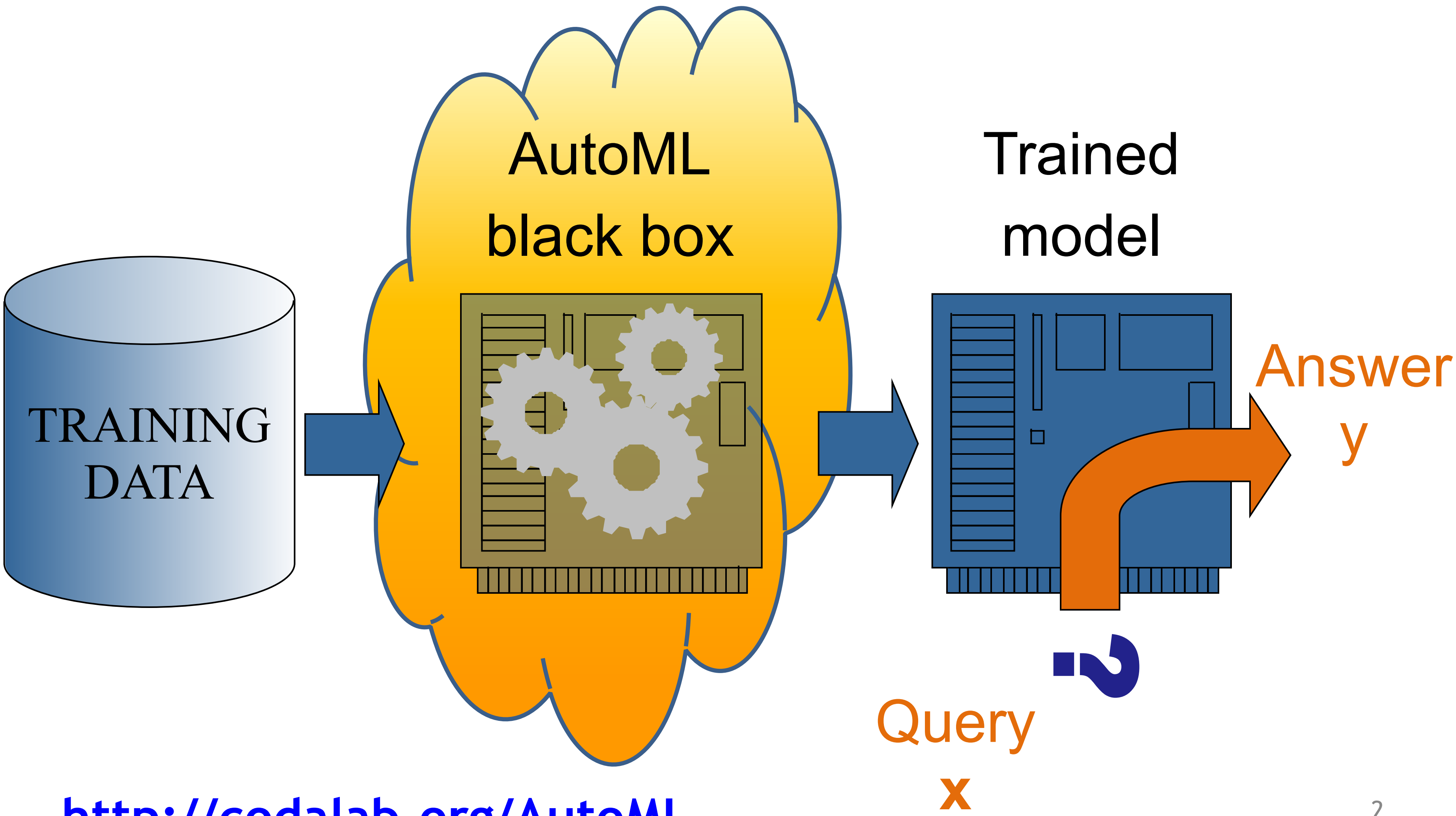
Introduction to Machine Learning

Adapted from Isabelle Guyon

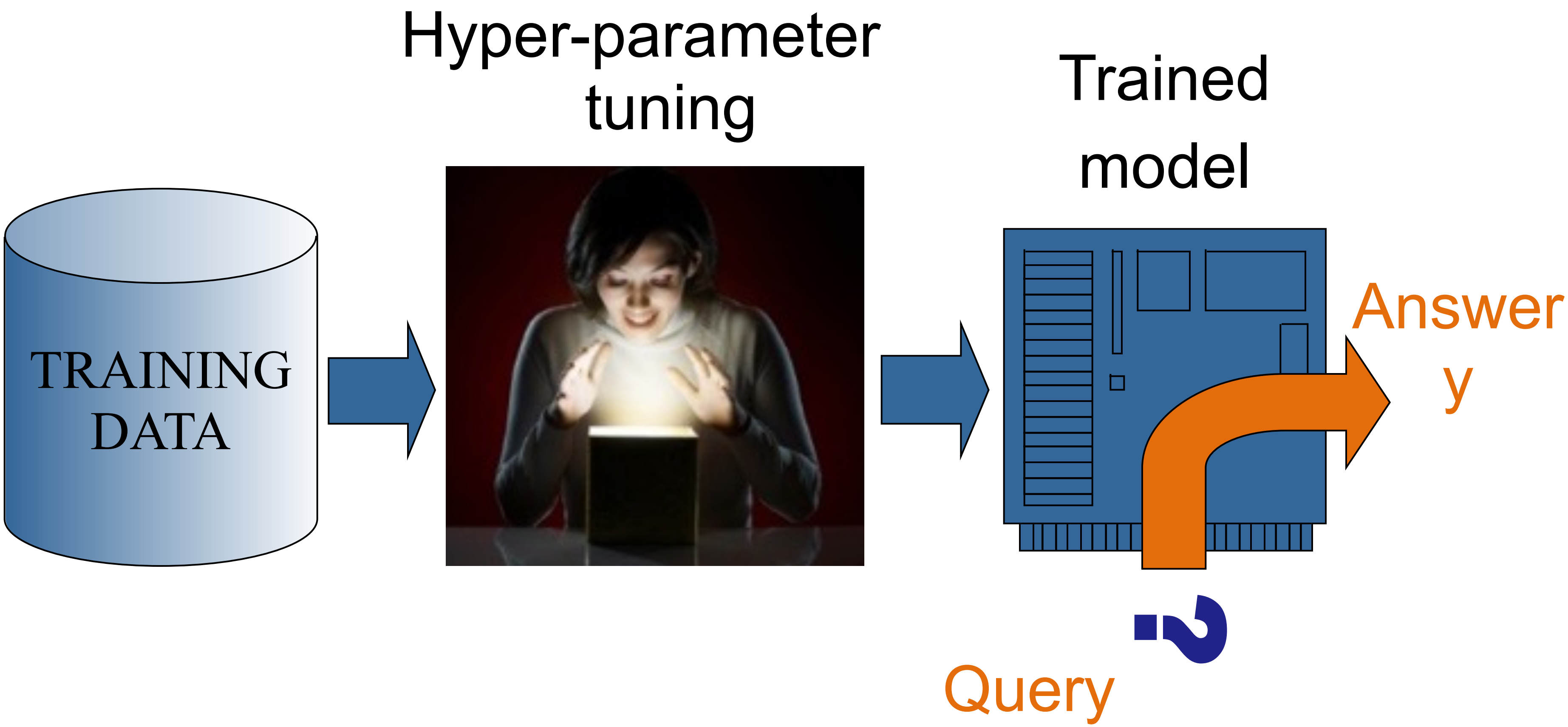
UCB - CS189

The DREAM



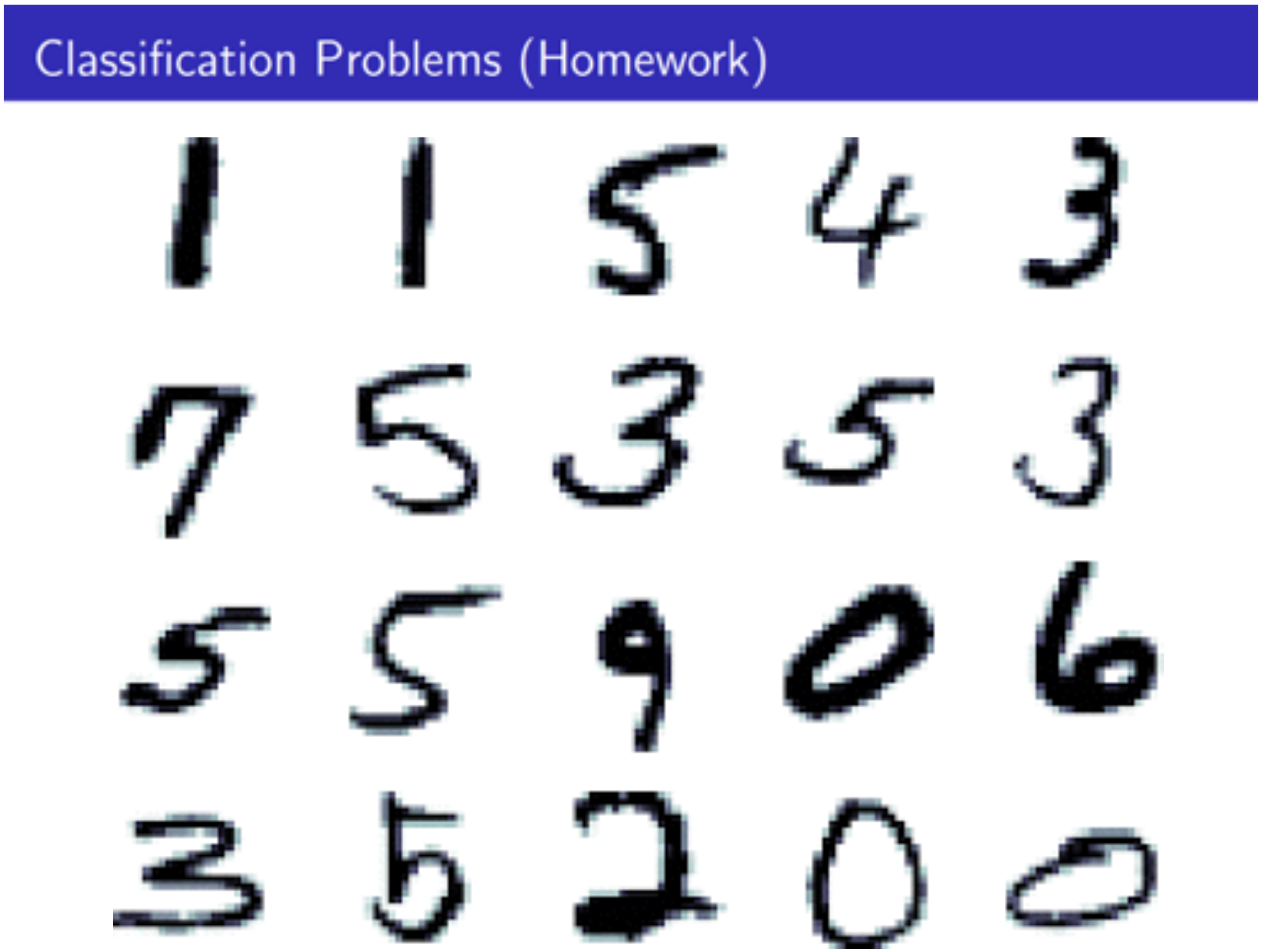
http://codalab.org/AutoML

The REALITY

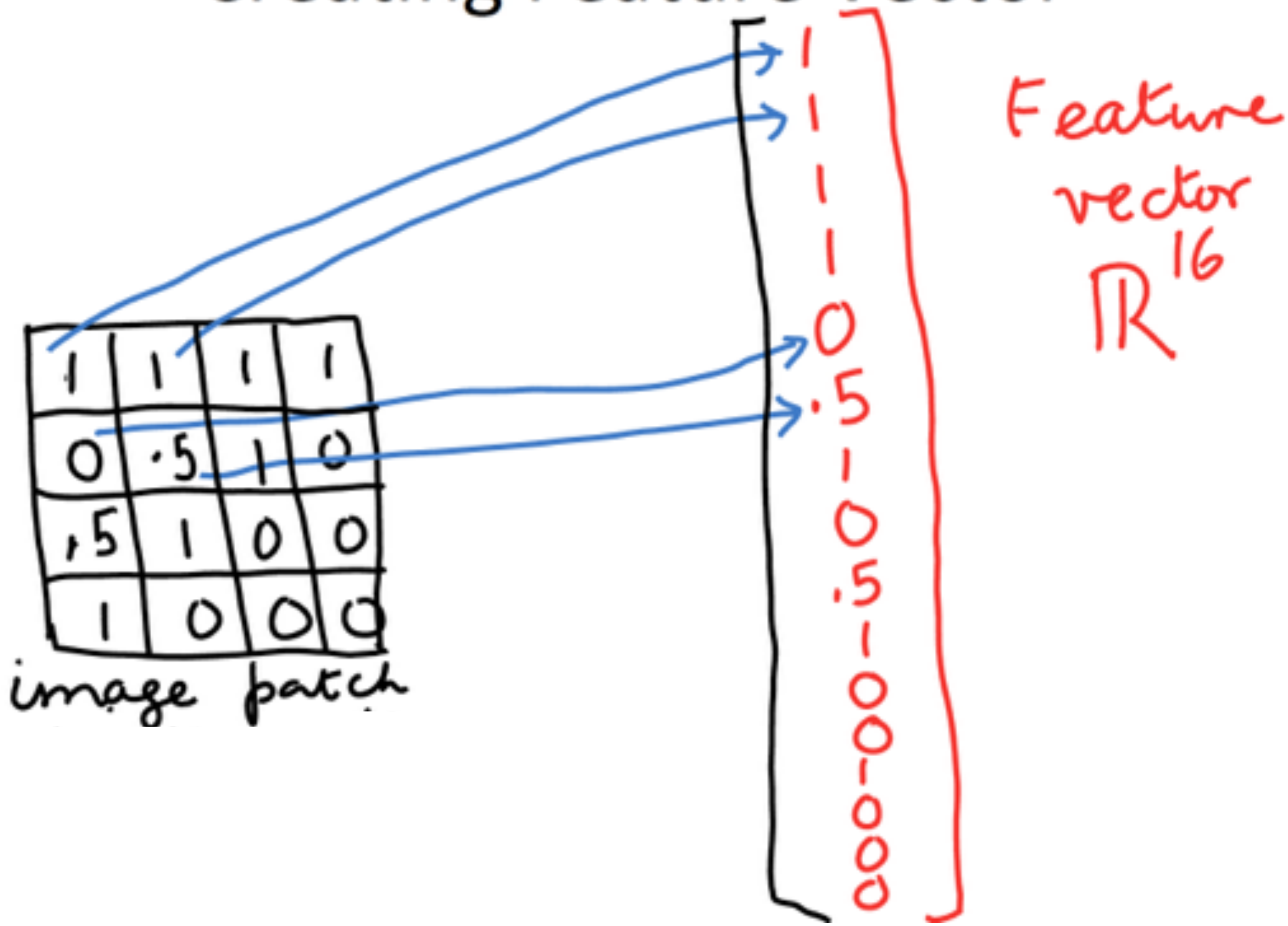


http://codalab.org/AutoML

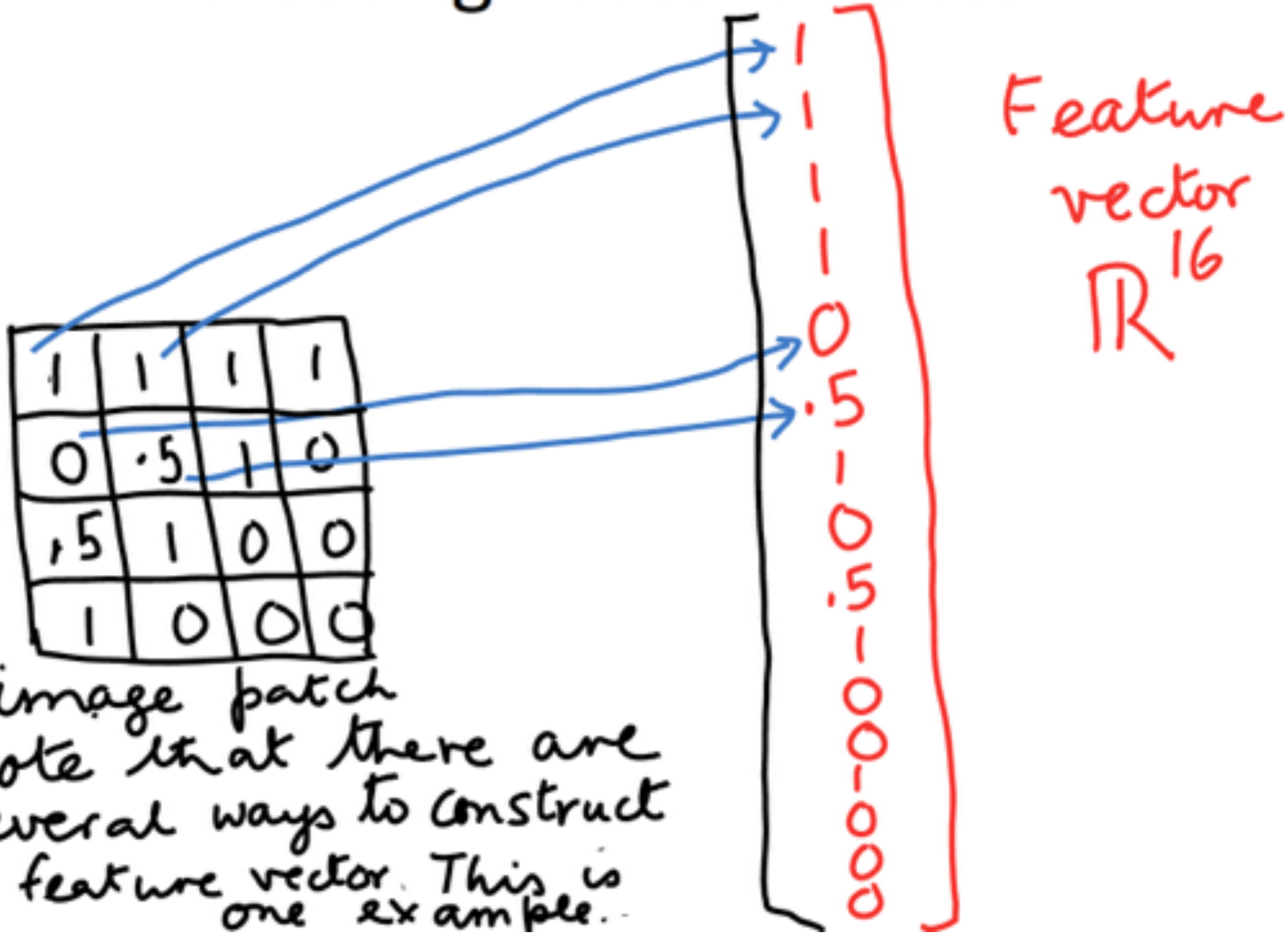
MNIST data



Creating Feature Vector

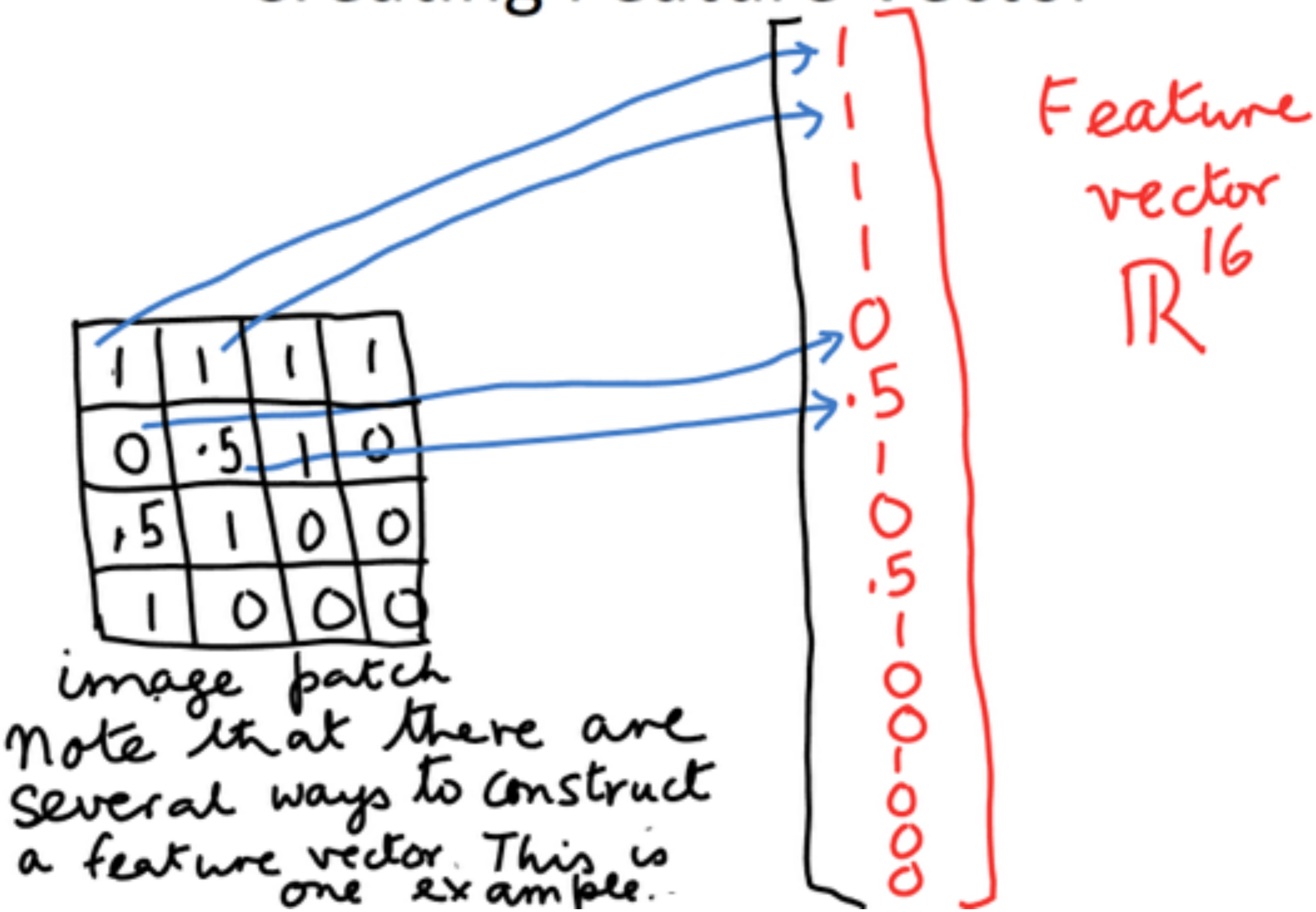


Creating Feature Vector



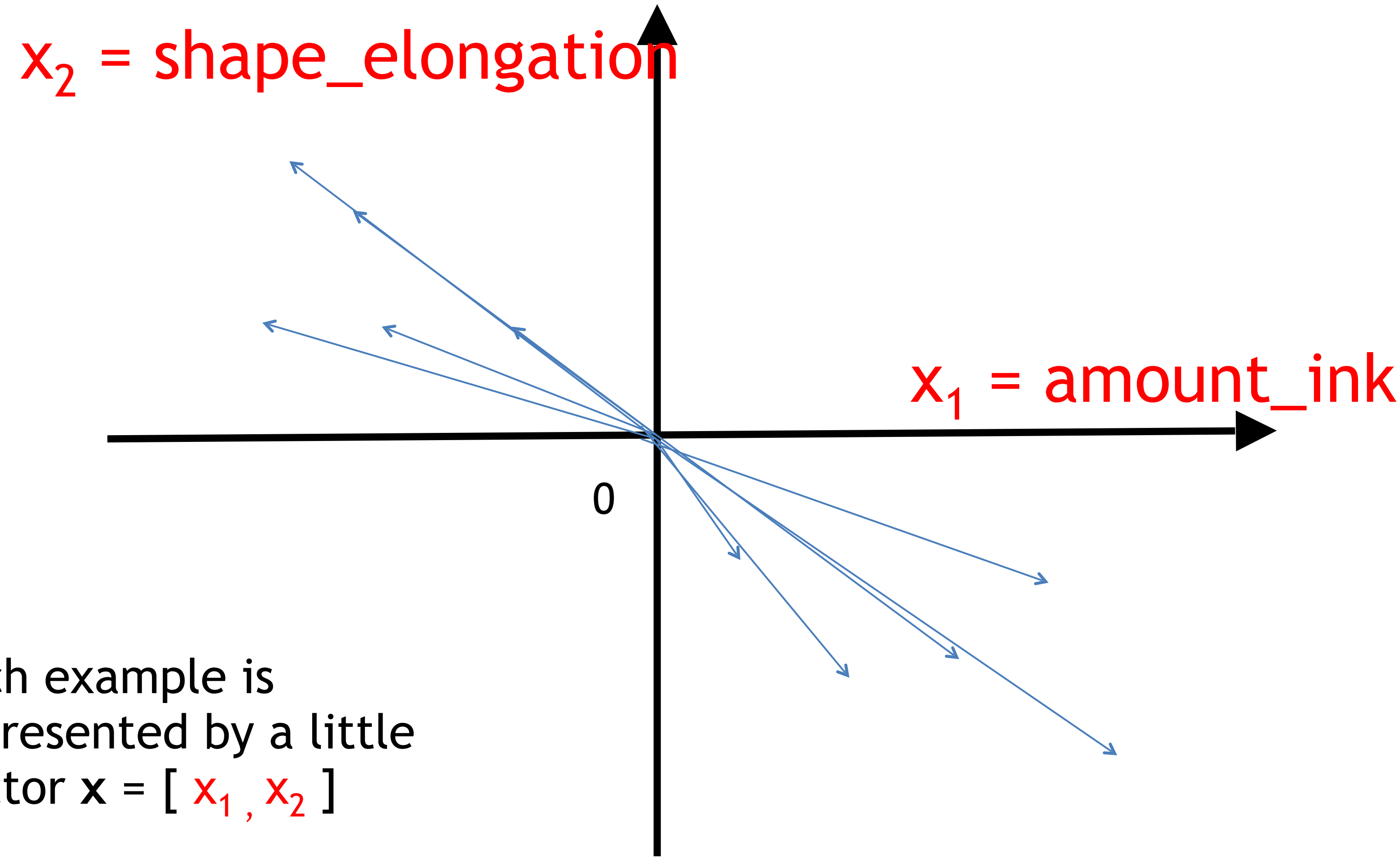
Note that there are several ways to construct a feature vector. This is one example.

Creating Feature Vector



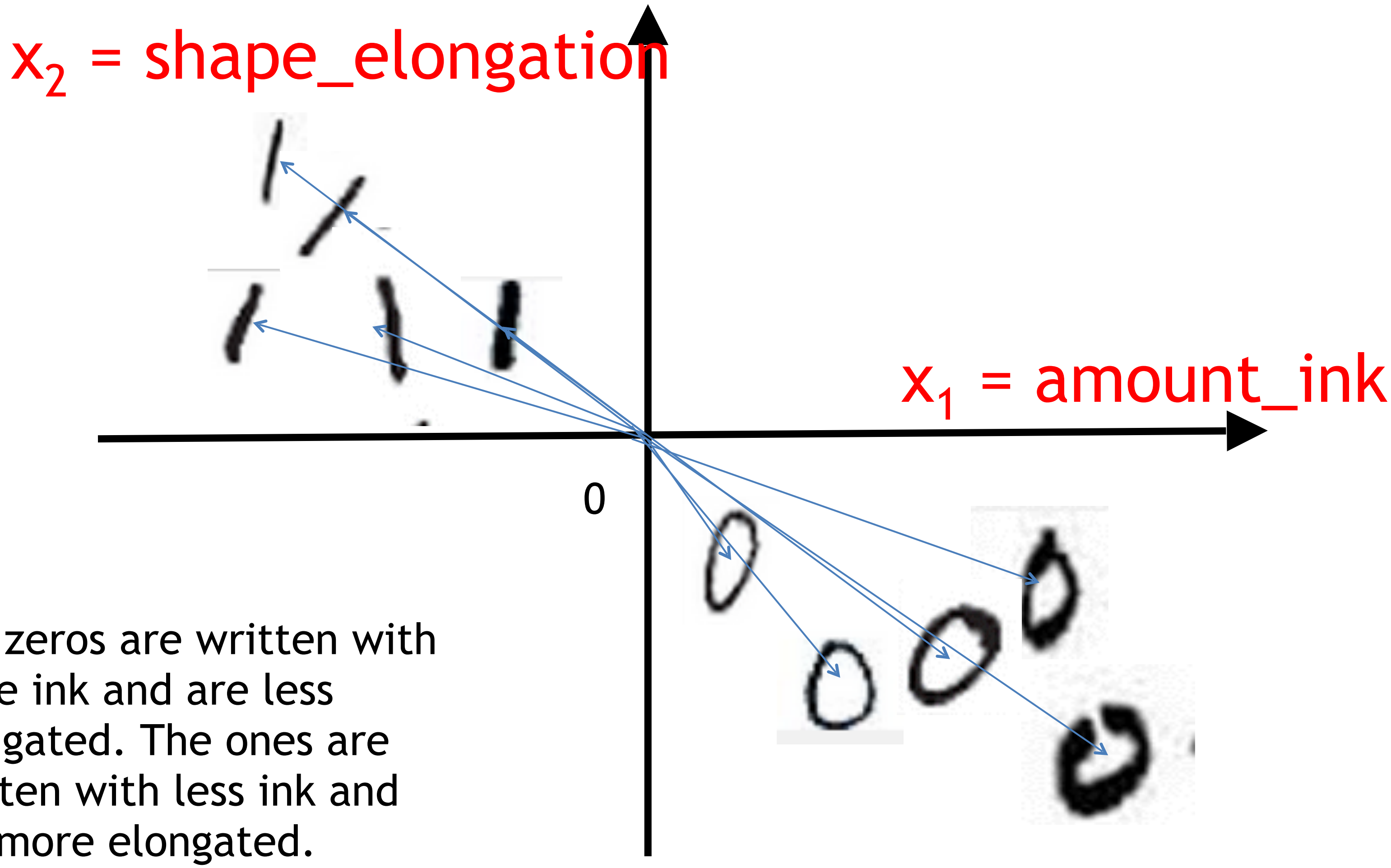
Another example, extract features: **x** = [amount_ink, shape_elongation]

Separate "0" and "1"
in 2 dimensions



Each example is represented by a little vector $\mathbf{x} = [\, x_1 , x_2 \,]$

Separate “0” and “1”
in 2 dimensions

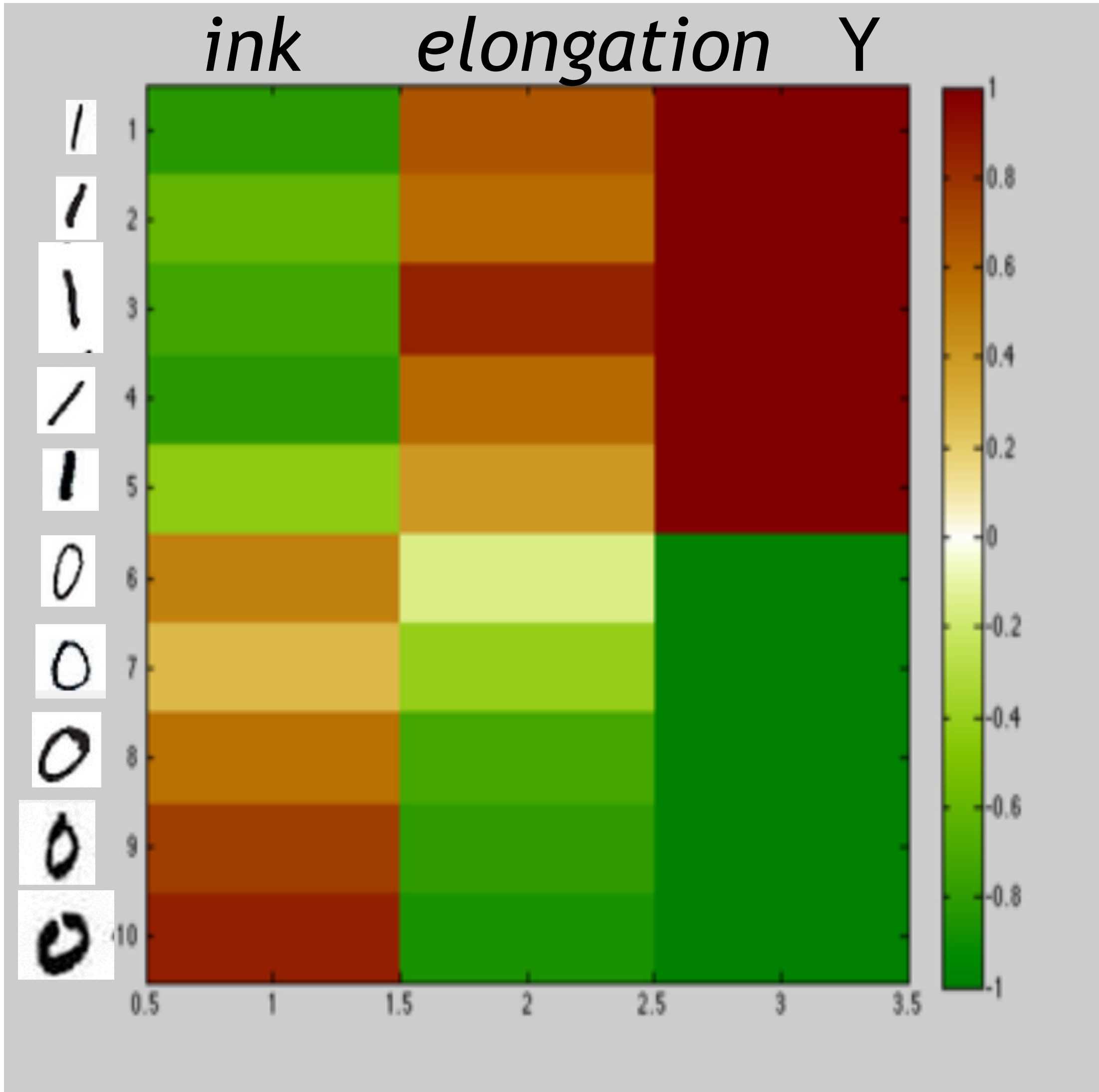


The zeros are written with more ink and are less elongated. The ones are written with less ink and are more elongated.

Separate “0” and “1”
matrix representation

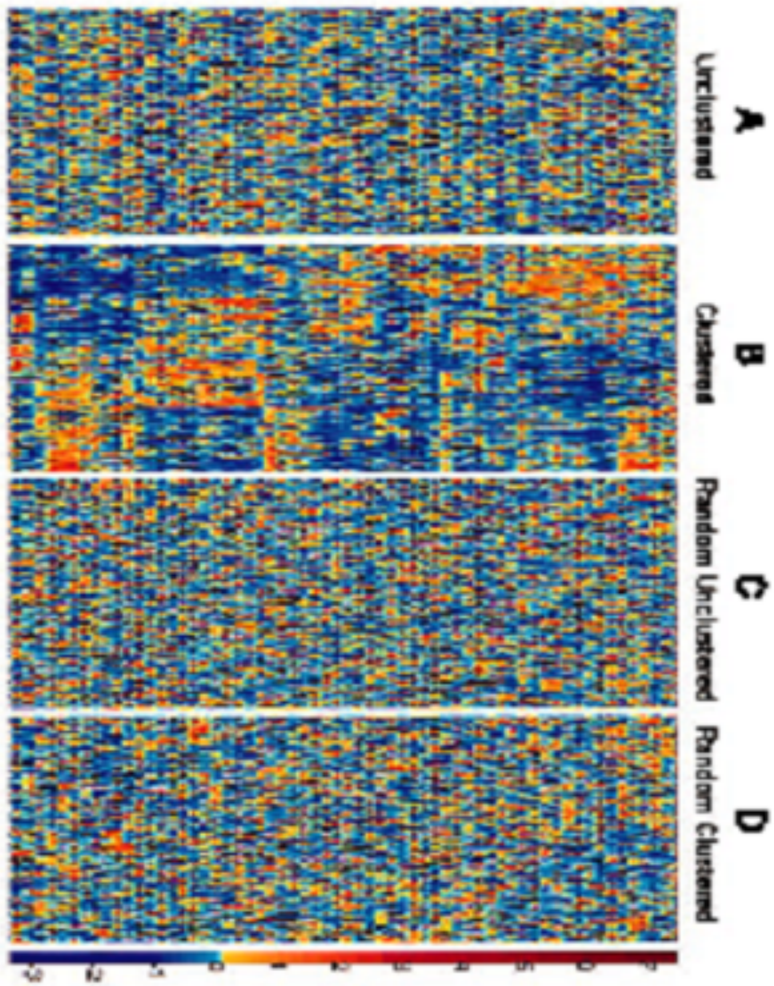
$$
\begin{array}{cc}
 & \begin{array}{cc} \textit{ink} & \textit{elongation} \end{array} \\
X = & \left[\begin{array}{cc}
-0.8147 & 0.6576 \\
-0.6058 & 0.5706 \\
-0.7270 & 0.8572 \\
-0.8134 & 0.5854 \\
-0.4324 & 0.4003 \\
0.4975 & -0.1419 \\
0.2785 & -0.4218 \\
0.5469 & -0.7157 \\
0.7575 & -0.7922 \\
0.8649 & -0.8595
\end{array} \right]
\end{array}
\qquad
\mathbf{y} = \left[\begin{array}{c}
1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1
\end{array} \right]
$$

Heat map



ink
elongation
Y
1
2
3
4
5
6
7
8
9
10
0.5
1
1.5
2
2.5
3
3.5
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Learning problem



Colon cancer, Alon et al 1999

Data matrix: X

N lines = patterns (data points, examples): samples, patients, documents, images, …

d columns = features: (attributes, input variables): genes, proteins, words, pixels, …

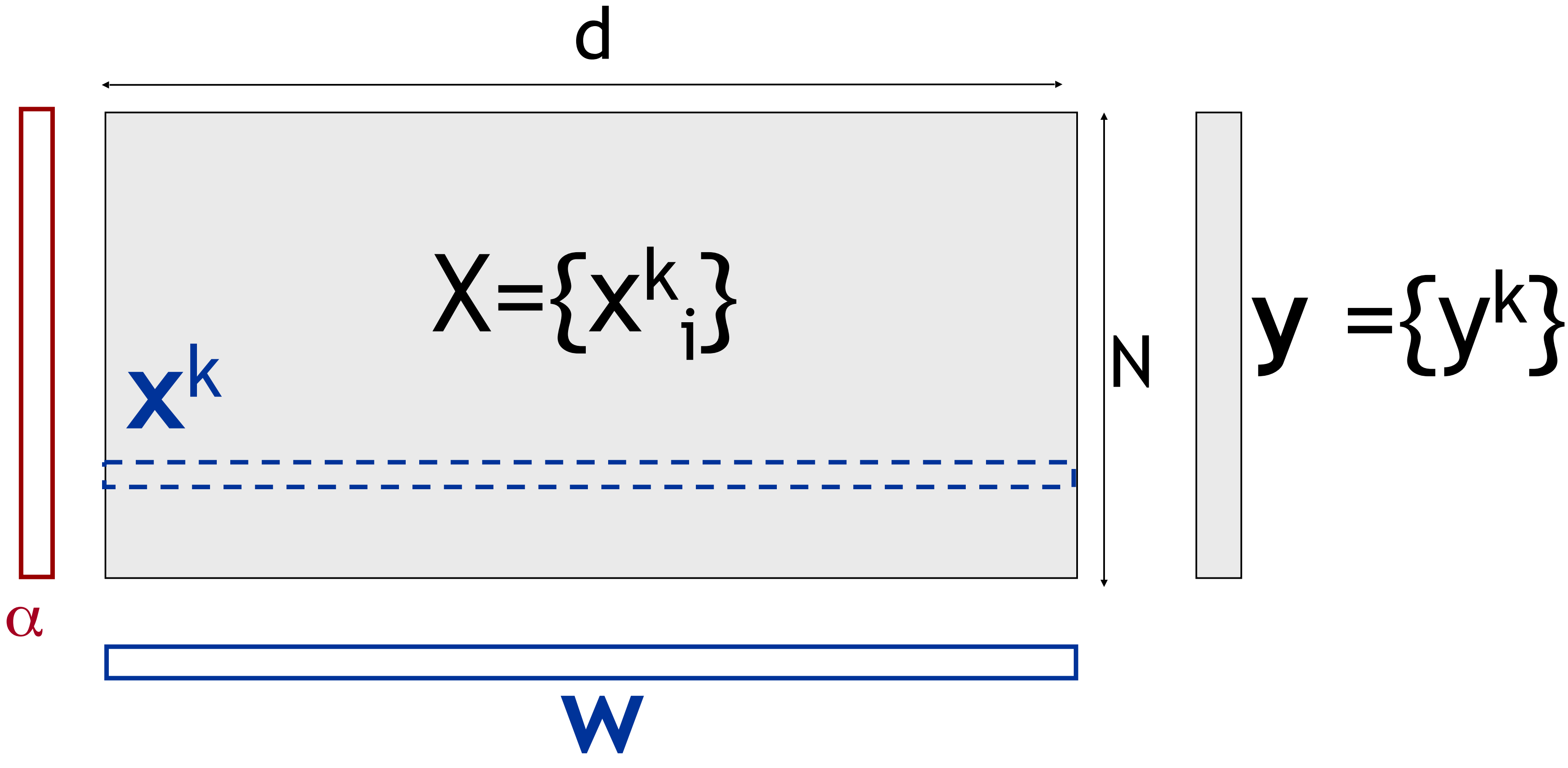
Unsupervised learning

Is there structure in data?

Supervised learning

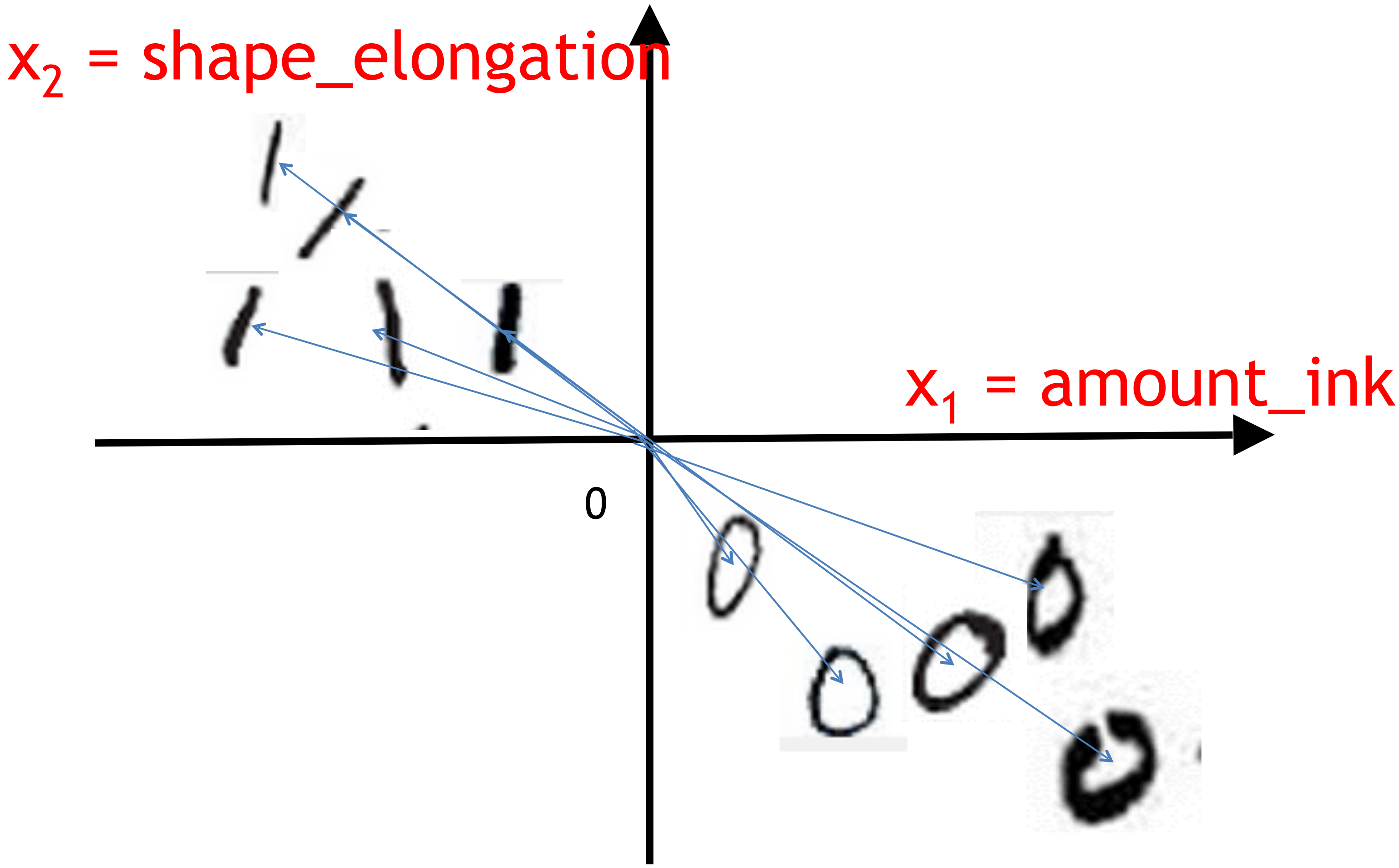
Predict an outcome **y**.

Conventions



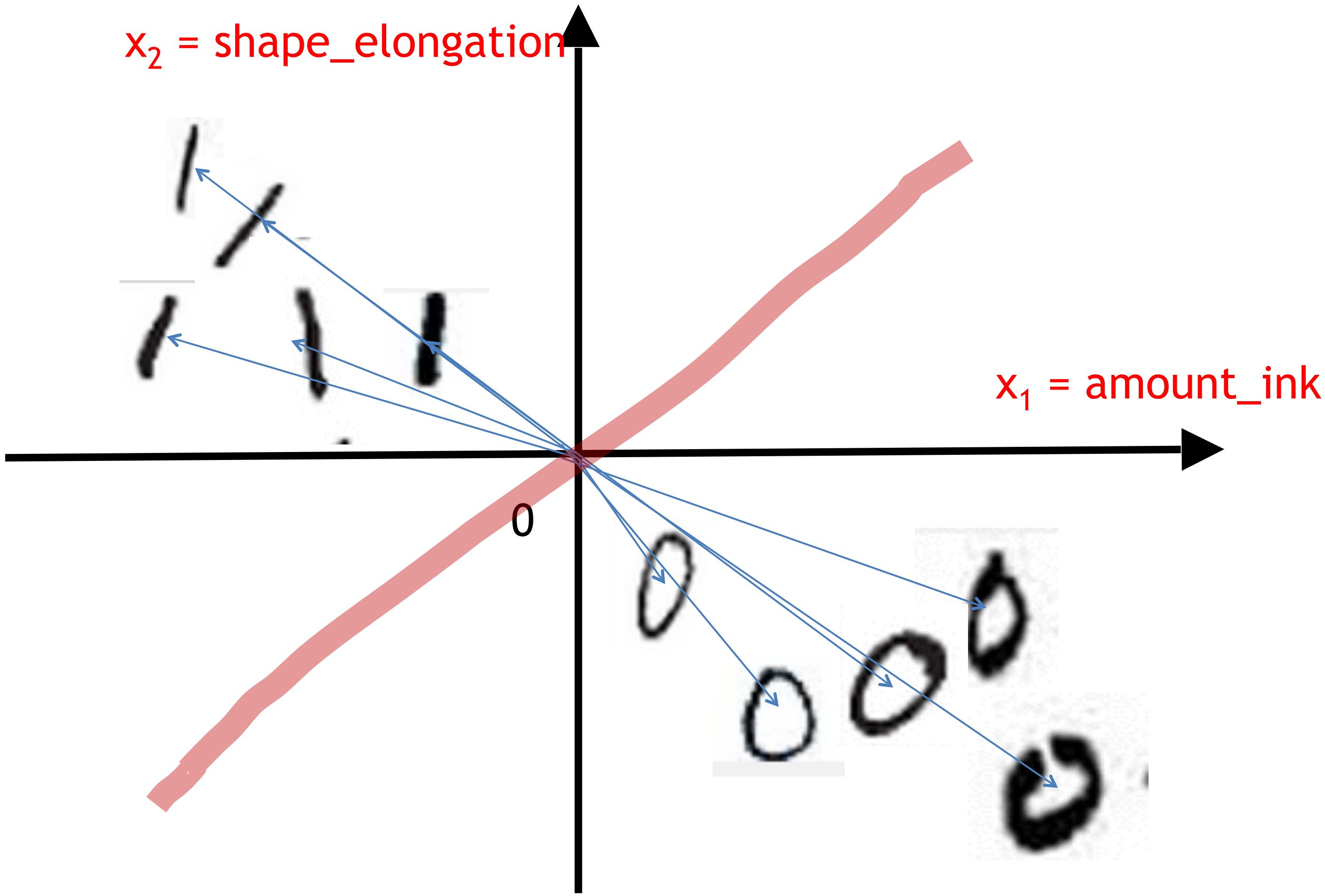
Separate “0” and “1”

How to build your first linear classifier?



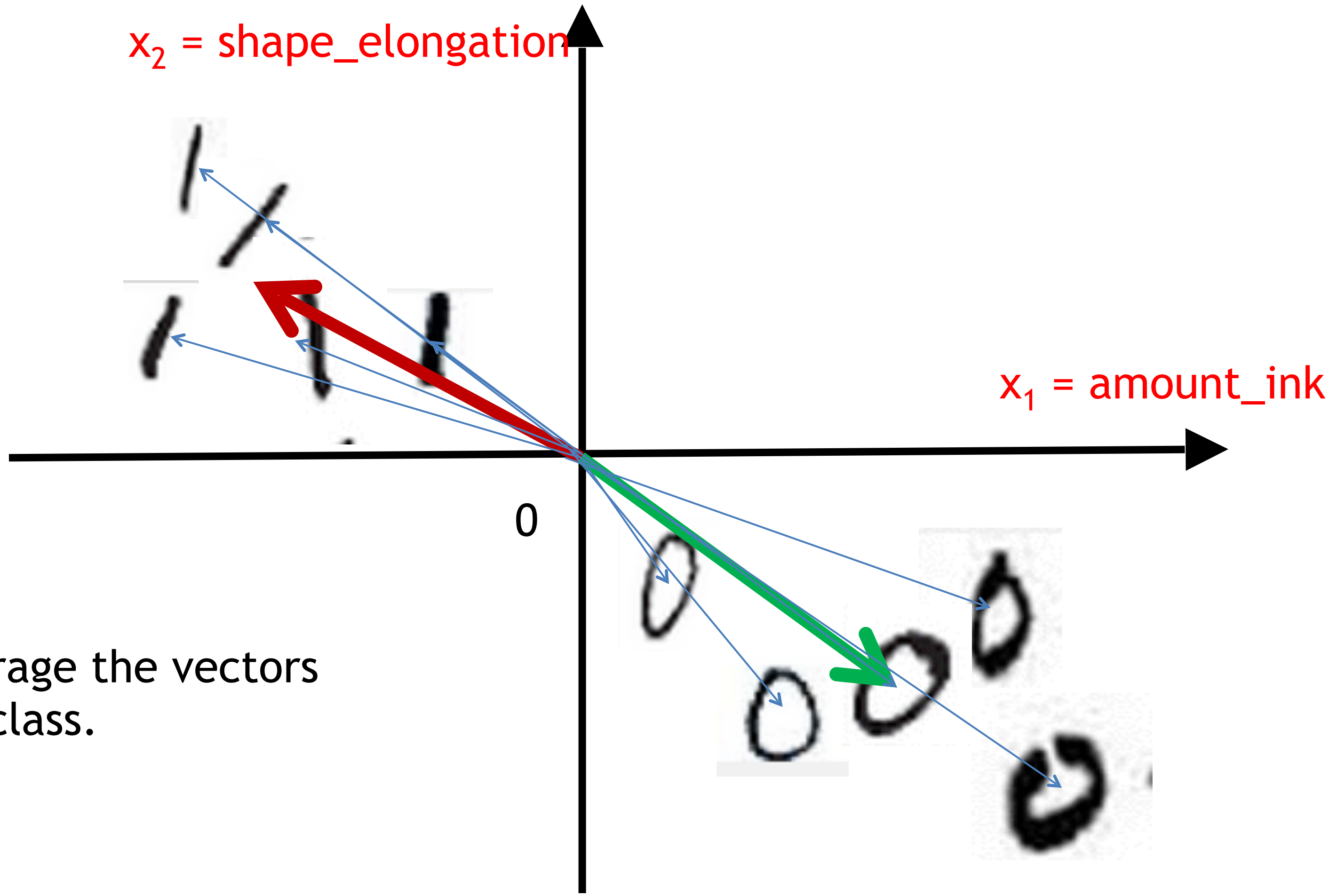
Separate “0” and “1”

How to build your first linear classifier?



Separate "0" and "1"

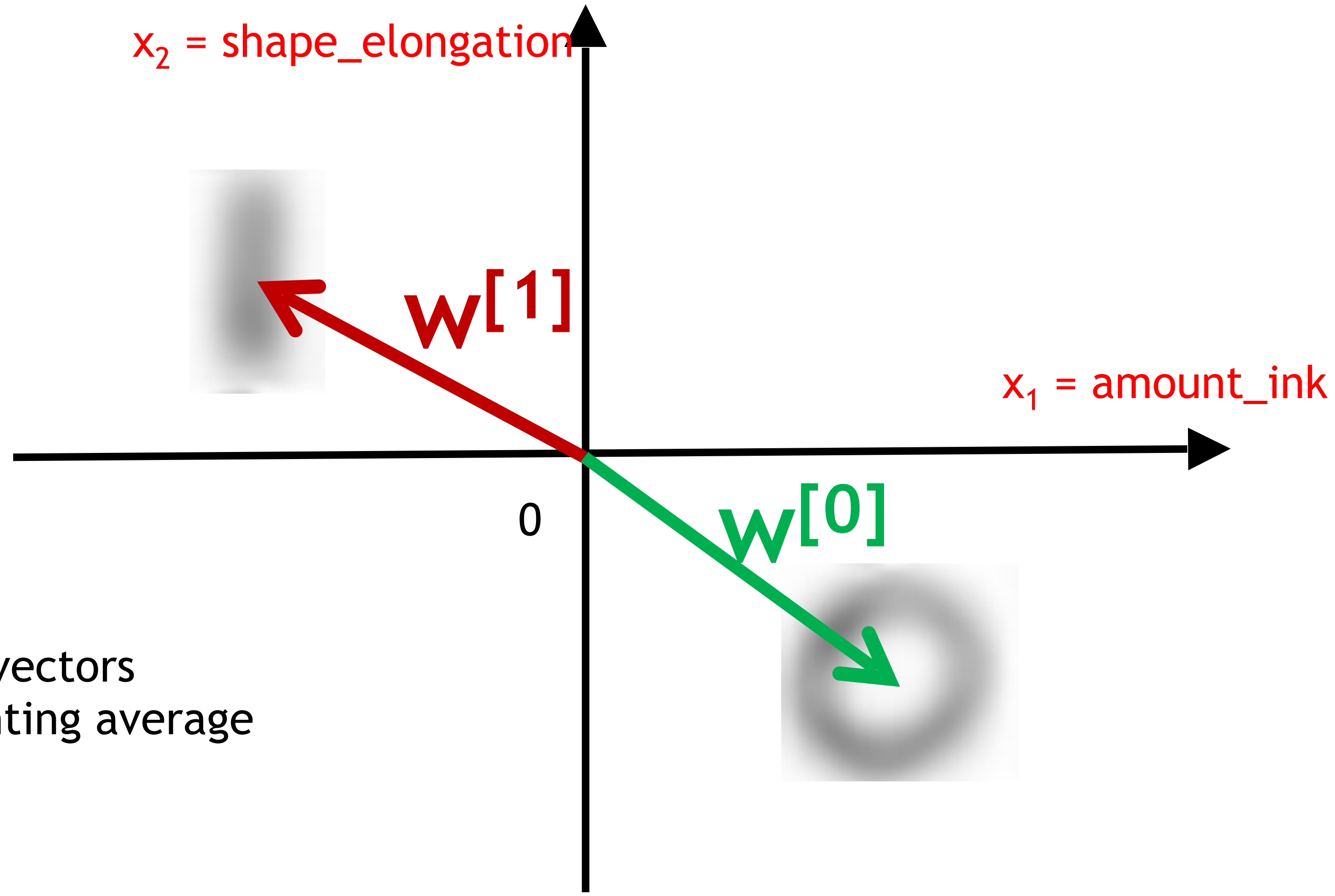
1) Find the class "centroids"



Just average the vectors of each class.

Separate “0” and “1”

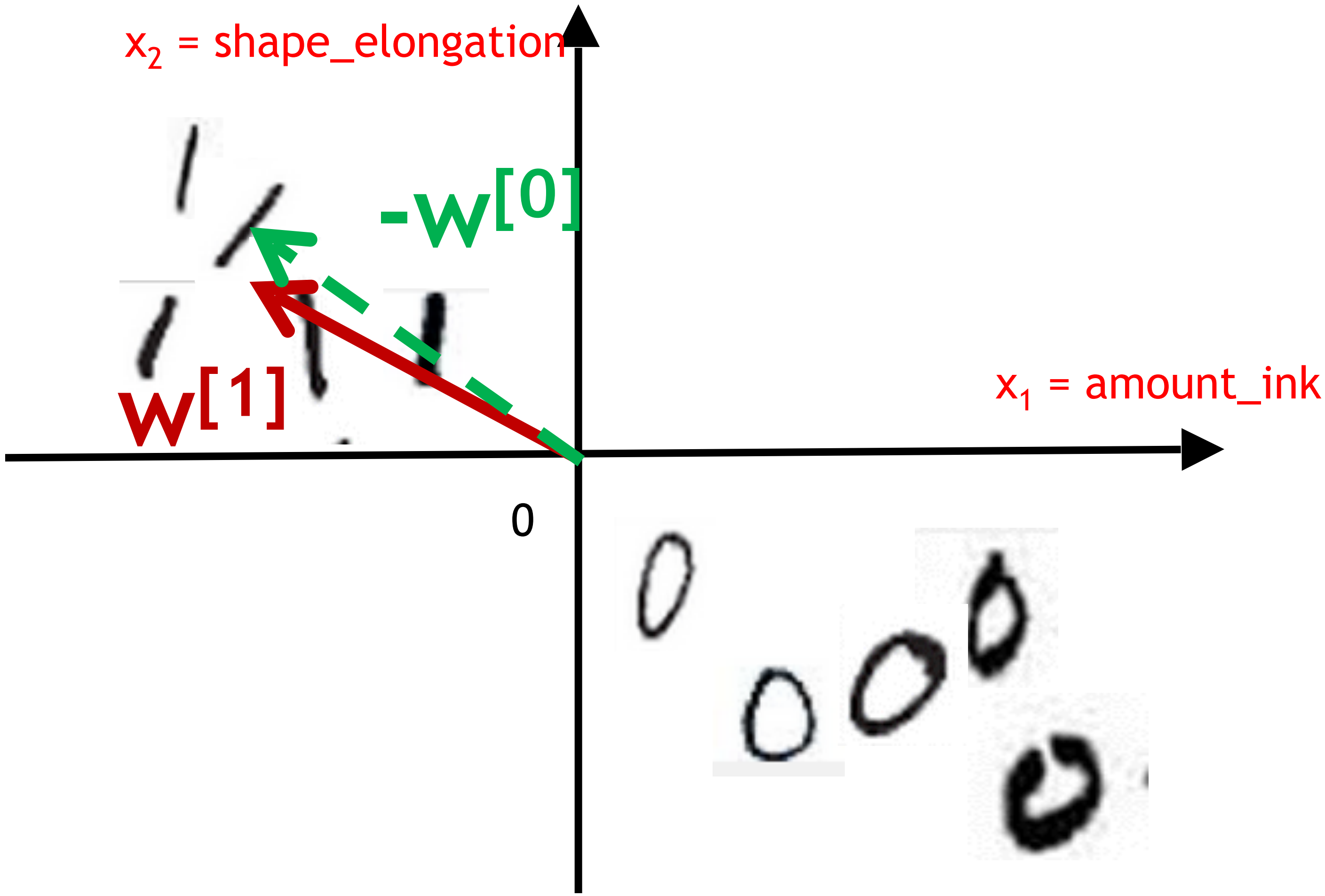
1) Find the class “centroids”



You get vectors representing average shapes.

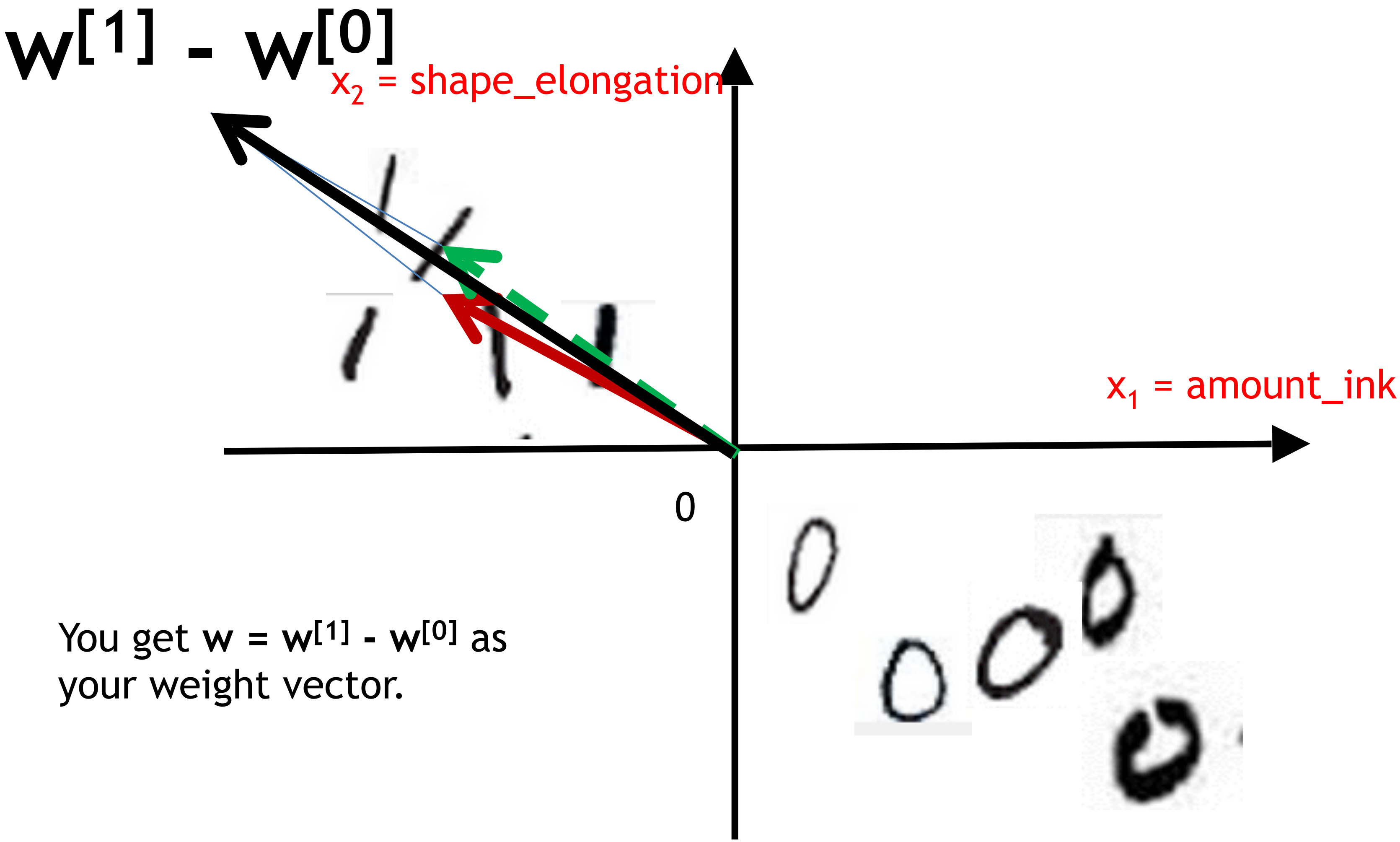
Separate "0" and "1"

2) Subtract $w^{[0]}$ from $w^{[1]}$



Separate “0” and “1”

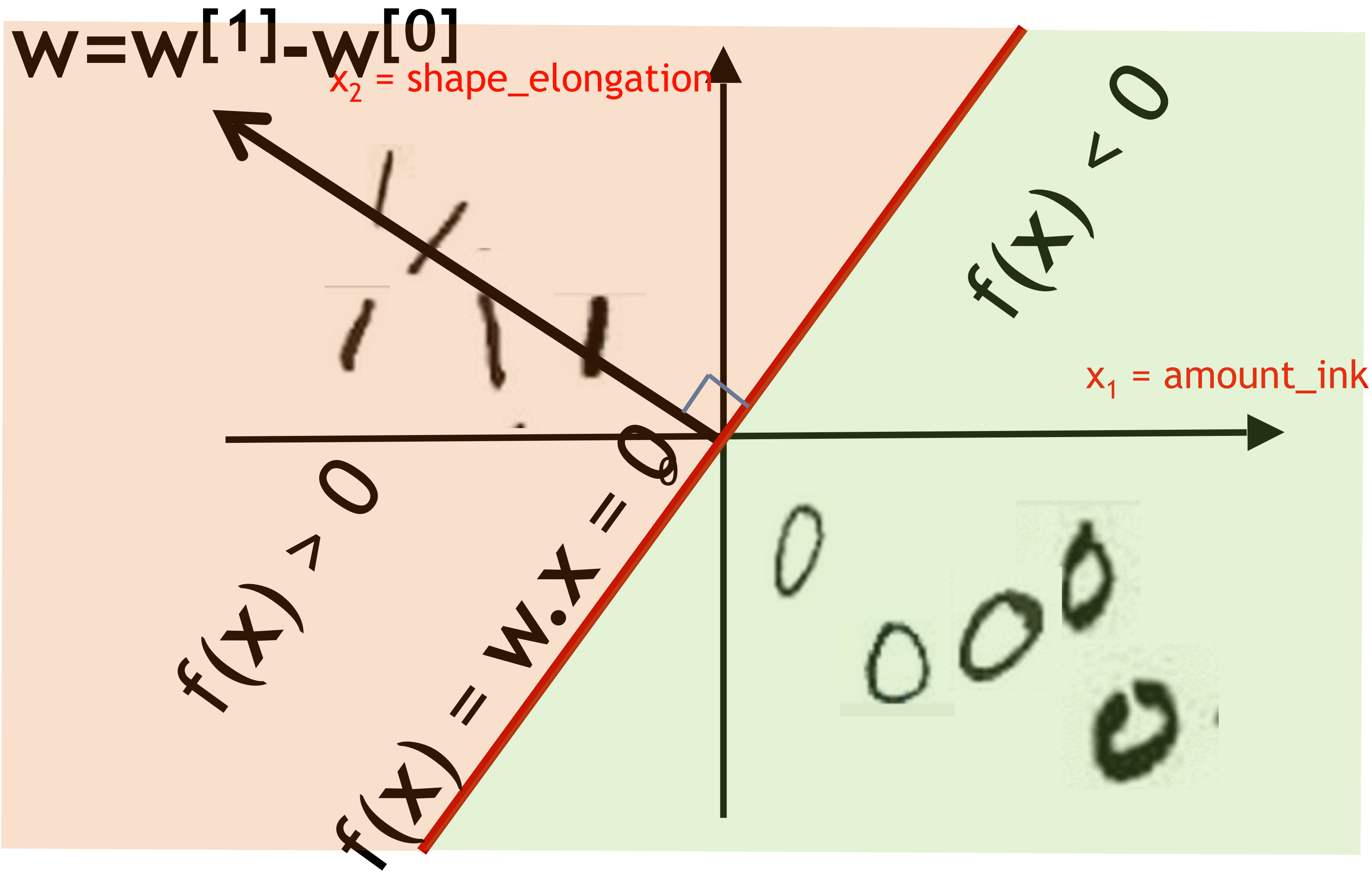
2) Subtract $w^{[0]}$ from $w^{[1]}$



You get $w = w^{[1]} - w^{[0]}$ as your weight vector.

Separate “0” and “1”

3) Get the separaring “hyperplane”



Separate “0” and “1”

4) ALL the “math”

Input vector $\mathbf{x} = [x_1, x_2]$

x_1 = amount_ink

x_2 = shape_elongation

Target value $y = \pm 1$

+1 = “one”; -1 = “zero”

Training examples:

$\{ (\mathbf{x}^1, y^1), (\mathbf{x}^2, y^2), \ldots , (\mathbf{x}^N, y^N) \}$

Class centroids:

$\mathbf{w}^{[0]} \sim \Sigma_{\{y^k==-1\}} \mathbf{x}^k$ (~ means “proportional”, omitting to divide by class cardinality)

$\mathbf{w}^{[1]} \sim \Sigma_{\{y^k==+1\}} \mathbf{x}^k$

Weight vector:

$\mathbf{w} = \mathbf{w}^{[1]} - \mathbf{w}^{[0]} \sim \Sigma_k\, y^k \mathbf{x}^k$ (for “balanced” classes)

Separate "0" and "1"

4) ALL the "math" (continued)

Decision function:

$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$

$f(\mathbf{x}) > 0$, decide that this is a "one"

$f(\mathbf{x}) < 0$, decide that this is a "zero"

Dot product:

$\mathbf{w} \cdot \mathbf{x} = w_1 x_1 + w_2 x_2$ This is a weighted sum.

Equivalent "centroid" method:

$f(\mathbf{x}) = \mathbf{w}^{[1]} \cdot \mathbf{x} - \mathbf{w}^{[0]} \cdot \mathbf{x}$ This is because $\mathbf{w} = \mathbf{w}^{[1]} - \mathbf{w}^{[0]}$.

Decide "one" if $\mathbf{w}^{[1]} \cdot \mathbf{x} > \mathbf{w}^{[0]} \cdot \mathbf{x}$ and "zero" otherwise

A dot product is a similarity measure.

Equivalent "kernel" method: This is because $\mathbf{w} = \Sigma_k\, y^k\, \mathbf{x}^k$

$f(\mathbf{x}) = \Sigma_k\, y^k\, \mathbf{x}^k \cdot \mathbf{x} = \Sigma_k\, \alpha^k\, k(\mathbf{x}^k, \mathbf{x})$ (in the case of identical

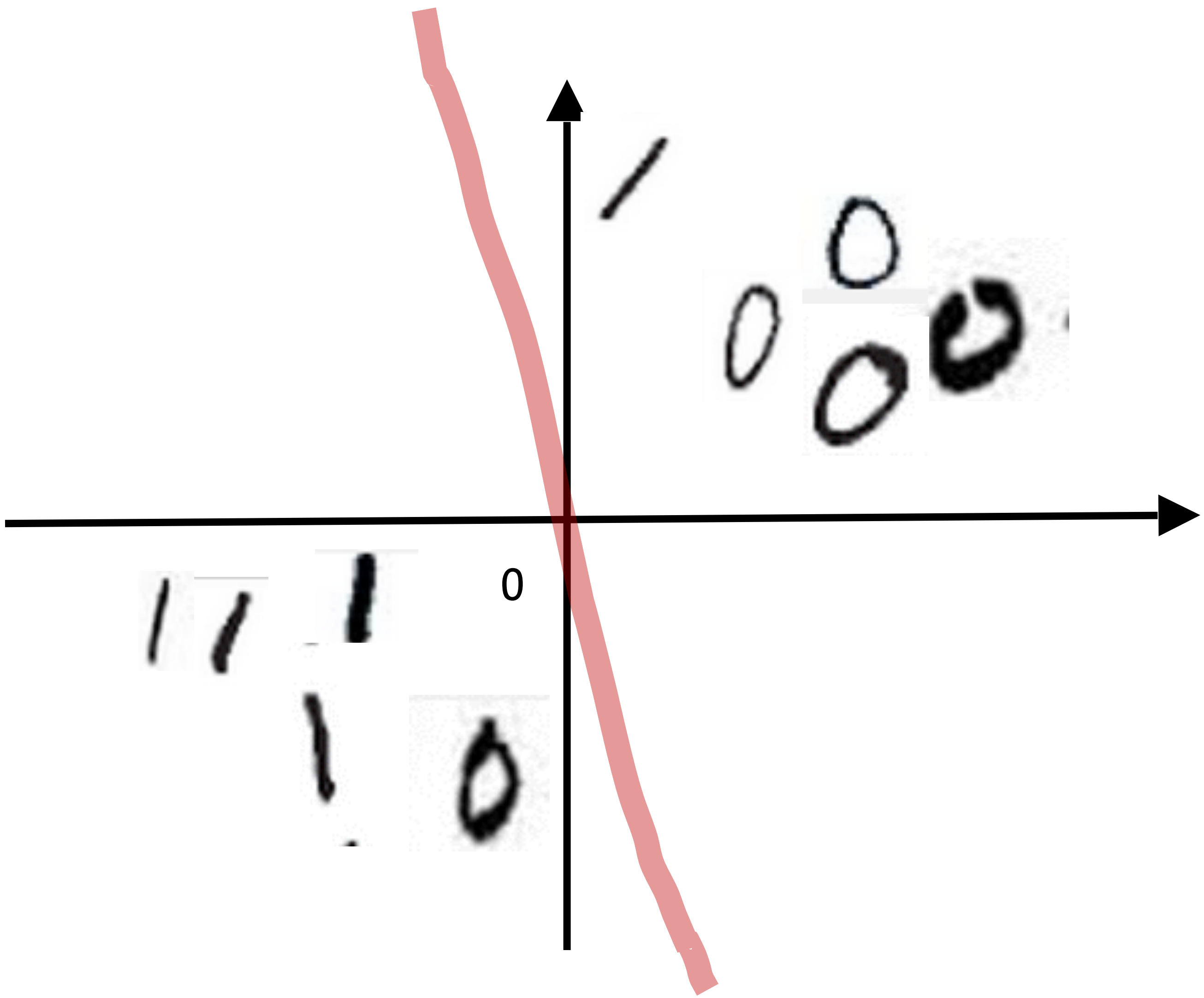
number of

classes)

examples for the two

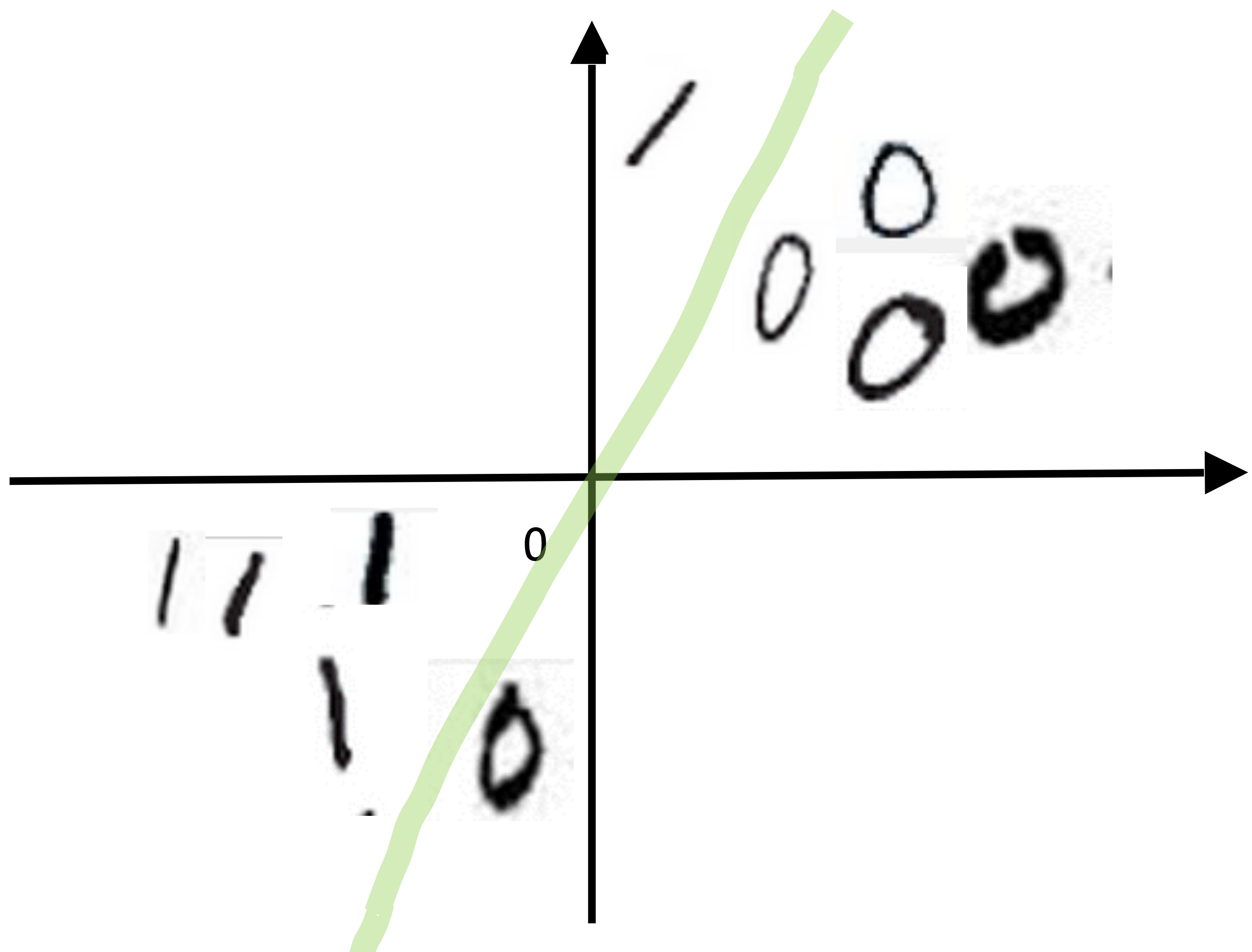
Draw the Boundaries: Centroids

Centroid methods don’t always work…

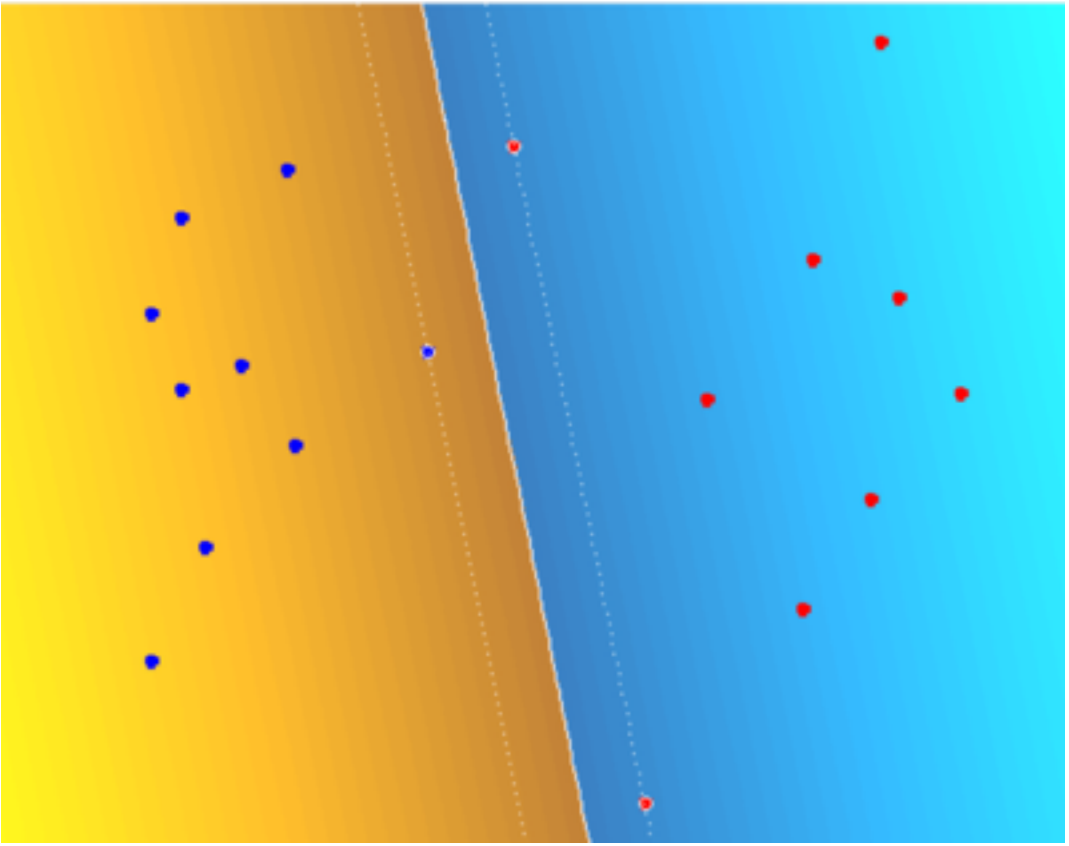


Draw the Boundaries: SVM

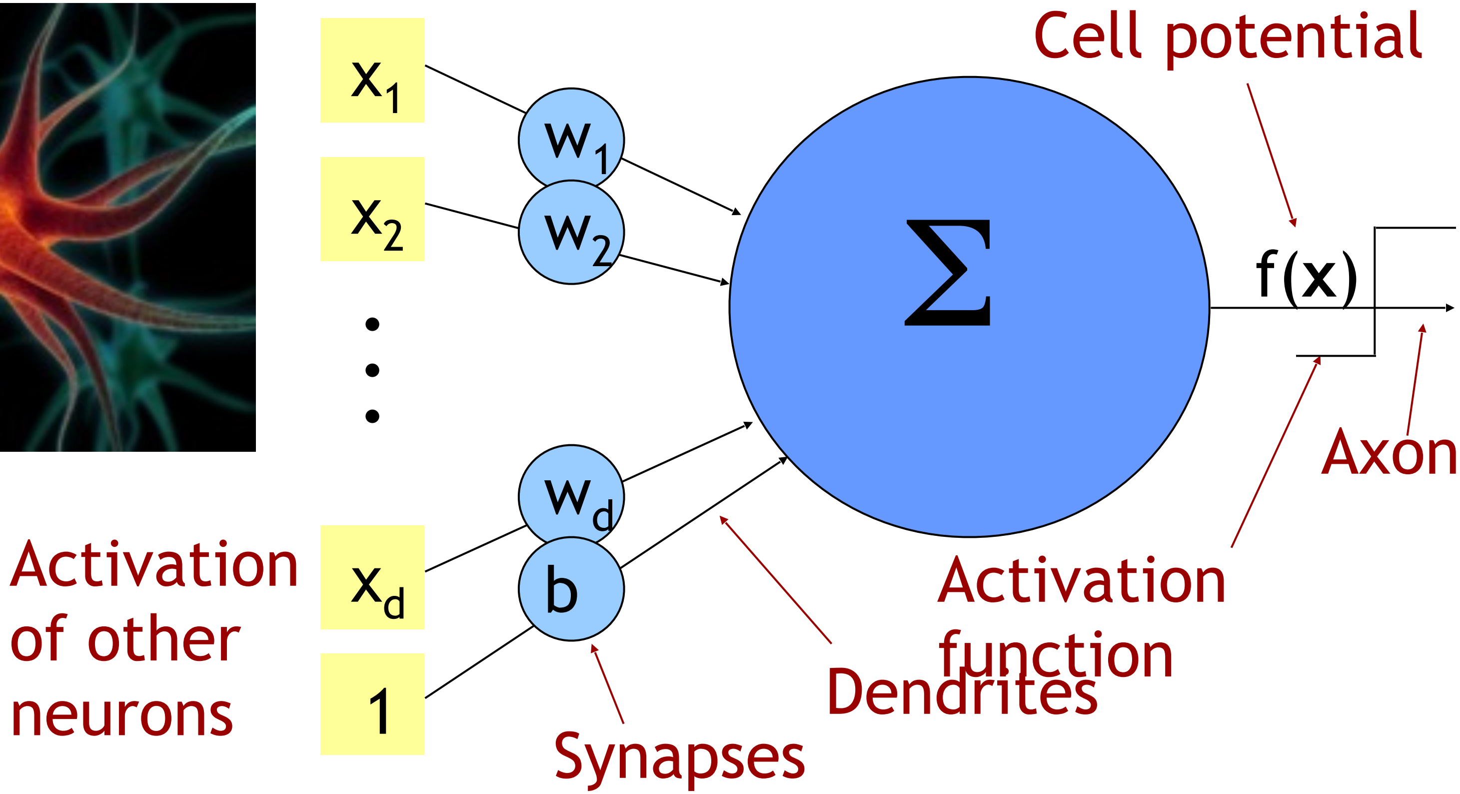
Centroid methods don't always work… but SVM words here…



Demo of SVM



Artificial Neurons

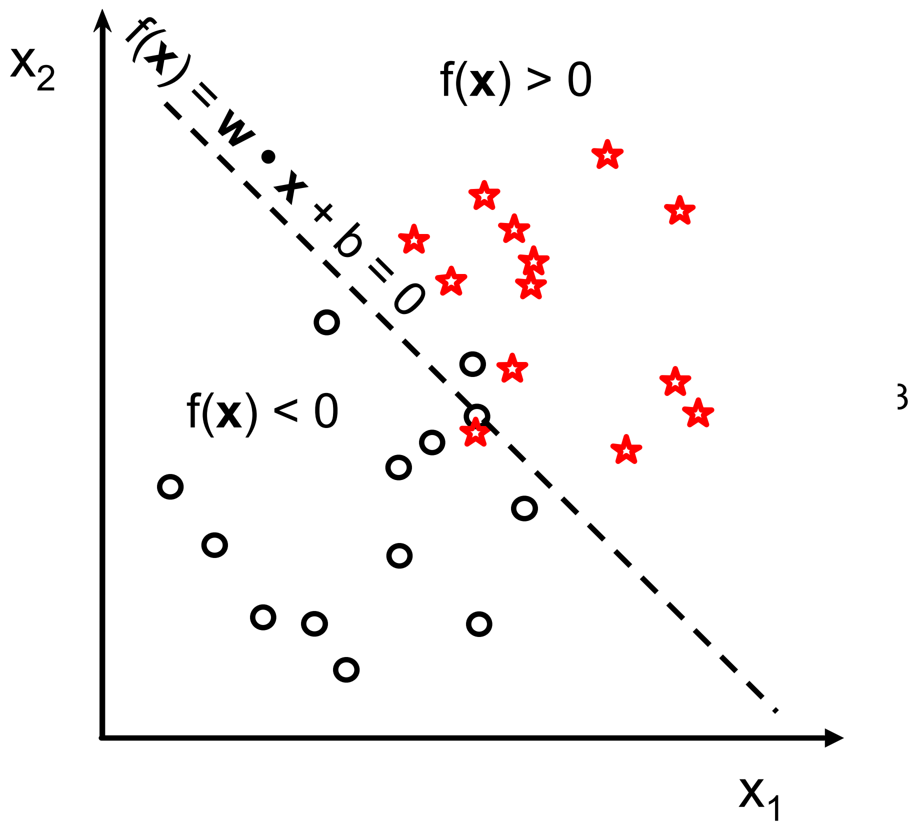


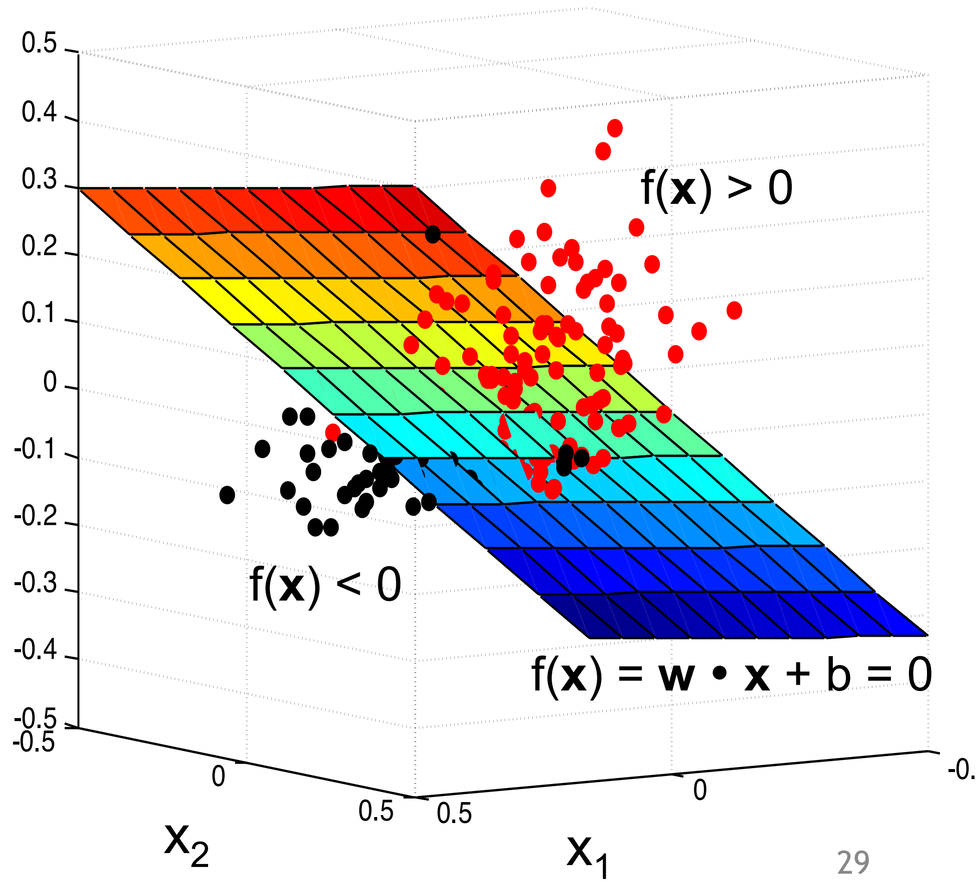
McCulloch and Pitts, 1943

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$$

Linear decision boundary

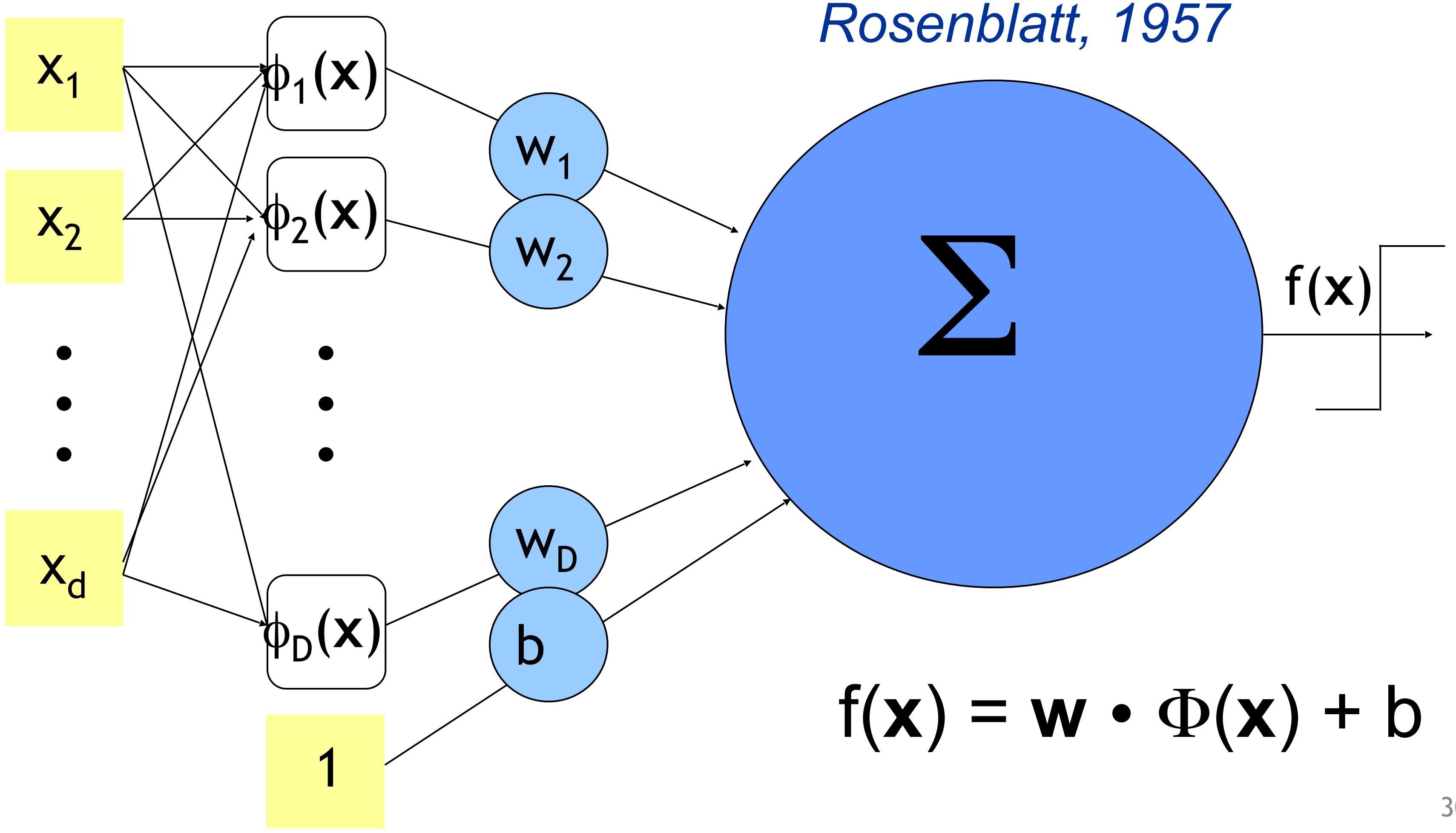
hyperplane





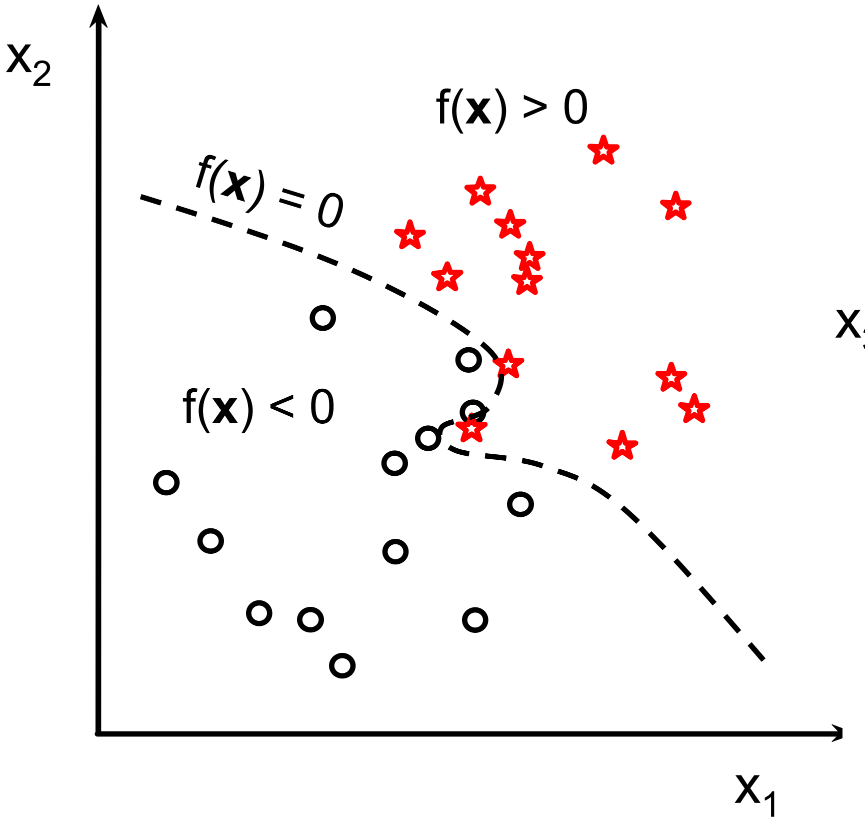
Perceptron

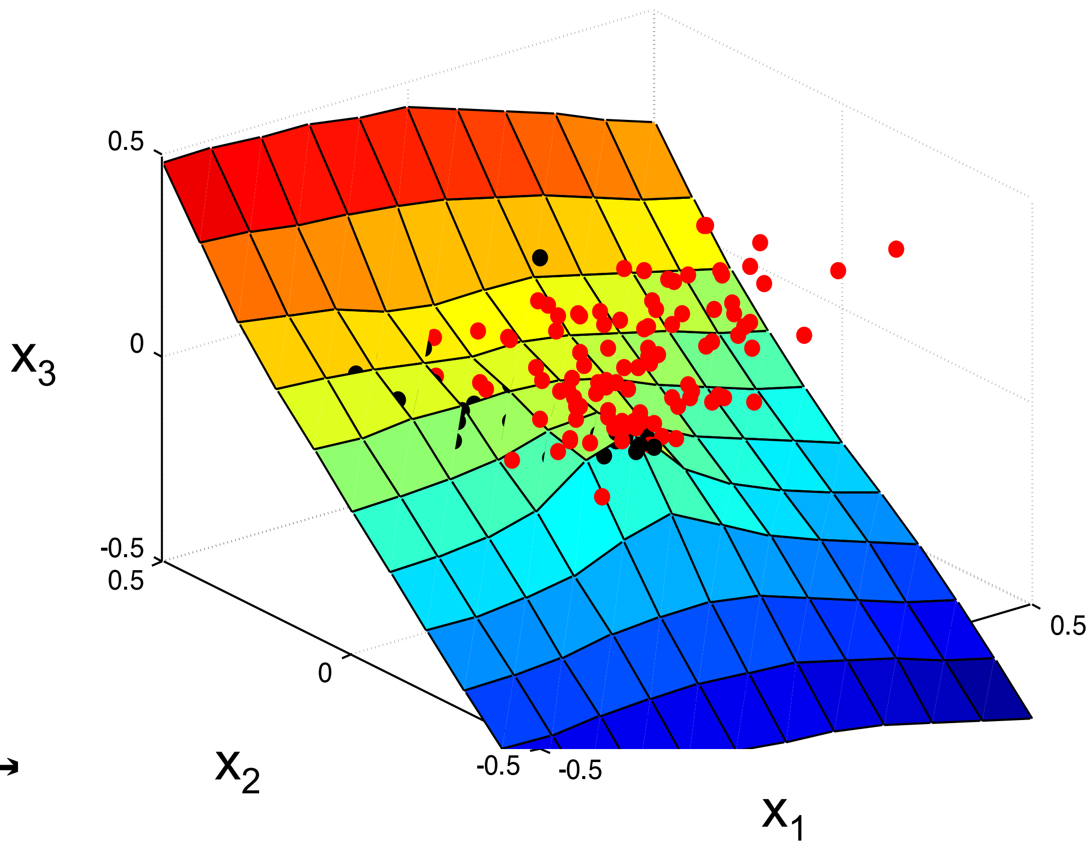
Rosenblatt, 1957



$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b$$

Non-linear decision boundary





Summary

- We represent patterns as vectors **x** in a space of d dimensions.
- A "discriminant function" f(**x**) is a function such that f(**x**) > 0 for one class and f(**x**) < 0 for the other. f(**x**)=0 is the equation of the decision boundary.
- Given a weight vector **w**, f(**x**)=**w.x** is a linear discriminant function. The corresponding decision boundary **w.x**=0 is a hyperplane (a subspace of dimension (d-1)).