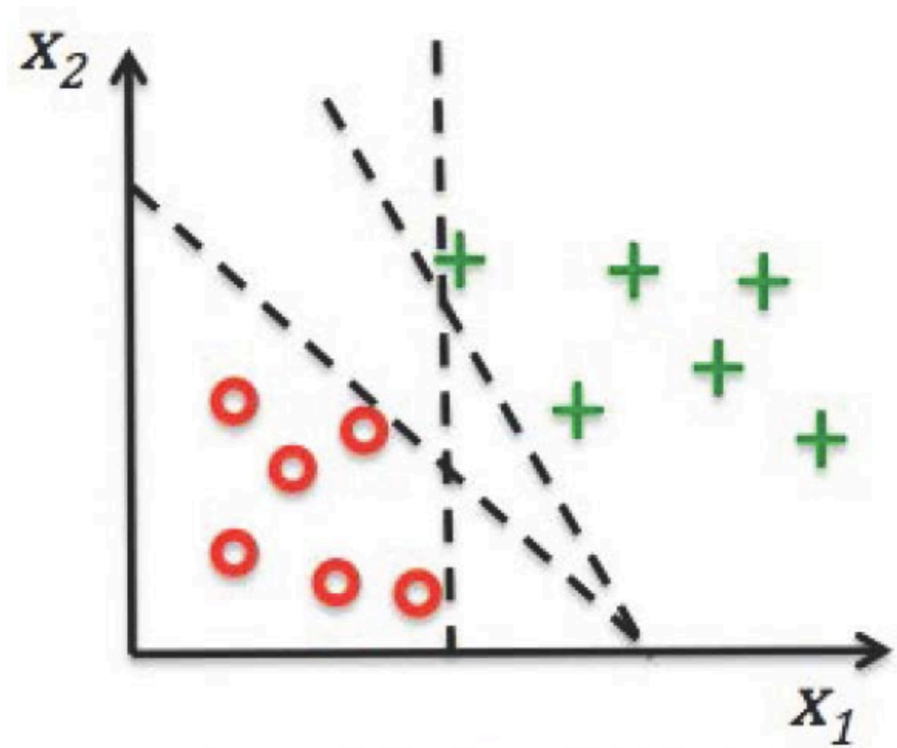
Support Vector Machines

Linearly separable case

START
01



Hard-margin classifications



Which hyperplane to use?

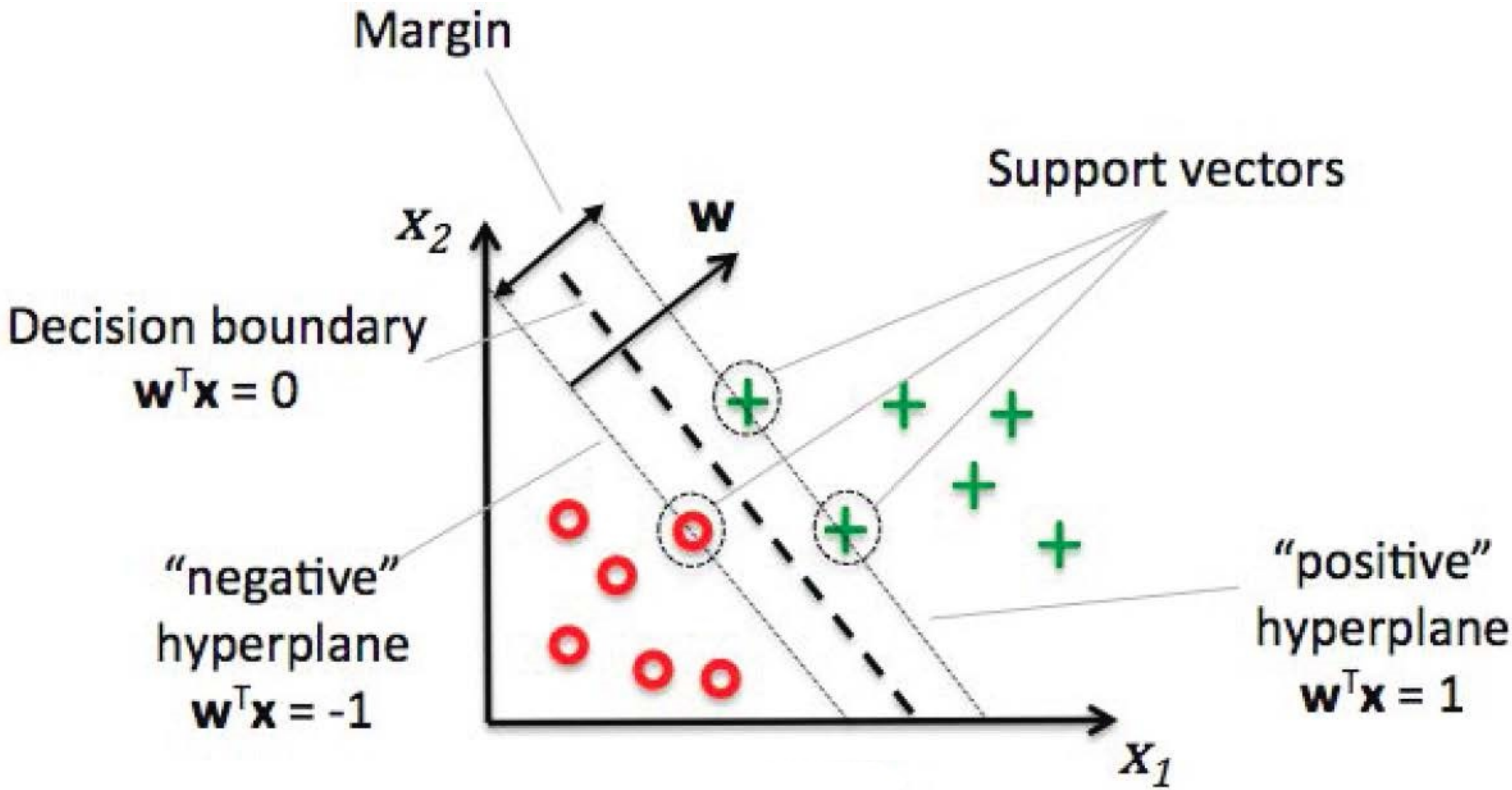
Perceptron	Separate the classes
Adaline	Minimize the cost function
Logistic Regression	Maximize the log-likelihood function
Source Vector Machine	Maximize the **Margin**

STOP

Hard-margin classification

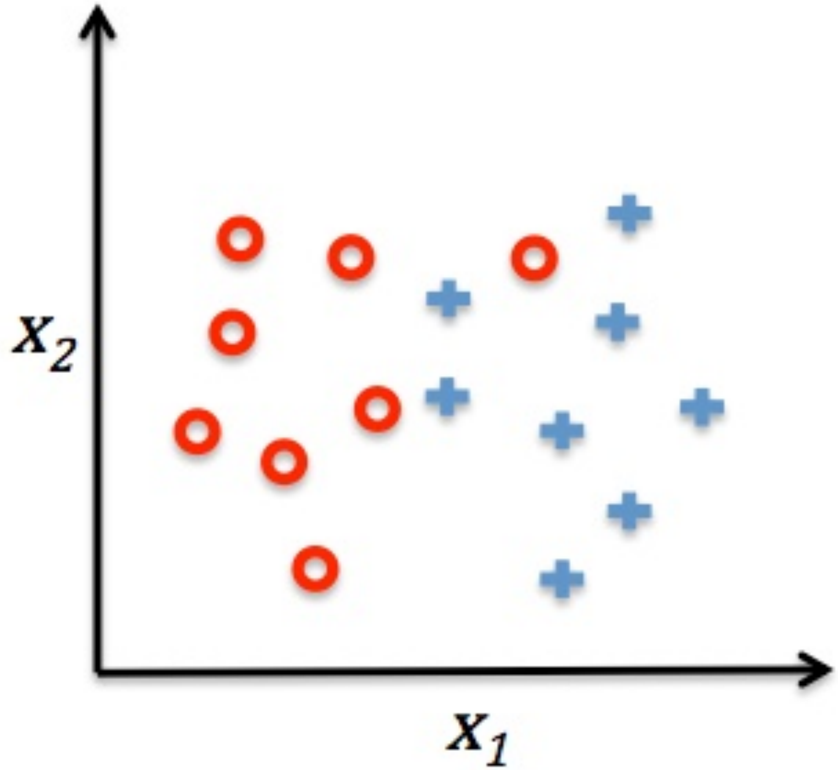
- Choose a pair of parallel hyperplanes such that the hyperplanes and the space between them separate the classes.
- Label one hyperplane as the **Positive** hyperplane and the other as the **Negative** hyperplane (in a OvA or OvA case, the Positive hyerplane includes at least one sample from the One class.
- Define the **Margin** to be the distance between the two hyperplanes
- Choose the two hyperplanes that maximize the **Margin.**
- Designate the hyperplane equidistant between the Positive and Negative hyperplanes as the **Decision Boundary.**
- Label the samples on the Positive and Negative hyperplanes as **Support Vectors.**

Maximize the margin



START
03

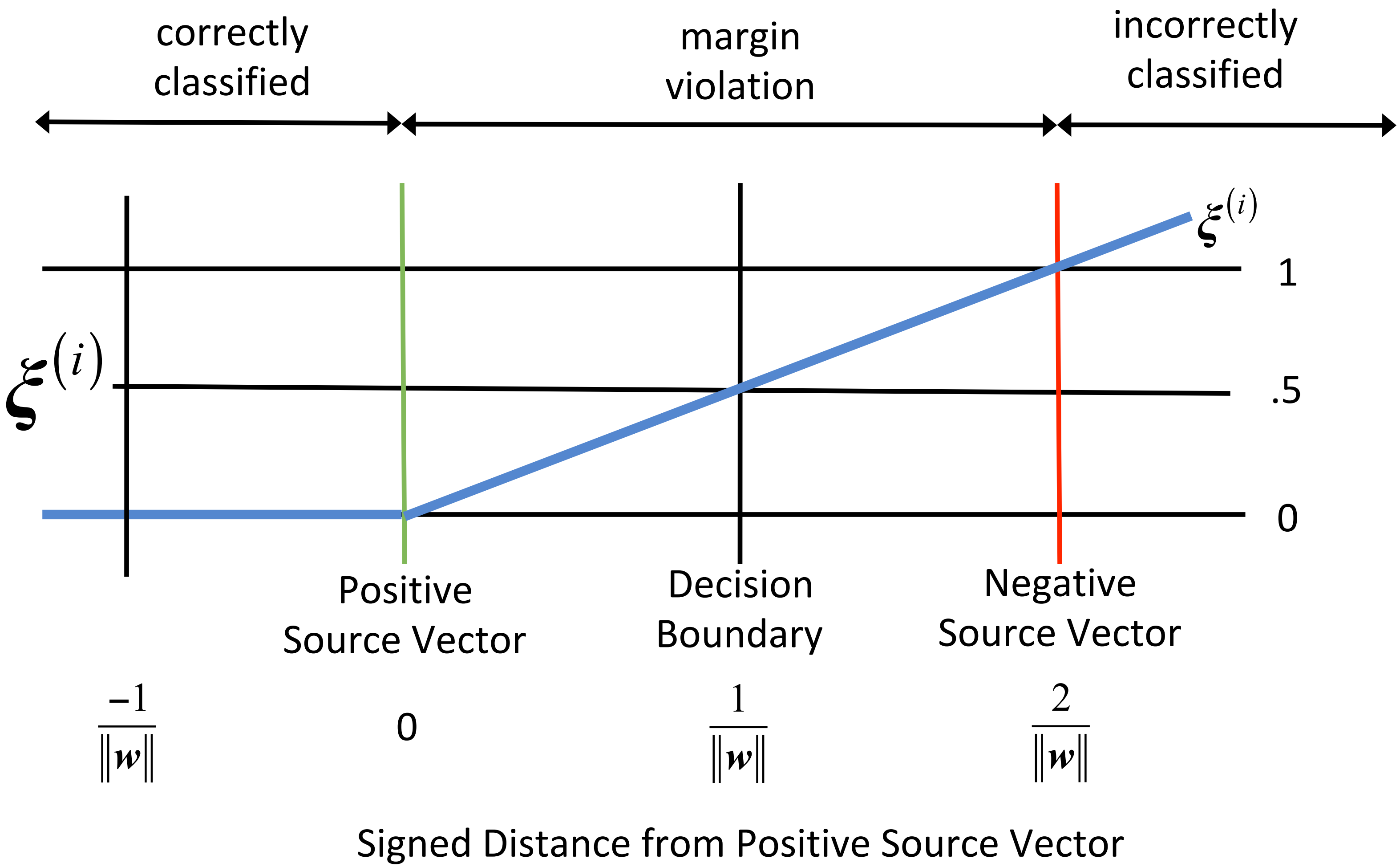
Linearly inseparable case



Soft-margin classifications

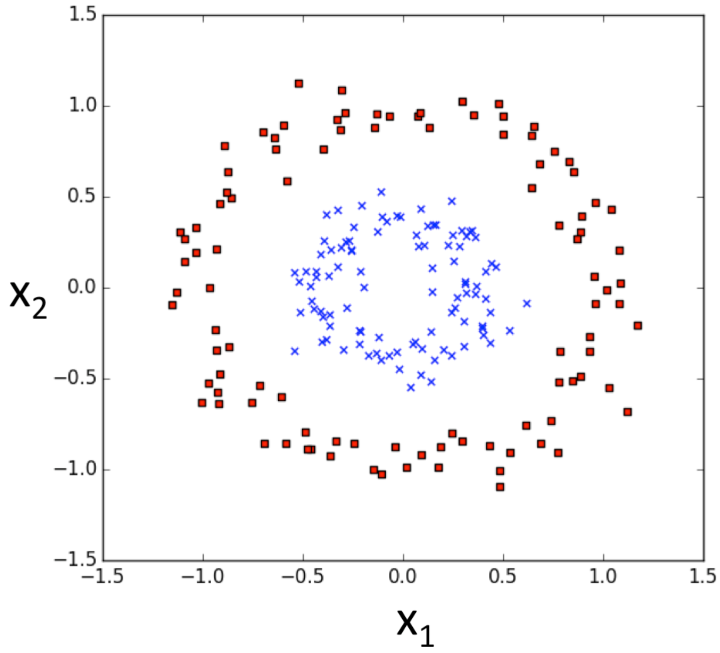
- Slack Variable using a Hinge Loss Function

Slack-variable hinge-loss function



Solving non-liner problems with a SVM

- So far, we've used hyperplanes to divide the regions corresponding to one class from those corresponding to others.
- But for some problems, surfaces other than hyperplanes are more appropriate.



Solving non-liner problems with a SVM

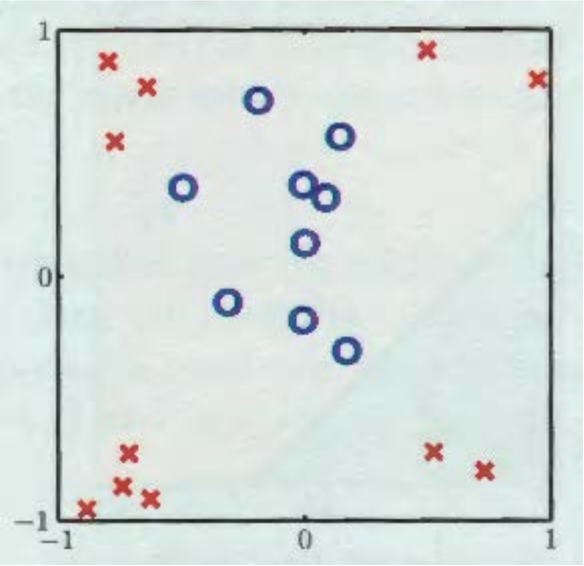
START
05

1. Transform the training data onto a higher dimensional feature space using a mapping function, $\phi(.)$.
2. Train a linear SVM model to classify the transformed data in this new feature space.
3. Transform new, previously unseen, data using the same mapping function, $\phi(.)$ as was used in Step 1 and classify the new, transformed data using the SVM trained in Step 2.

OR

3. Transform the decision boundary determined in Step 2 using $\phi^{-1}(.)$, the inverse of the mapping function used in Step 1, and classify new, previously unseen data using transformed decision boundary.

Non-linear example

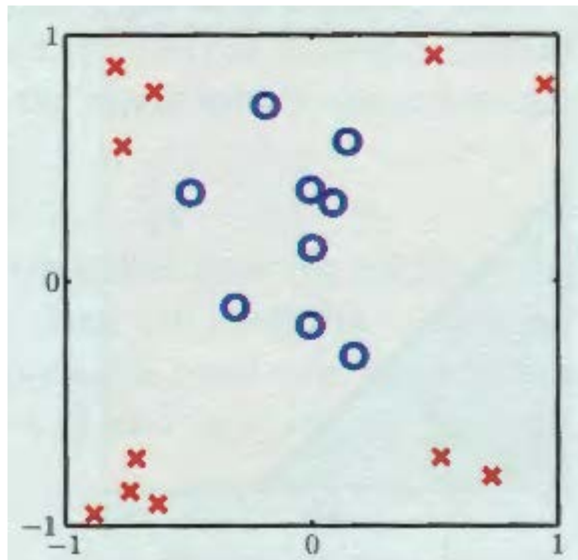


Original Data

$\boldsymbol{x}^{(i)} \in X$

Example is from **Learning From Data, A Short Course**, Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Liu, AMLBook.com, 2012; I highly recommend the book)

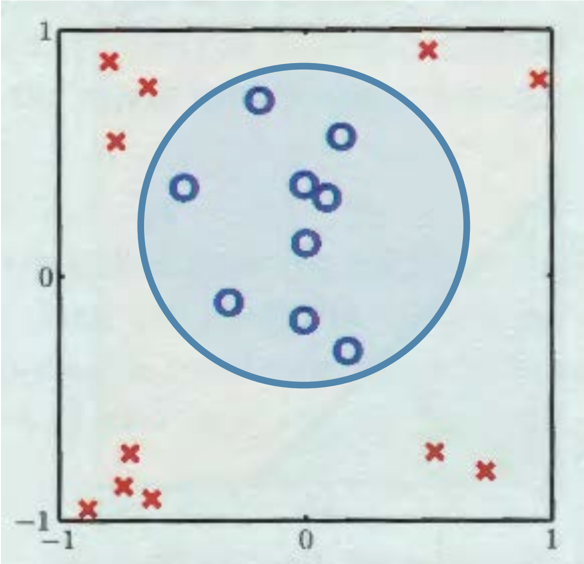
Data is not linearly separable; nor is it close to being so



Original Data

$\boldsymbol{x}^{(i)} \in X$

Possible non-linear decision boundary in *X*-space



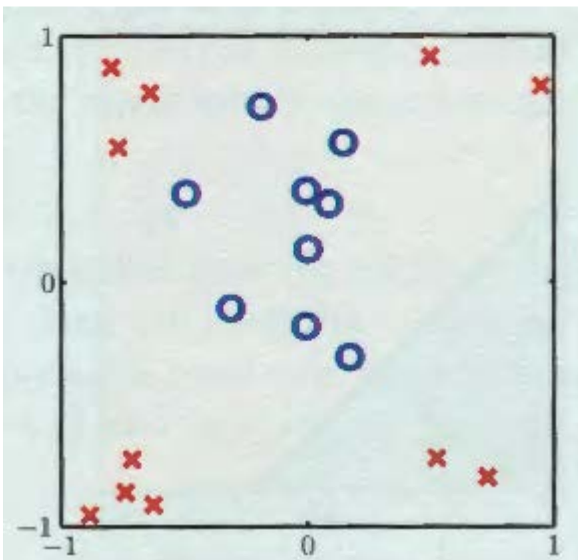
Original Data

$$\boldsymbol{x}^{(i)} \in X$$

with a possible non-linear decision boundary

Transform the data from *X*-space to *Z*-space

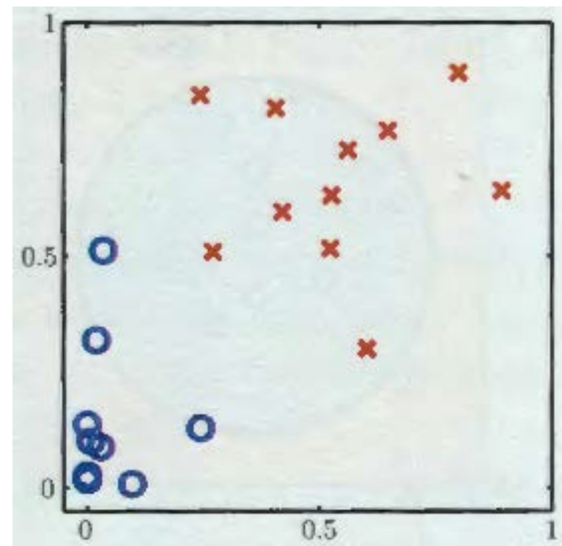
$$z^{(i)} = \Phi\left(1, x_1^{(i)}, x_2^{(i)}\right) = \left(1, \left(x_1^{(i)}\right)^2, \left(x_2^{(i)}\right)^2\right)$$



Original Data

$\boldsymbol{x}^{(i)} \in X$

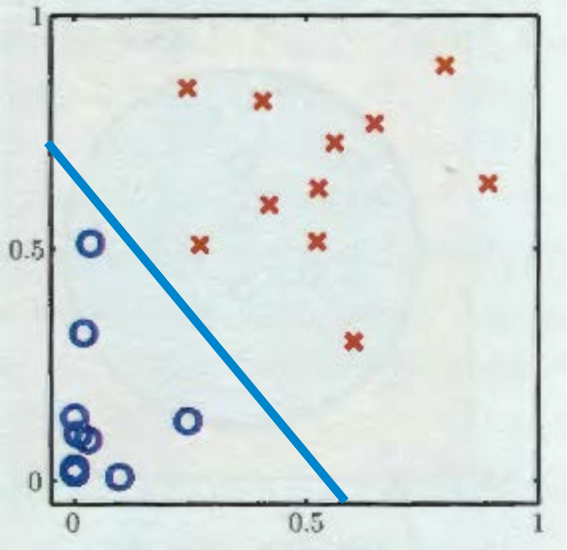
$\xrightarrow{\Phi}$



Transformed Data

$\boldsymbol{z}^{(i)} = \Phi\left(\boldsymbol{x}^{(i)}\right) \in Z$

Possible linear decision boundary in *Z*-space

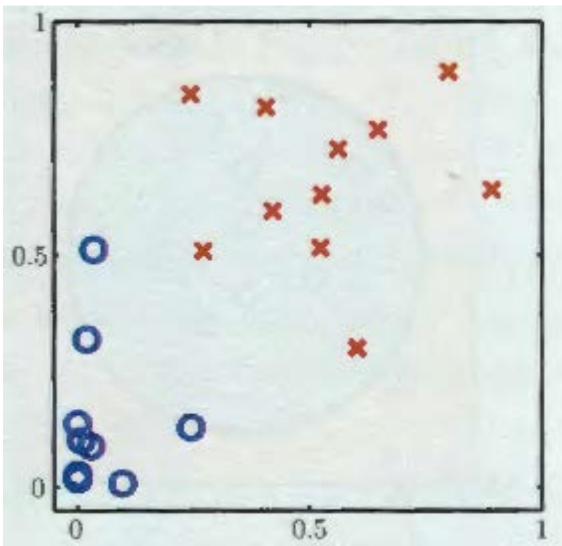


Transformed Data

$$z^{(i)} \in Z$$

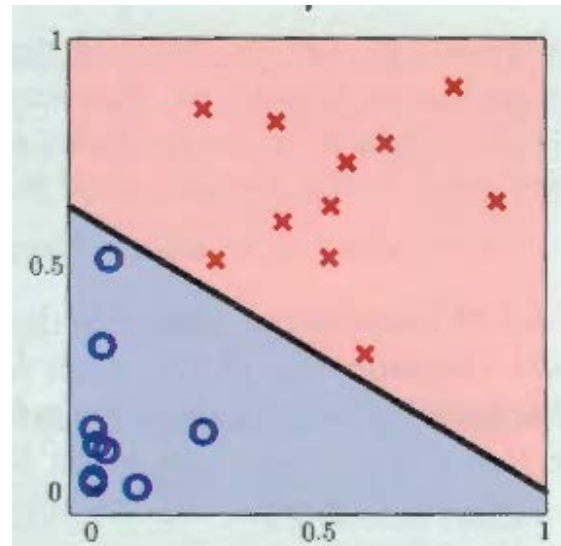
with Potential Decision Boundary

Find the decision boundary



Transformed Data

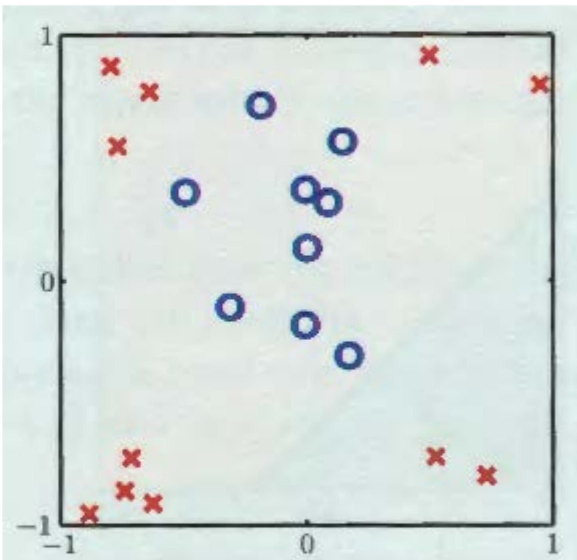
$z^{(i)} = \Phi\left(x^{(i)}\right) \in Z$



Transformed Data

$z^{(i)} = \Phi\left(x^{(i)}\right) \in Z$

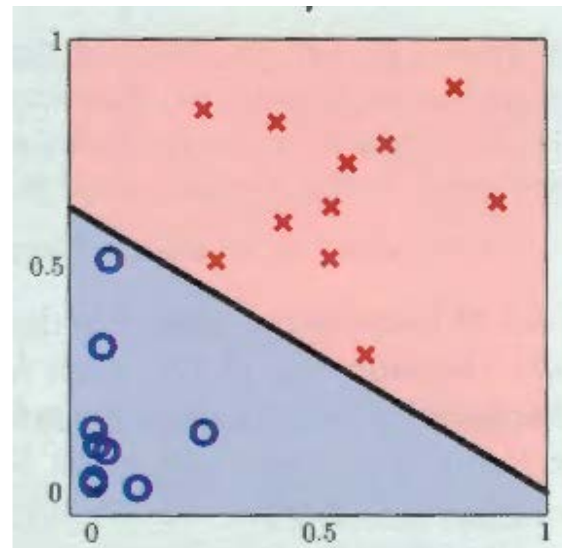
Option 1: Classify New Data in Z-Space



New Data

$\boldsymbol{x}^{(i)} \in X$

$\xrightarrow{\Phi}$

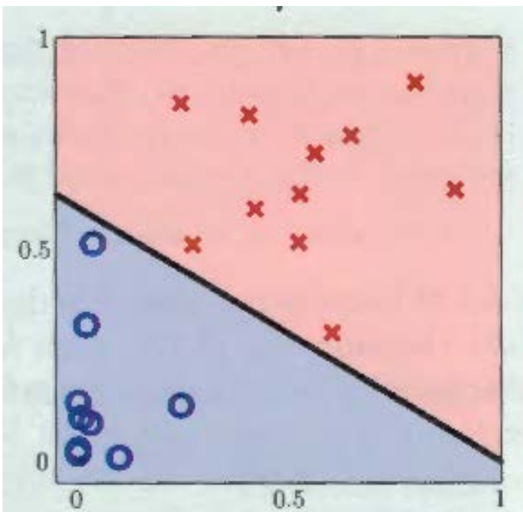


Transformed New Data

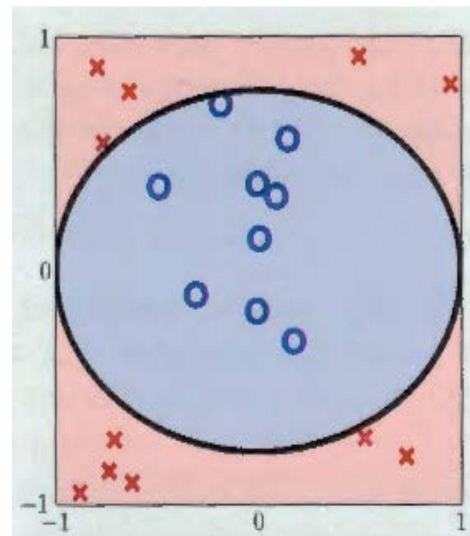
$\boldsymbol{z}^{(i)} = \Phi\left(\boldsymbol{x}^{(i)}\right) \in Z$

Option 2: Transform the decision boundary from *Z*-space to *X*-space

$$x^{DB} = \Phi^{-1}\left(1, z_1^{DB}, z_2^{DB}\right) = \left(1, \pm\sqrt{\left(z_1^{DB}\right)}, \pm\sqrt{\left(z_1^{DB}\right)}\right)$$



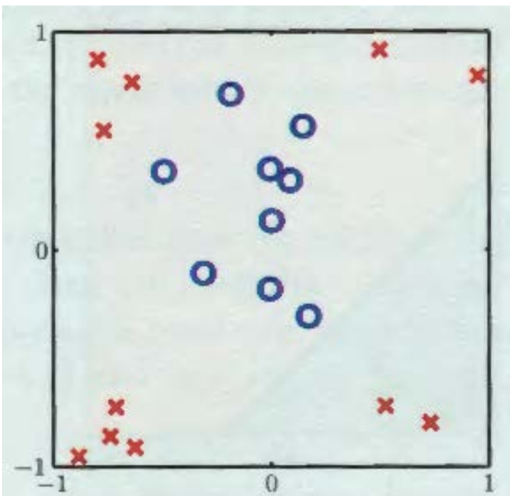
$\xrightarrow{\Phi^{-1}}$



Non-linear decision boundary in **Z**-Space

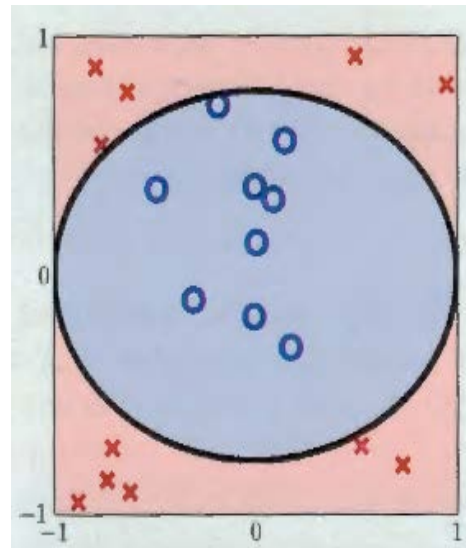
Non-linear decision boundary in **X**-Space

Option 2: Classify new data in *X*-space using the transformed decision boundary



New Data

$\boldsymbol{x}^{(i)} \in X$



New Data

$\boldsymbol{x}^{(i)} \in X$

STOP