

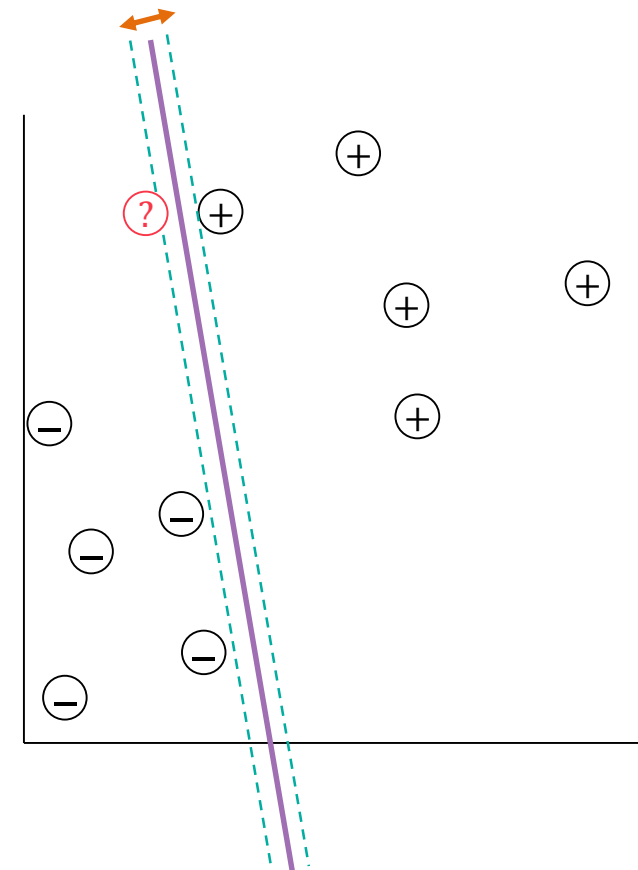
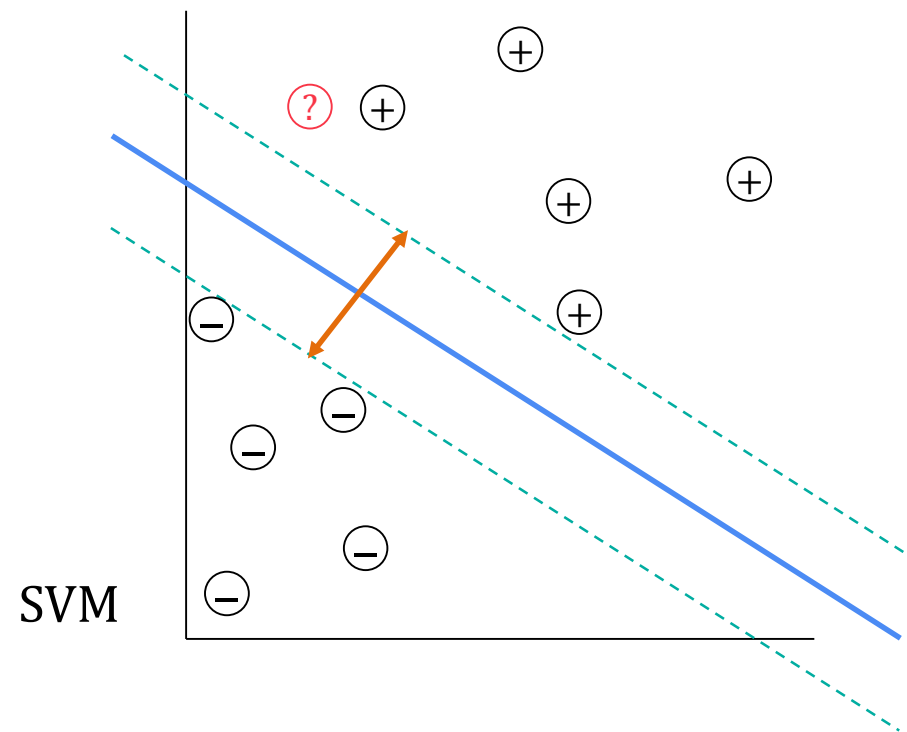
데이터사이언스응용 (Capstone design)

김응희

ehkim@sunmoon.ac.kr

Week 09

Last week,



Decision tree/
Perceptron

Goal of machine learning

Optimization

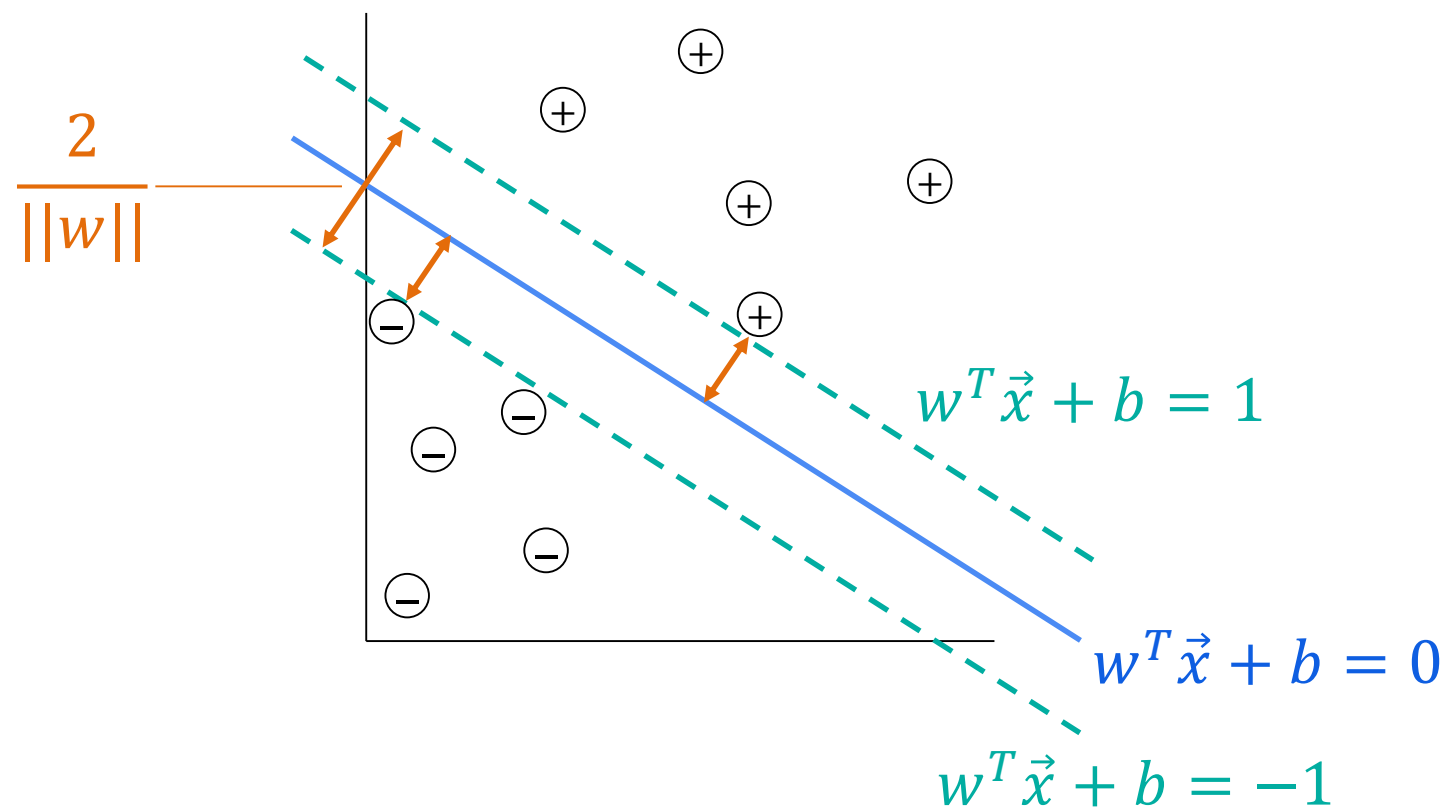
Generalization

Last week,

- Equation of lines and planes
- Distance from a point to a plane
- Support vector machine part I
- Constrained optimization
- Support vector machine part II

Last week,

$$\begin{aligned} &\text{Minimize } \frac{1}{2} ||w||^2 \\ &\text{subject to } y_i(w^T \vec{x}_i + b) \geq 1, \forall i \in n \end{aligned}$$



Last week,

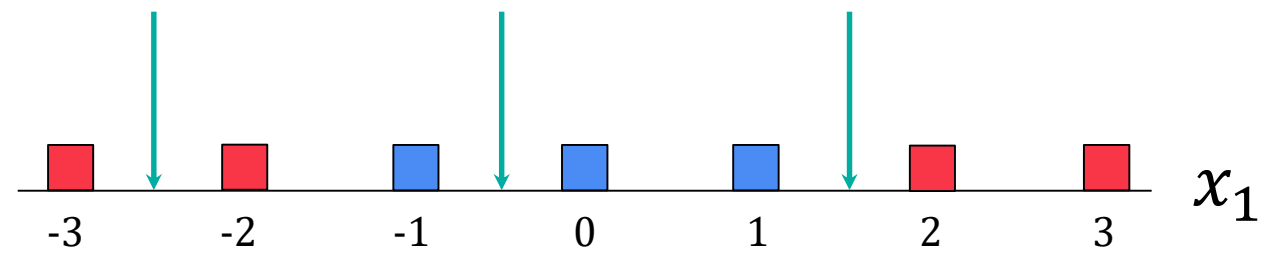
Main learning is about

SPACE

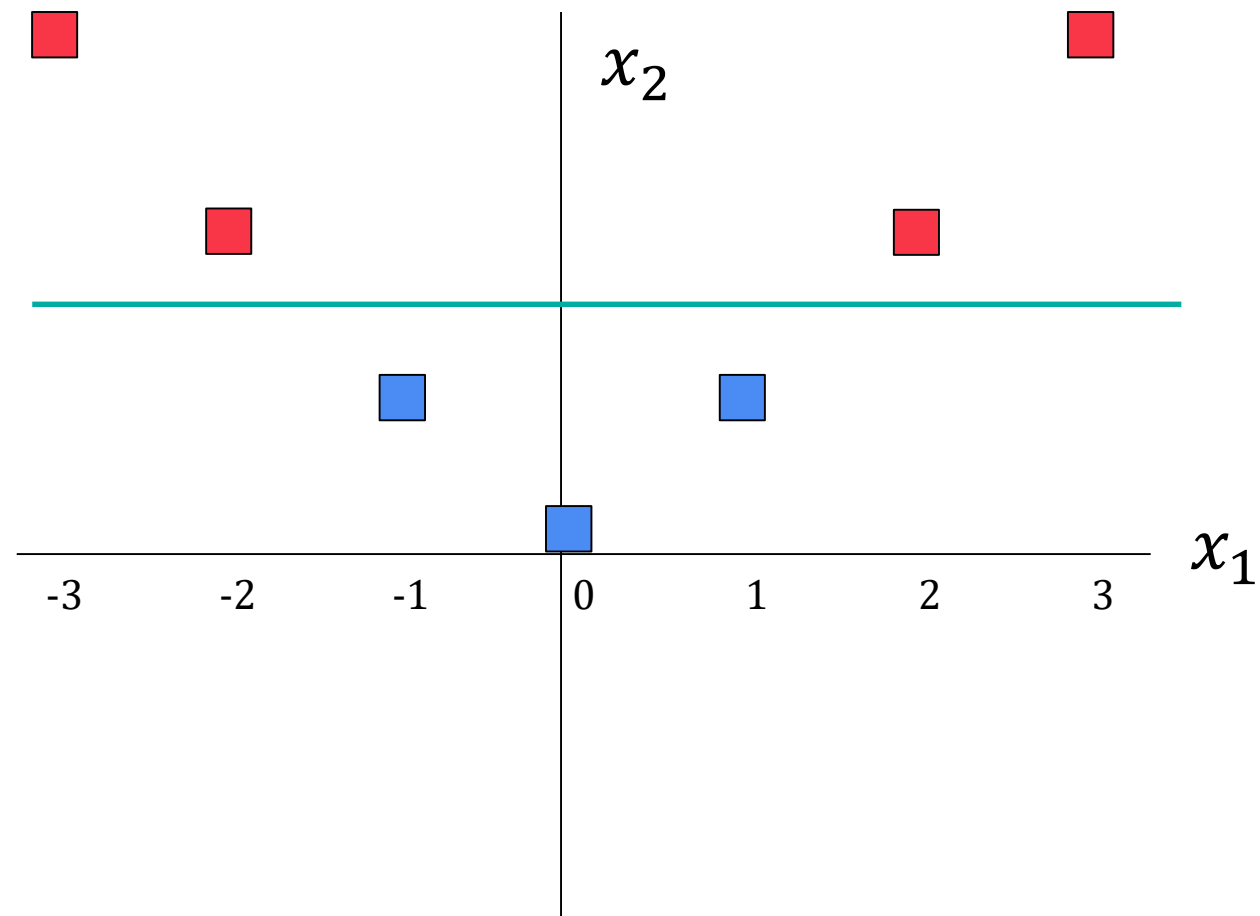
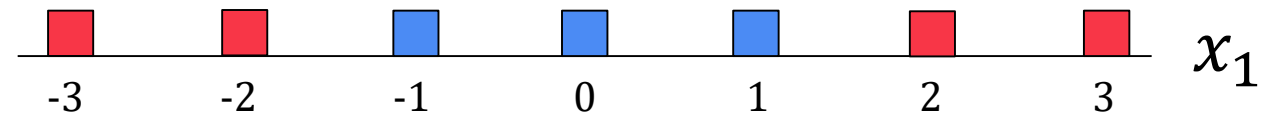
.

Even so,

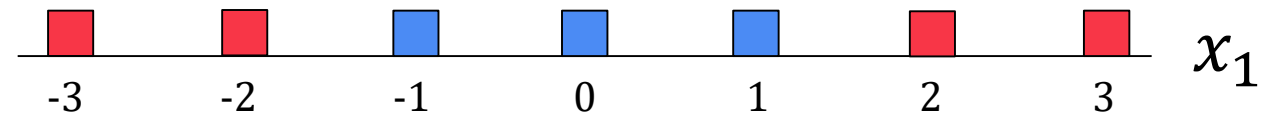
?



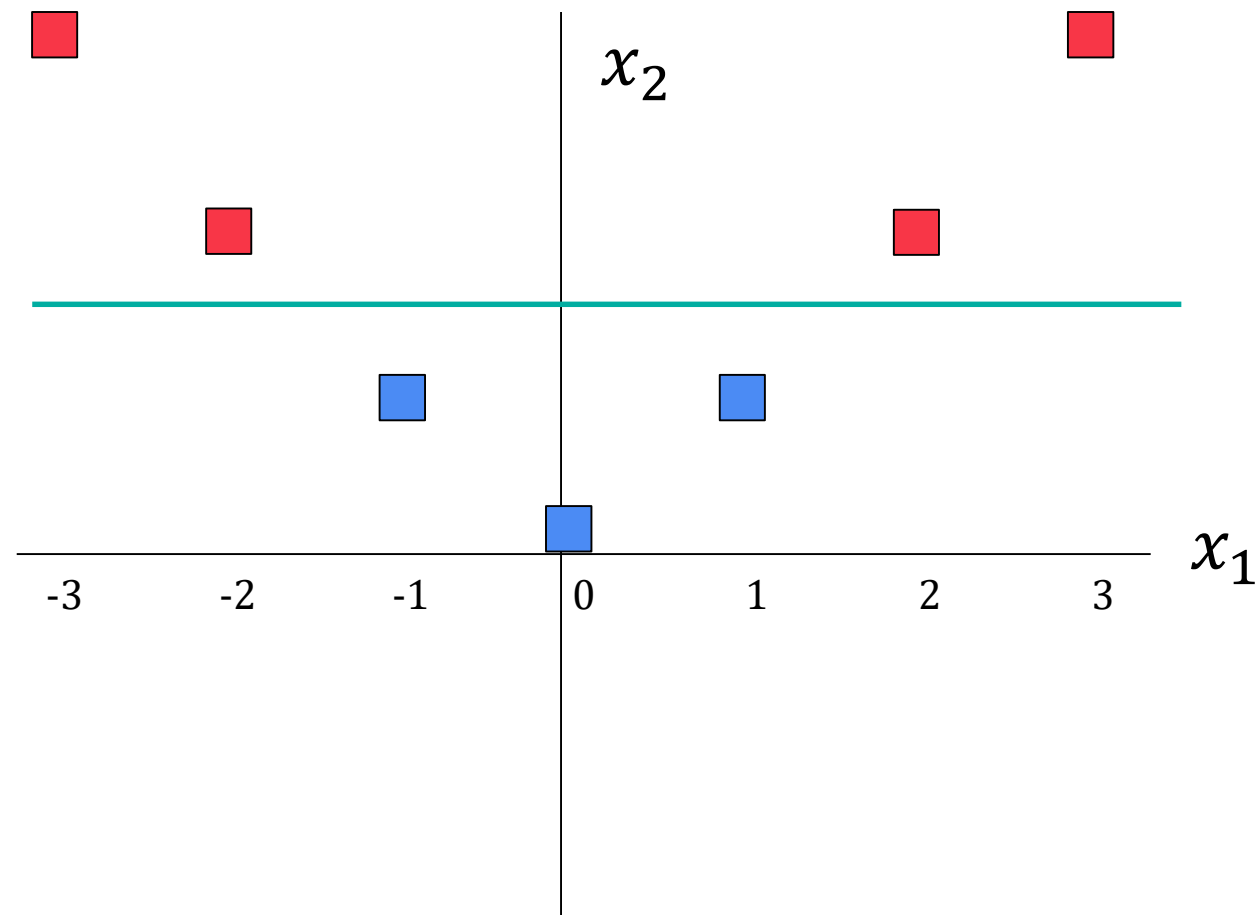
Even so,



Even so,



Kernel function/trick



Popular kernel functions (trick)

Fisher kernel

Graph kernels

Kernel smoother

Polynomial kernel

Radial basis function kernel (RBF)

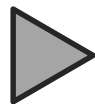
String kernels

Neural tangent kernel

Neural network Gaussian process (NNGP) kernel

Unfortunately, however,

- Equation of lines and planes
- Distance from a point to a plane
- Support vector machine part I
- **Constrained optimization**
- **Support vector machine part II**



Fisher **kernel**

Graph **kernels**

Kernel smoother

Polynomial **kernel**

Radial basis function **kernel** (RBF)

String **kernels**

Neural tangent **kernel**

Neural network Gaussian process (NNGP) **kernel**

Popular kernel functions (trick)

Fisher kernel

Graph kernels

Kernel smoother

Polynomial kernel

Radial basis function kernel (RBF)

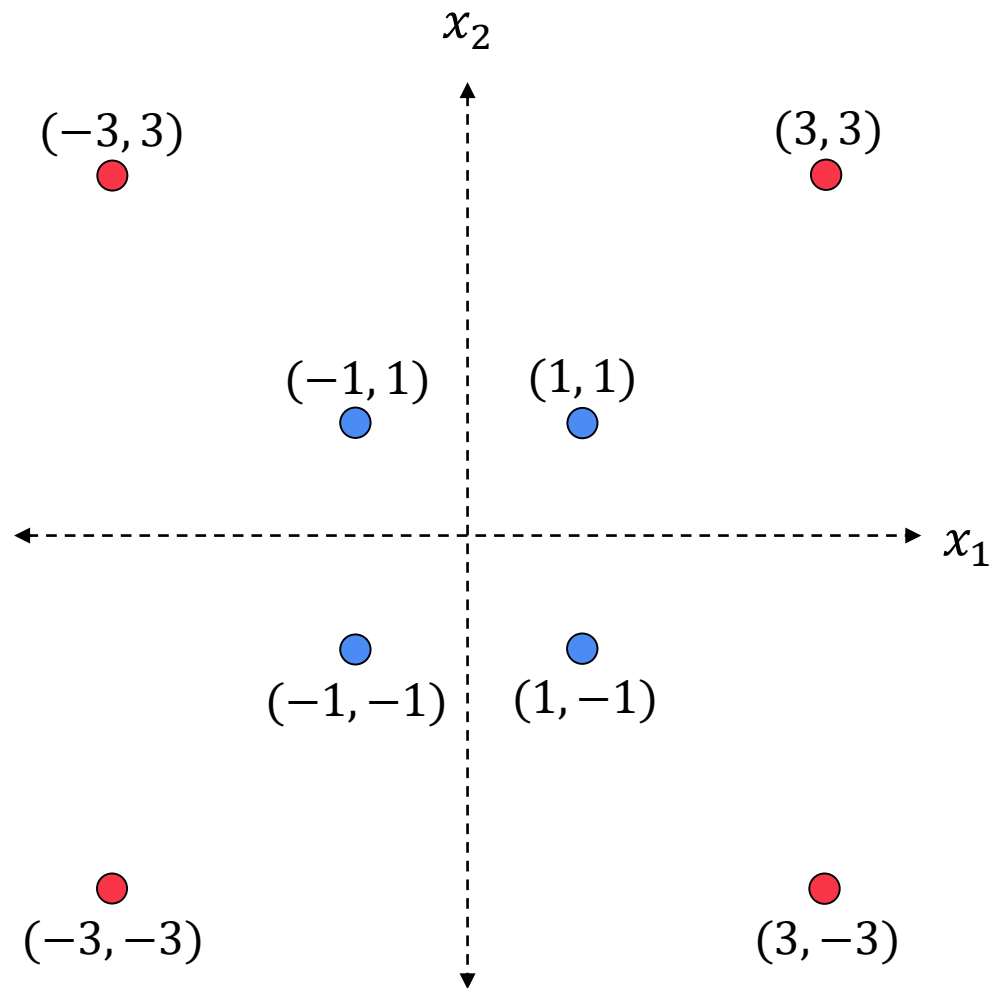
String kernels

Neural tangent kernel

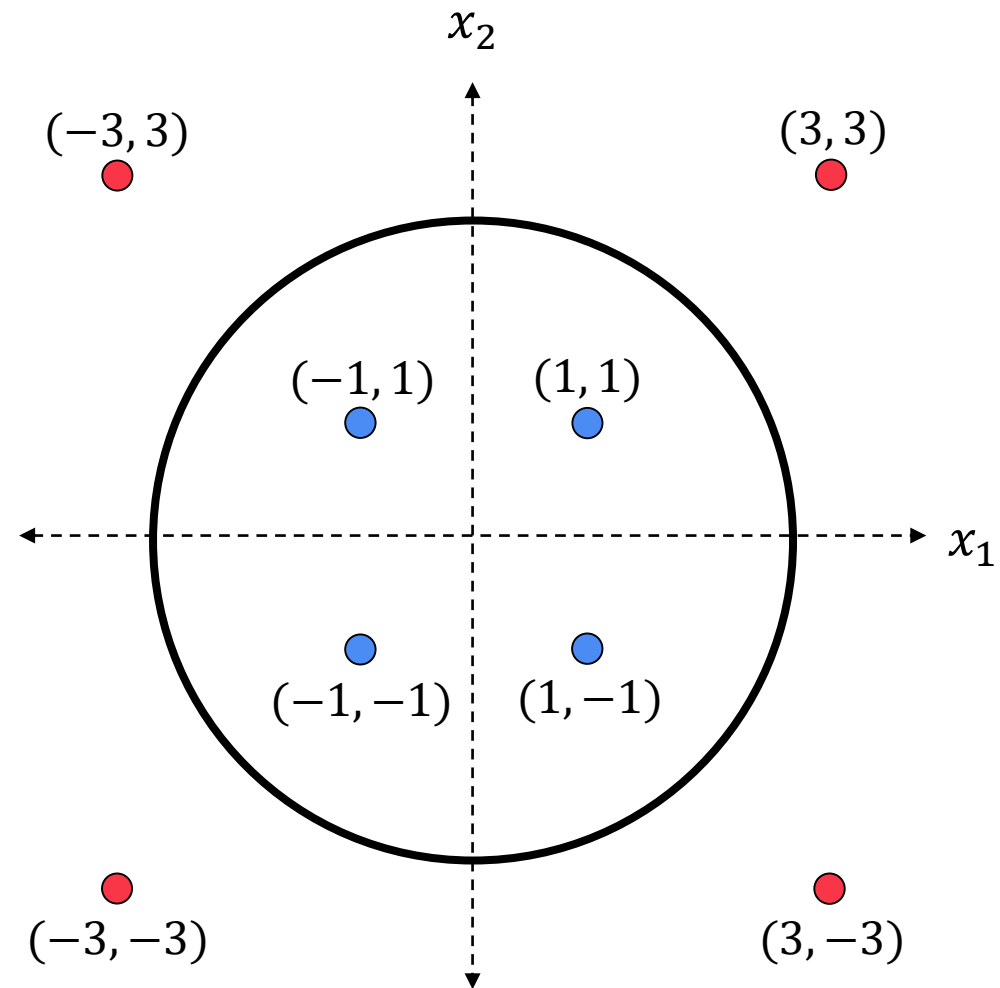
Neural network Gaussian process (NNGP) kernel

Concept of polynomial kernel function

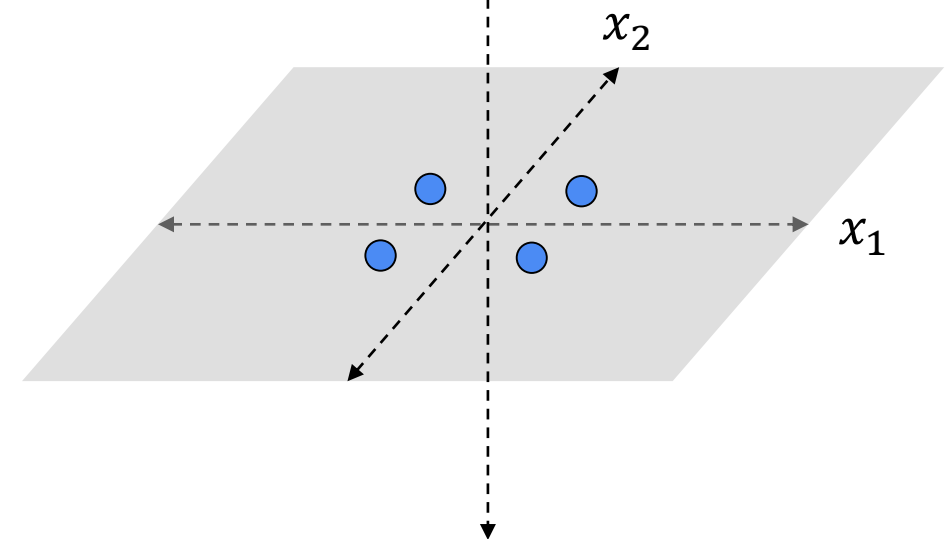
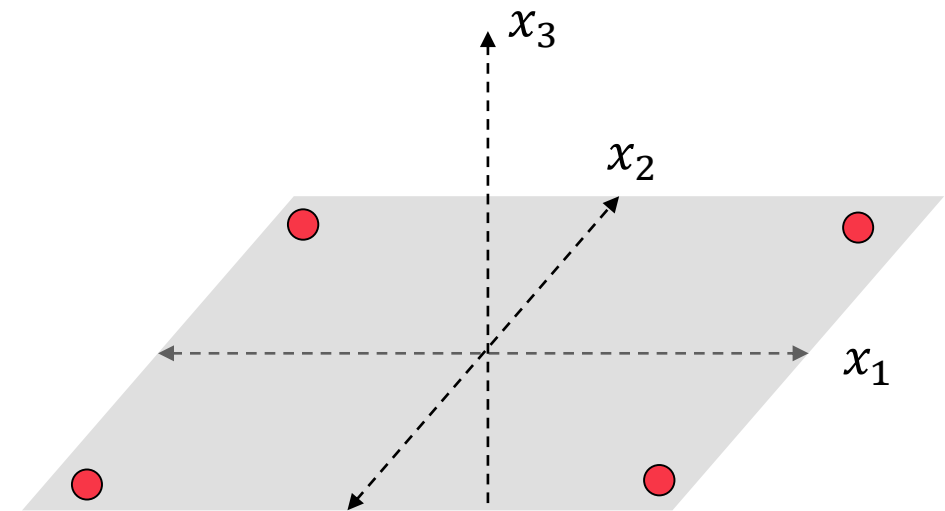
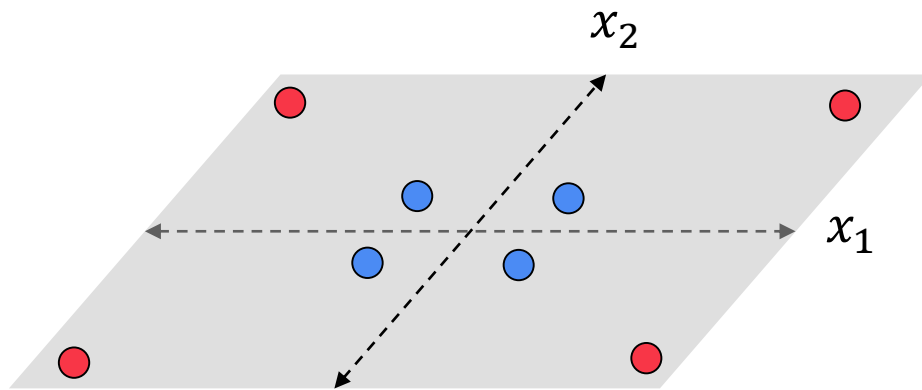
Nonlinearly separable data



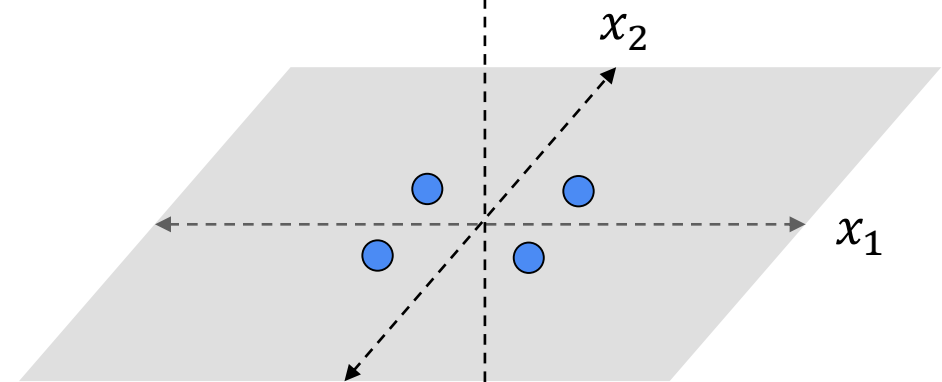
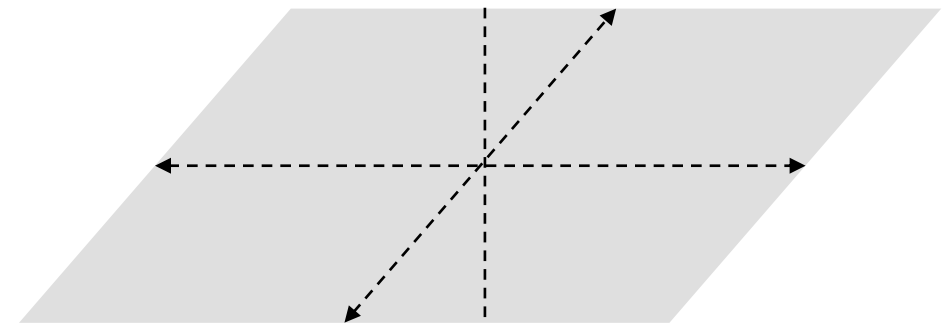
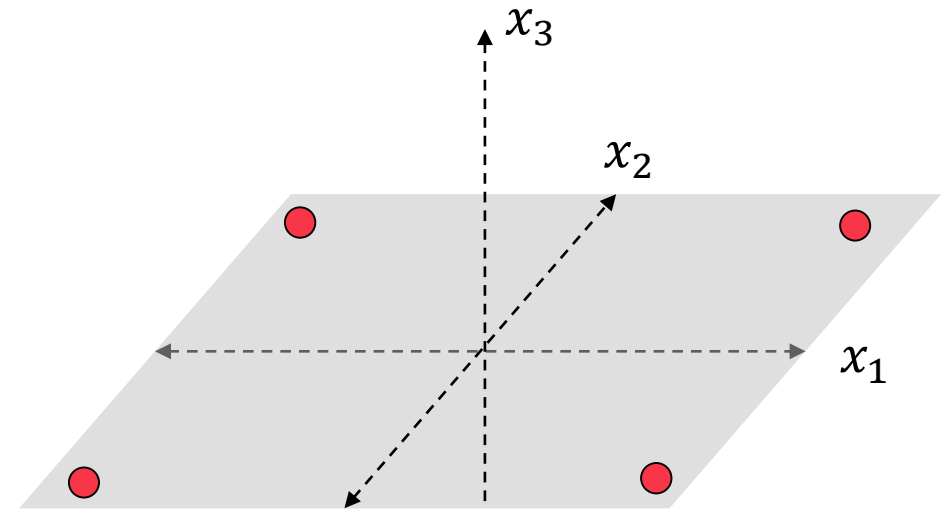
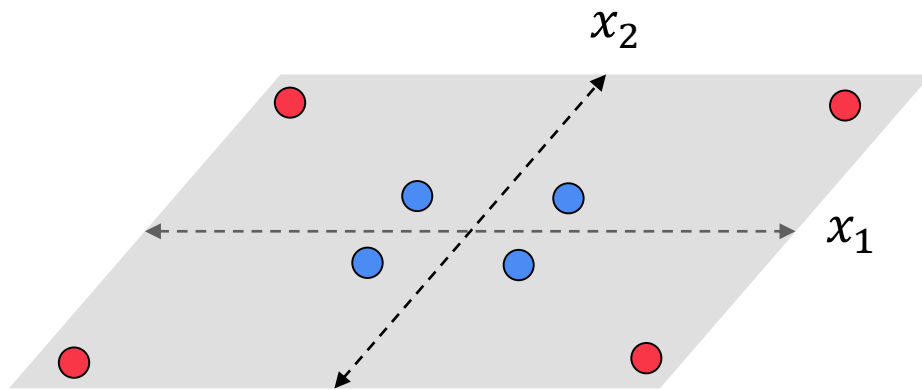
Nonlinear classifier



High dimensional space

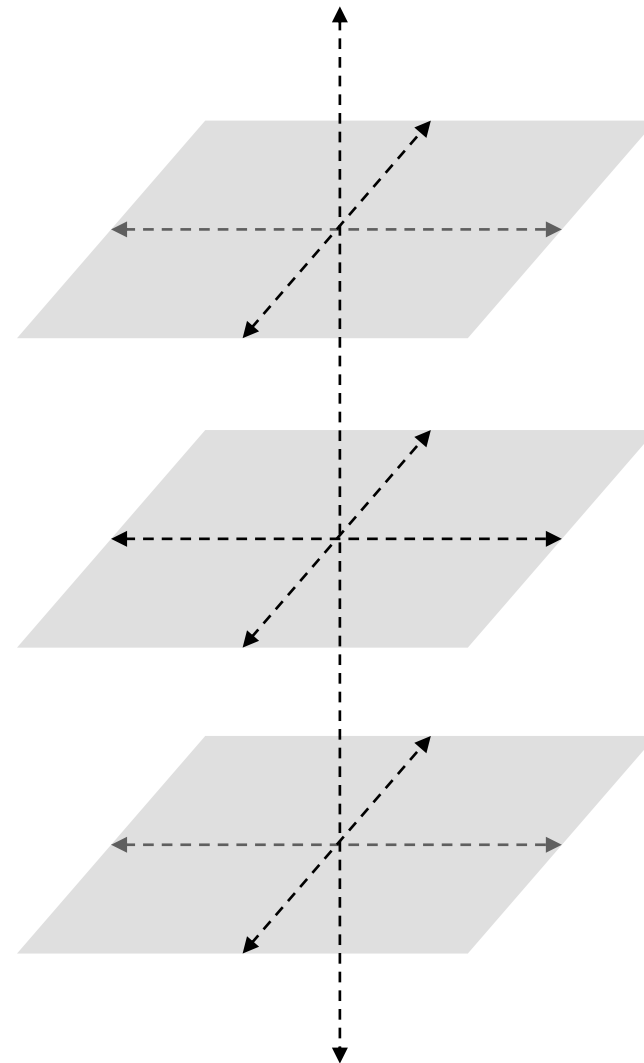
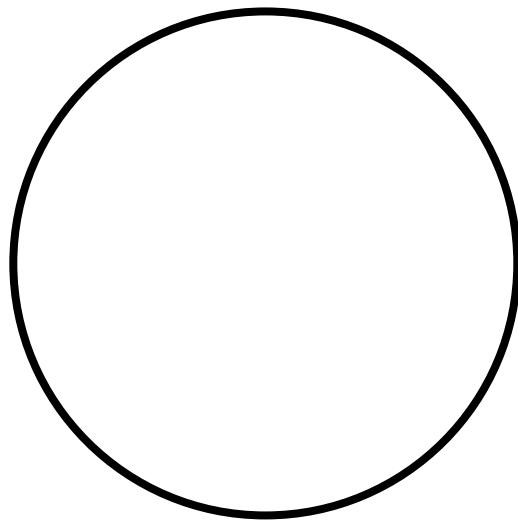


High dimensional space

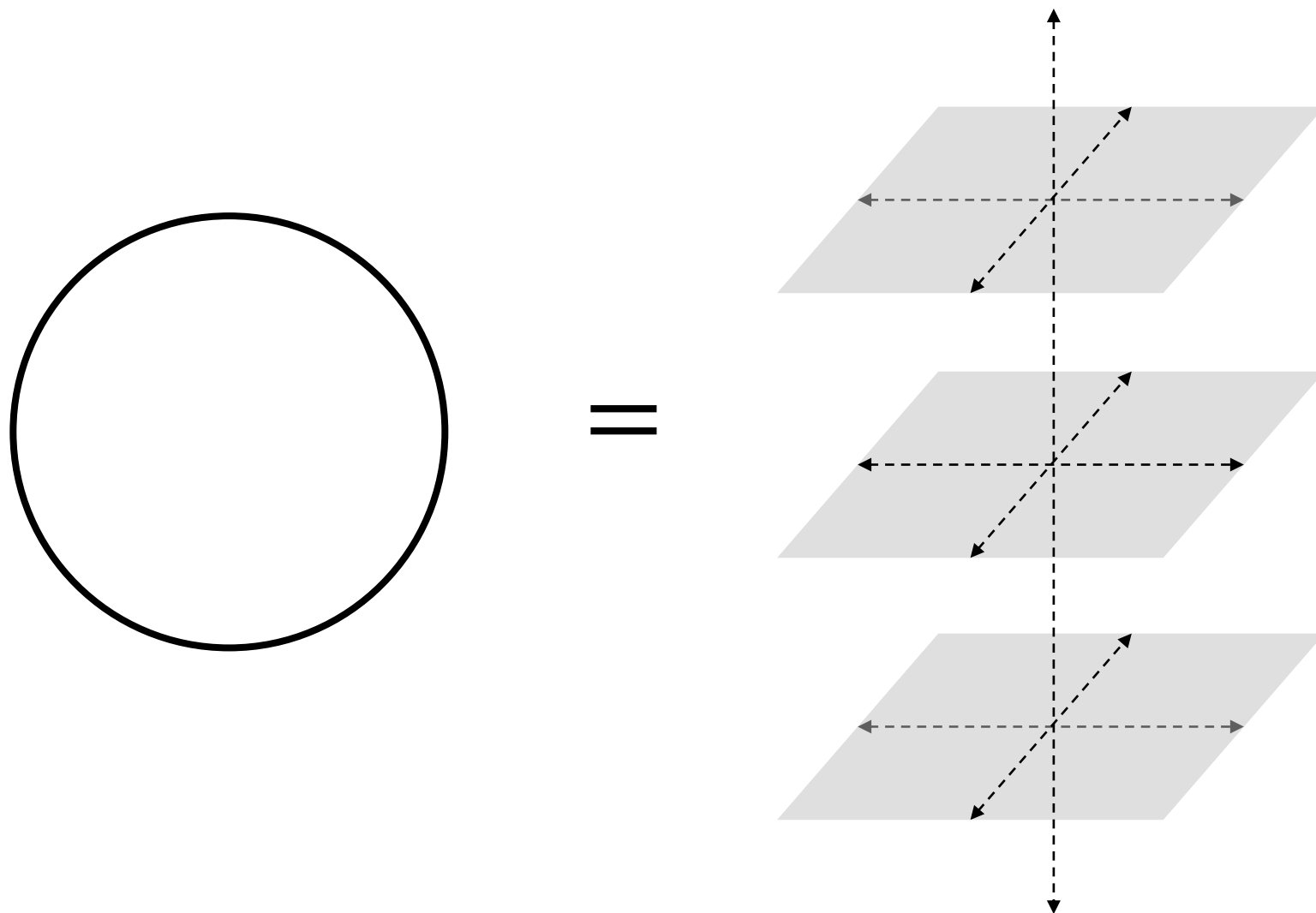


Nonlinear classifier vs. High dimensional space

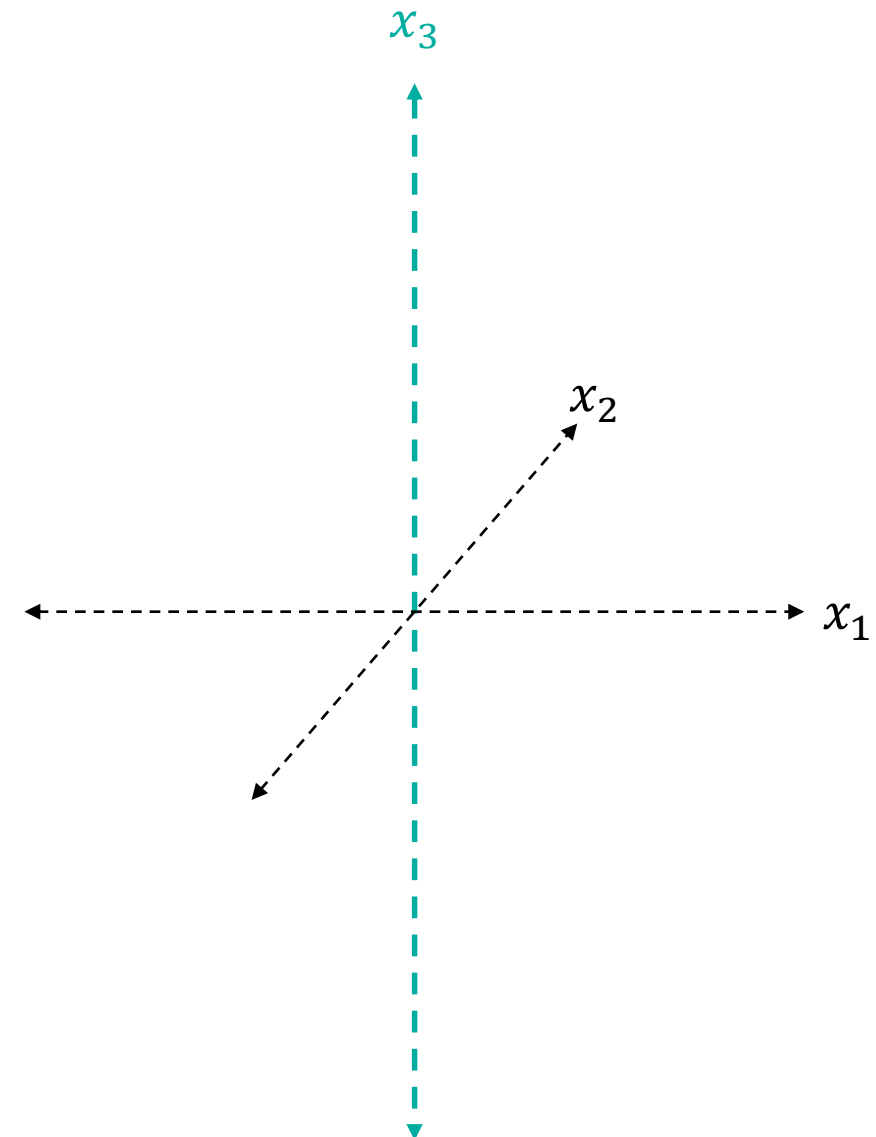
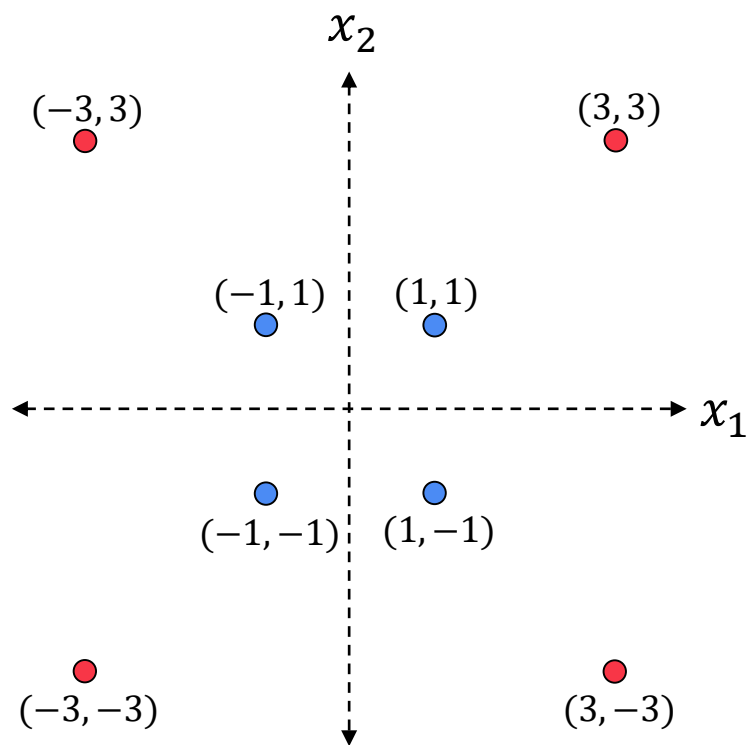
Which one do you more prefer?



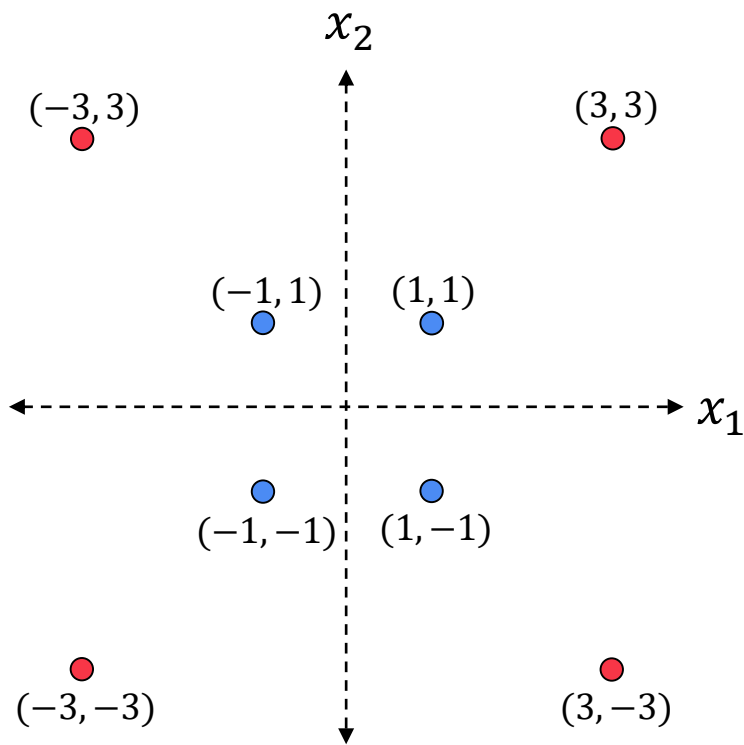
Nonlinear classifier vs. High dimensional space



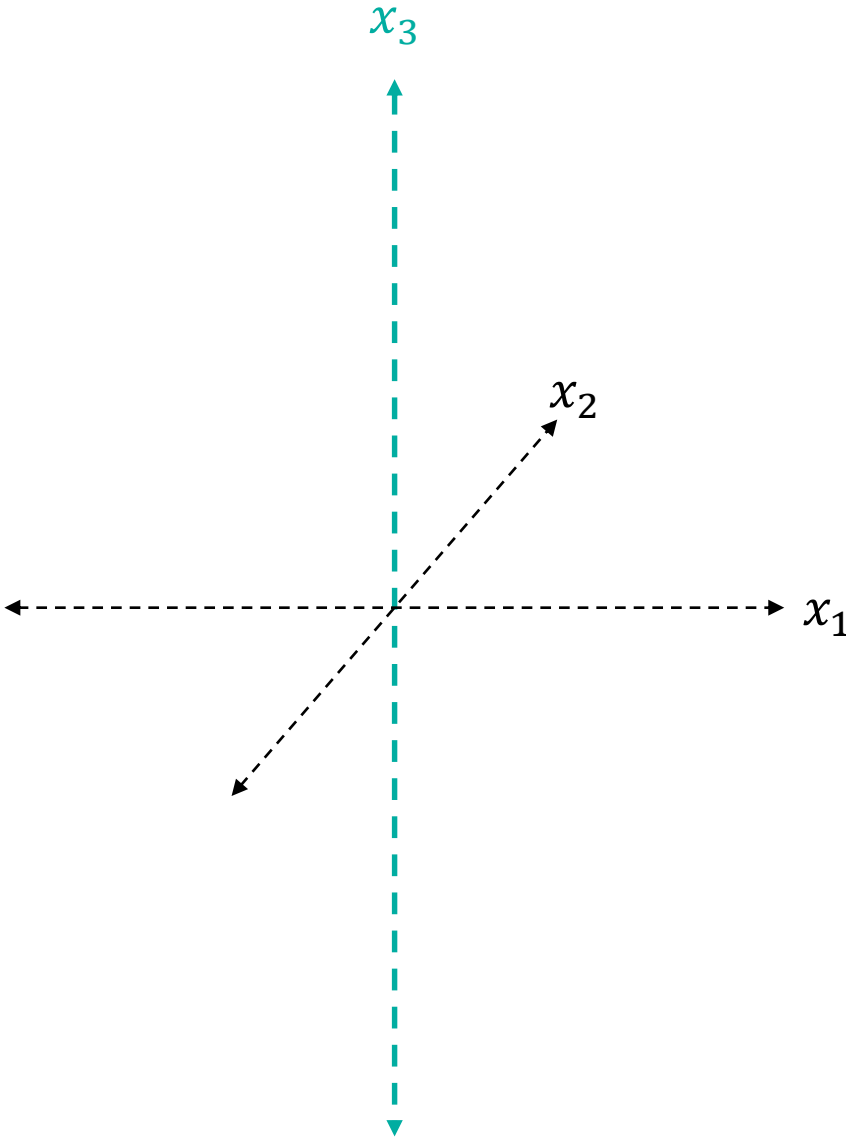
High dimensional space



High dimensional space



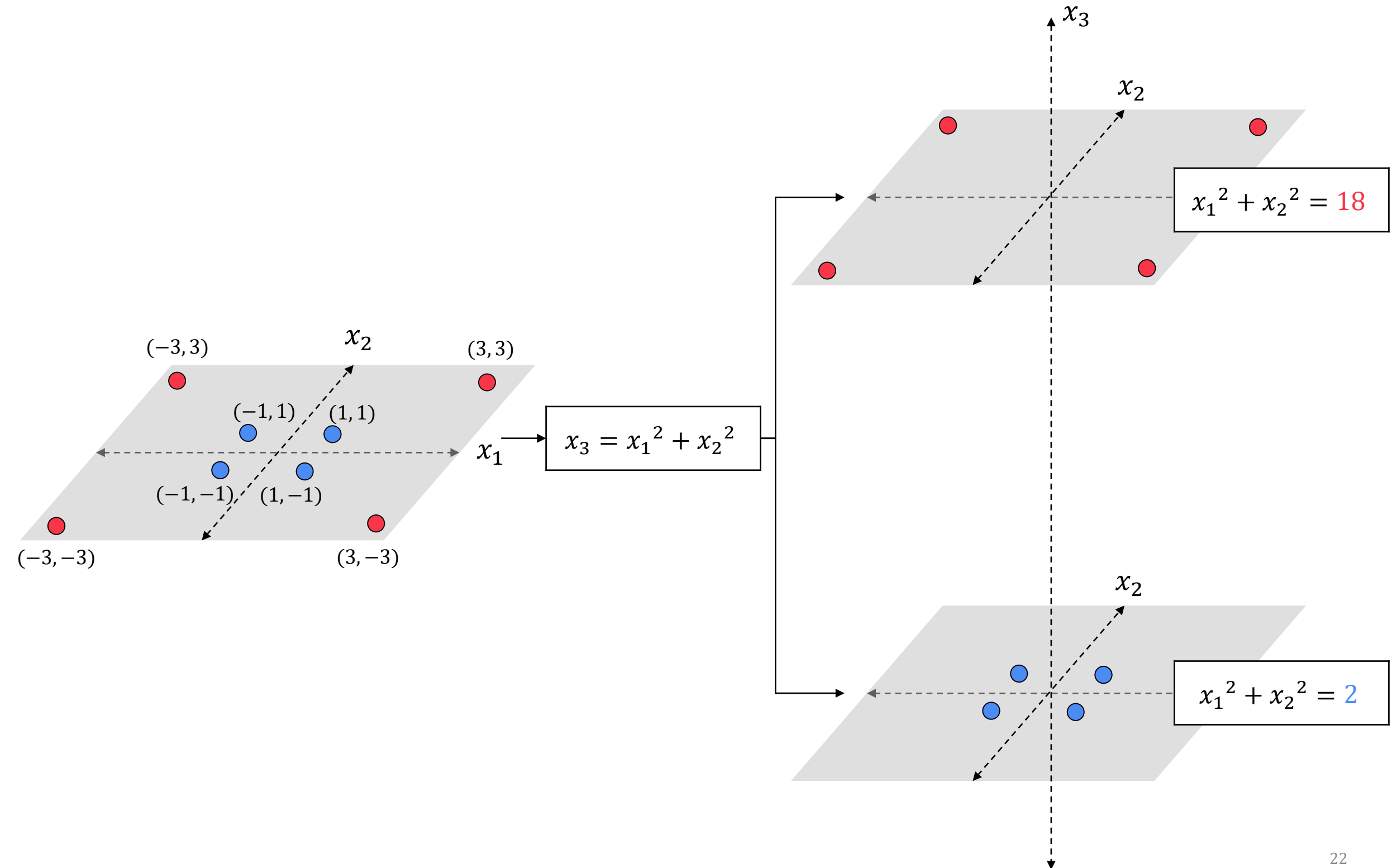
Equation for the new dimension x_3
$x_1 + x_2$
$x_1 x_2$
$x_1^2 + x_2^2$



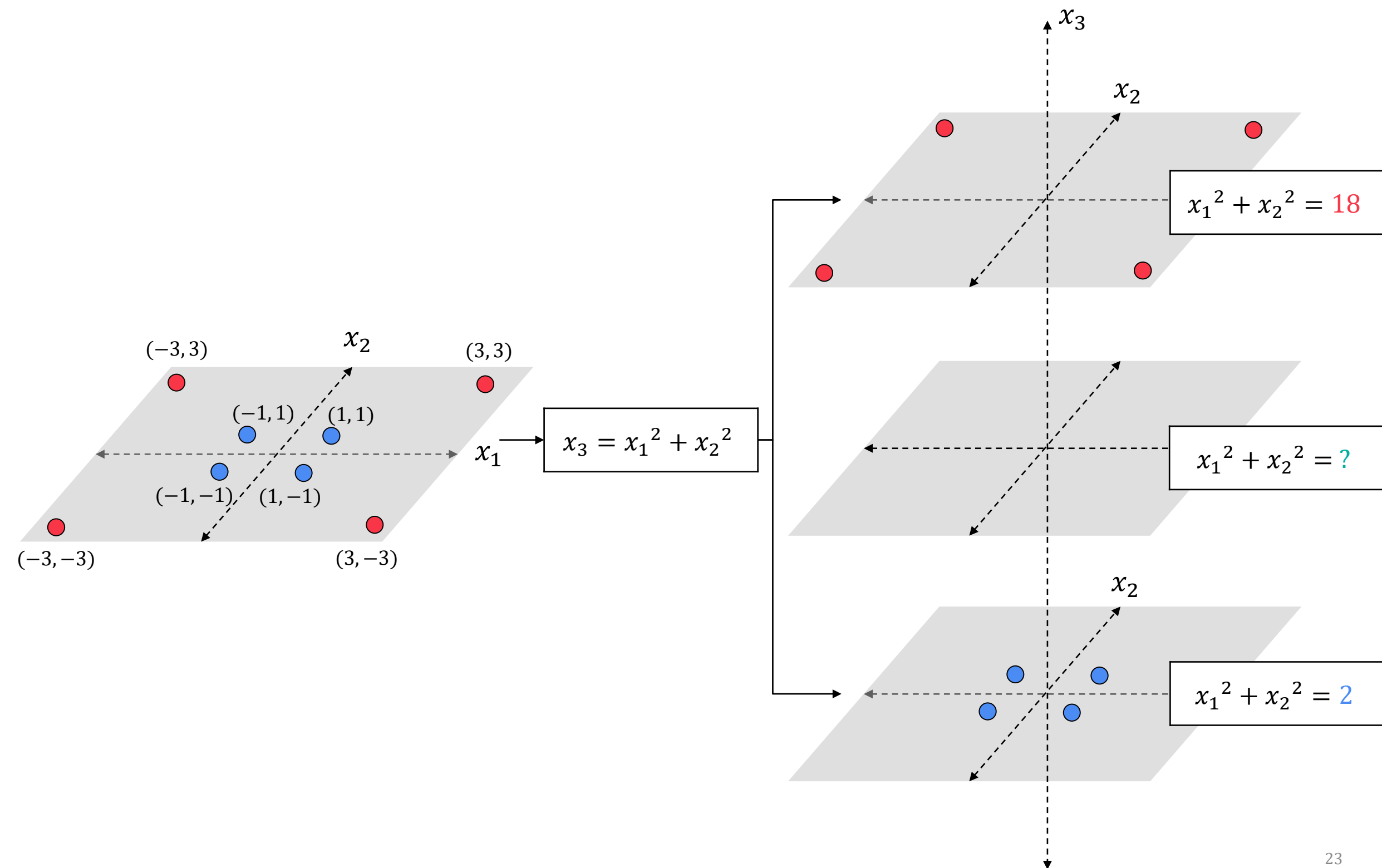
High dimensional space

	$(-3, -3)$	$(3, -3)$	$(-3, 3)$	$(3, 3)$	$(-1, -1)$	$(1, -1)$	$(-1, 1)$	$(1, 1)$
$x_3 = x_1 + x_2$	-6	0	0	6	-2	0	0	2
$x_3 = x_1 x_2$	9	-9	-9	9	1	-1	-1	1
$x_3 = x_1^2 + x_2^2$	18	18	18	18	2	2	2	2

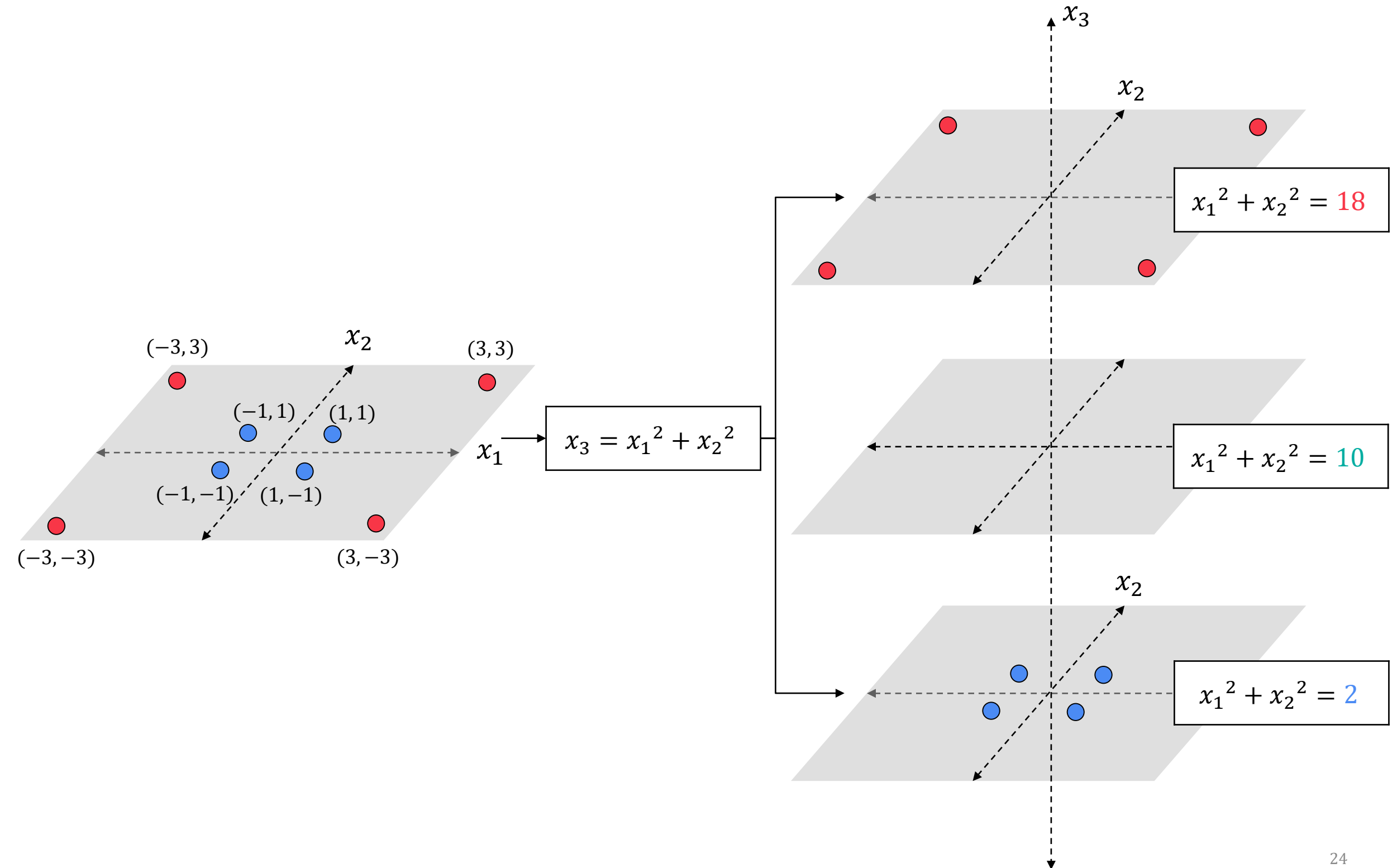
High dimensional space



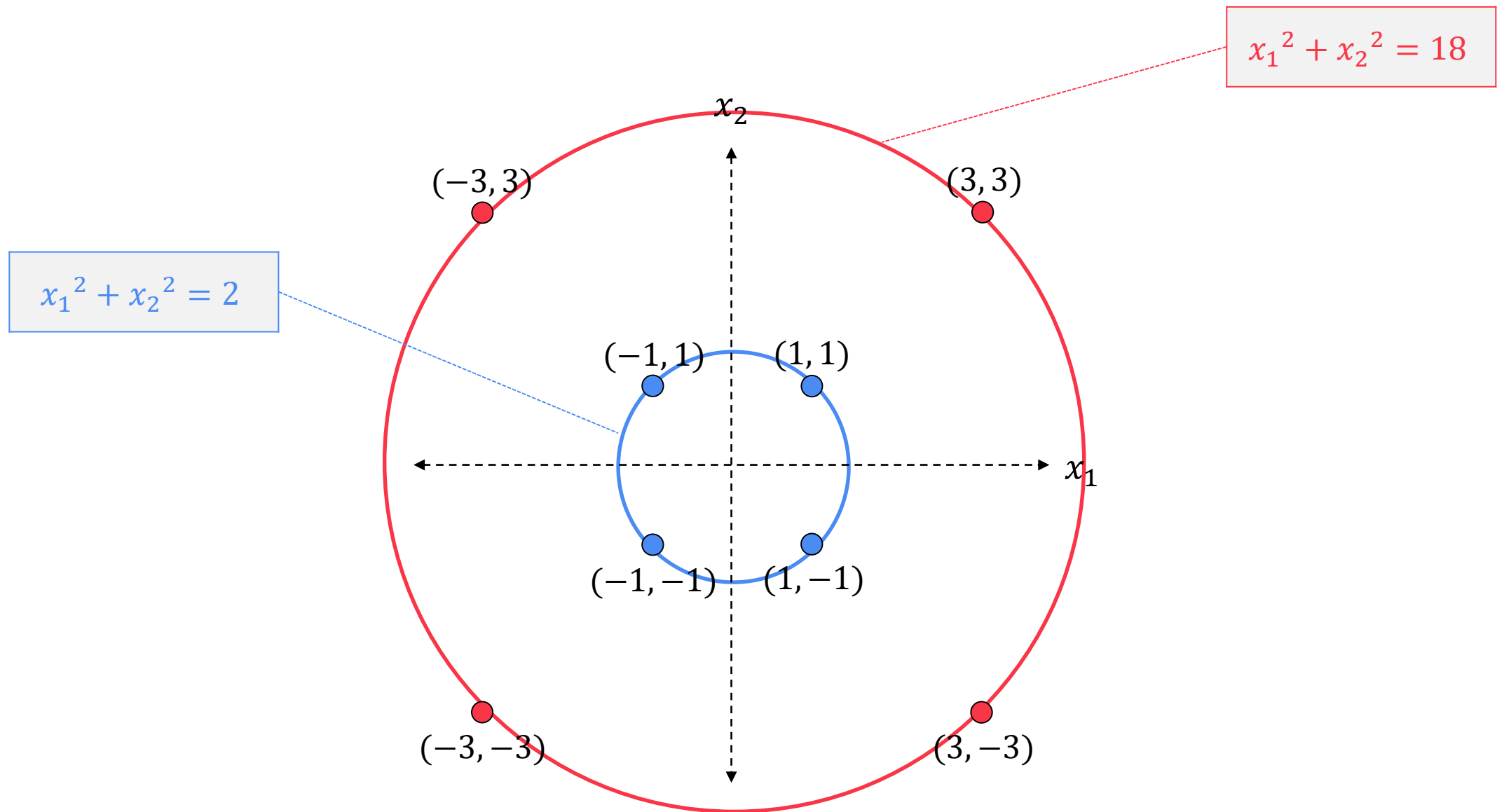
High dimensional space



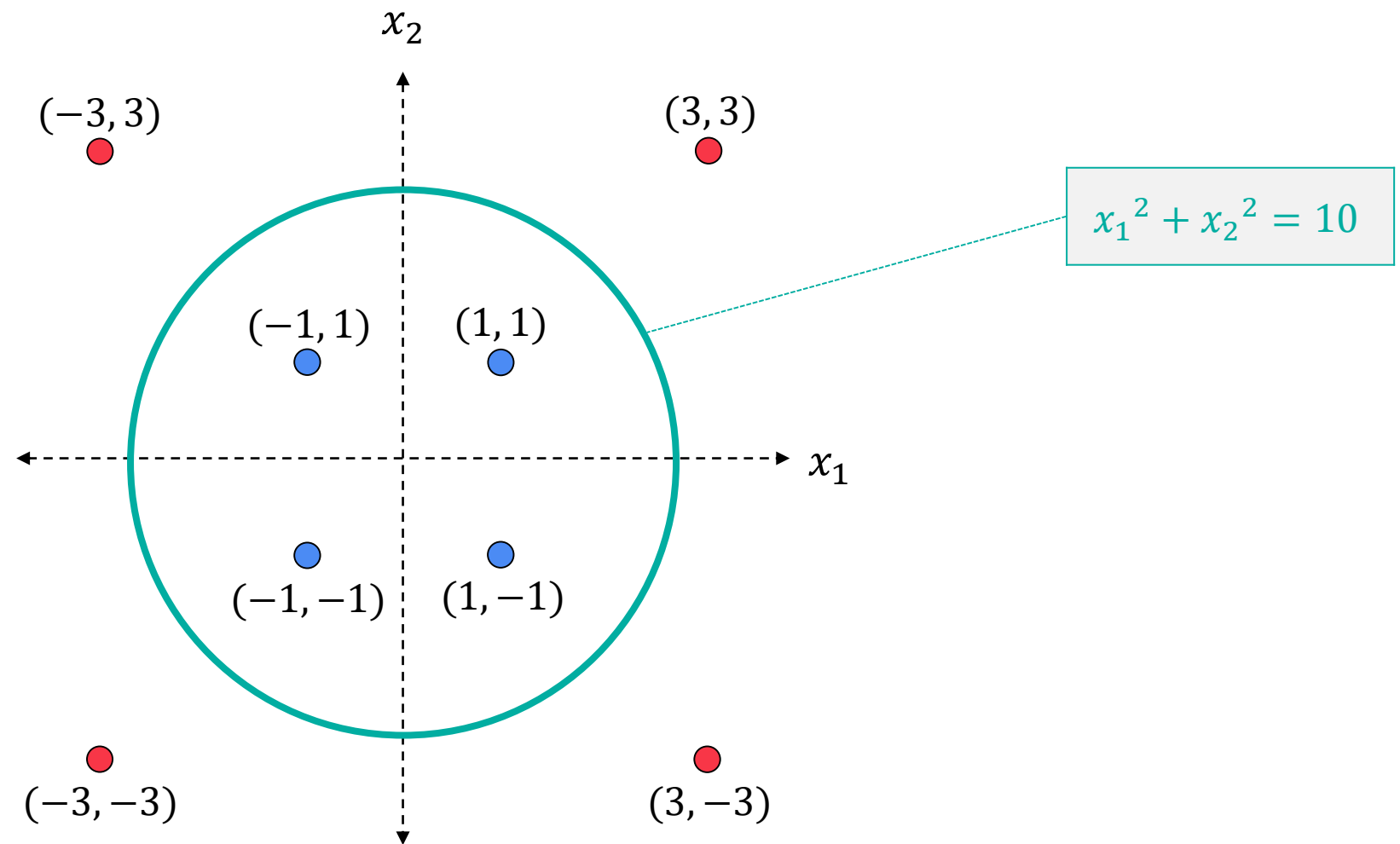
High dimensional space



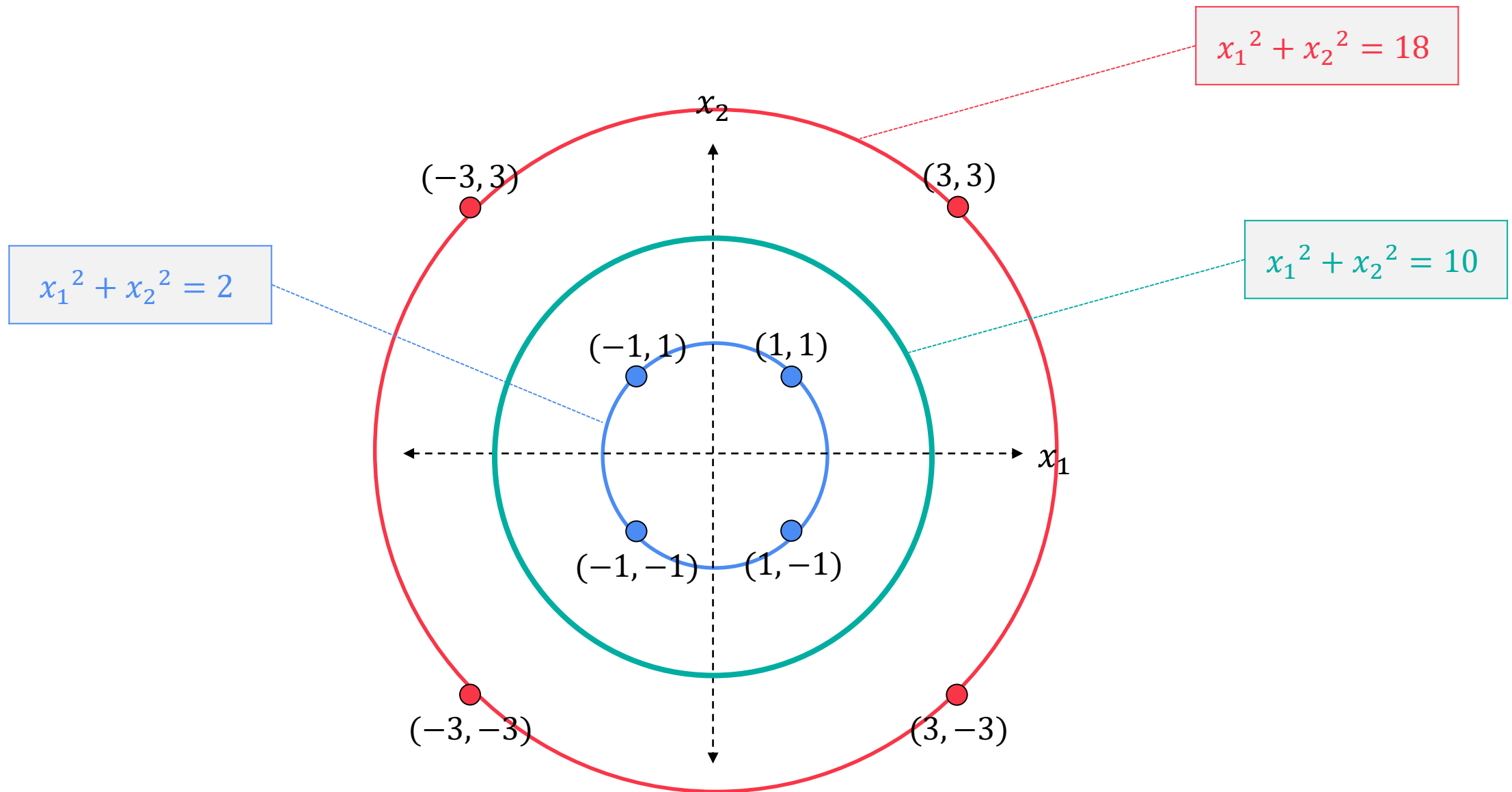
Nonlinear classifier



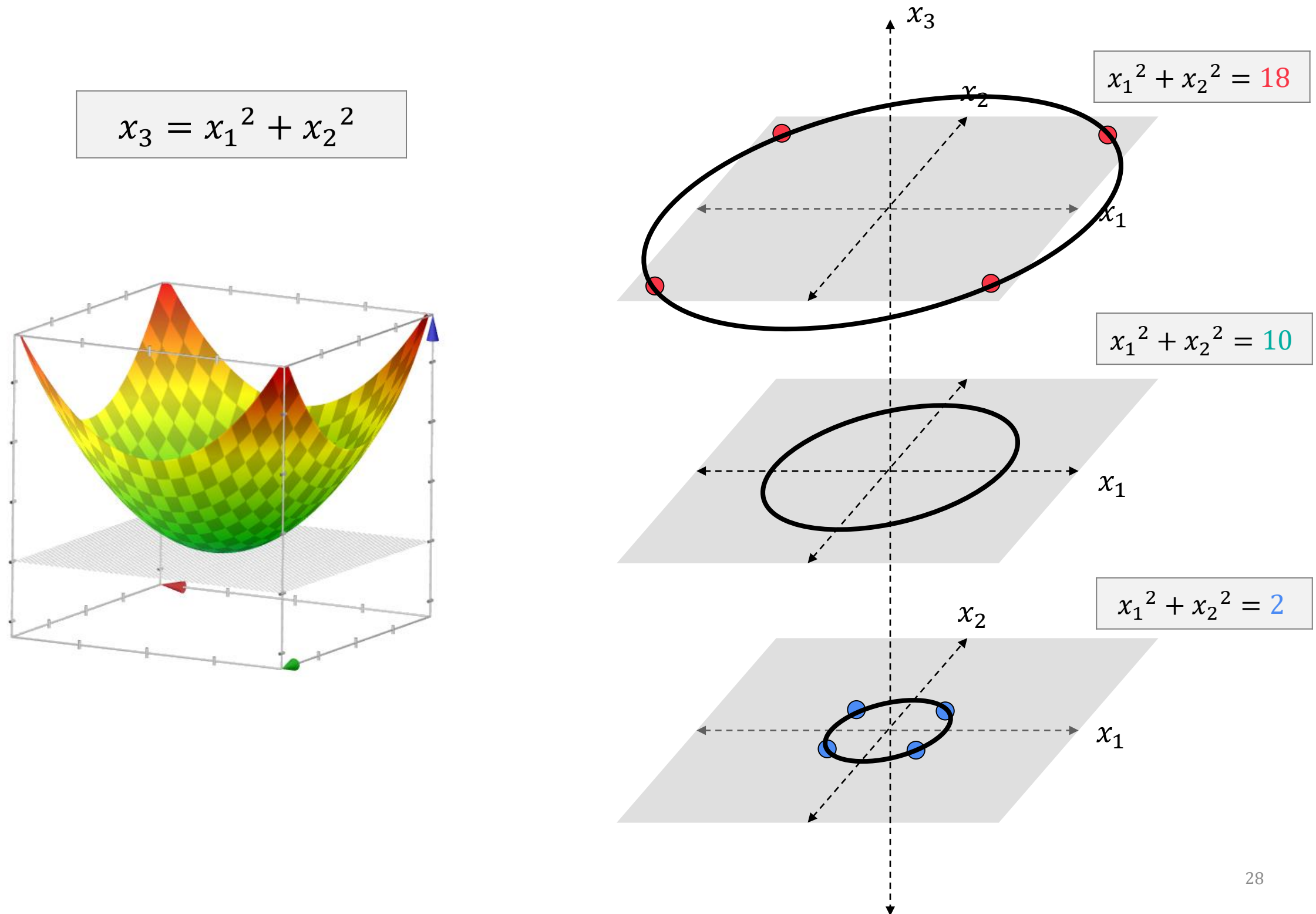
Nonlinear classifier



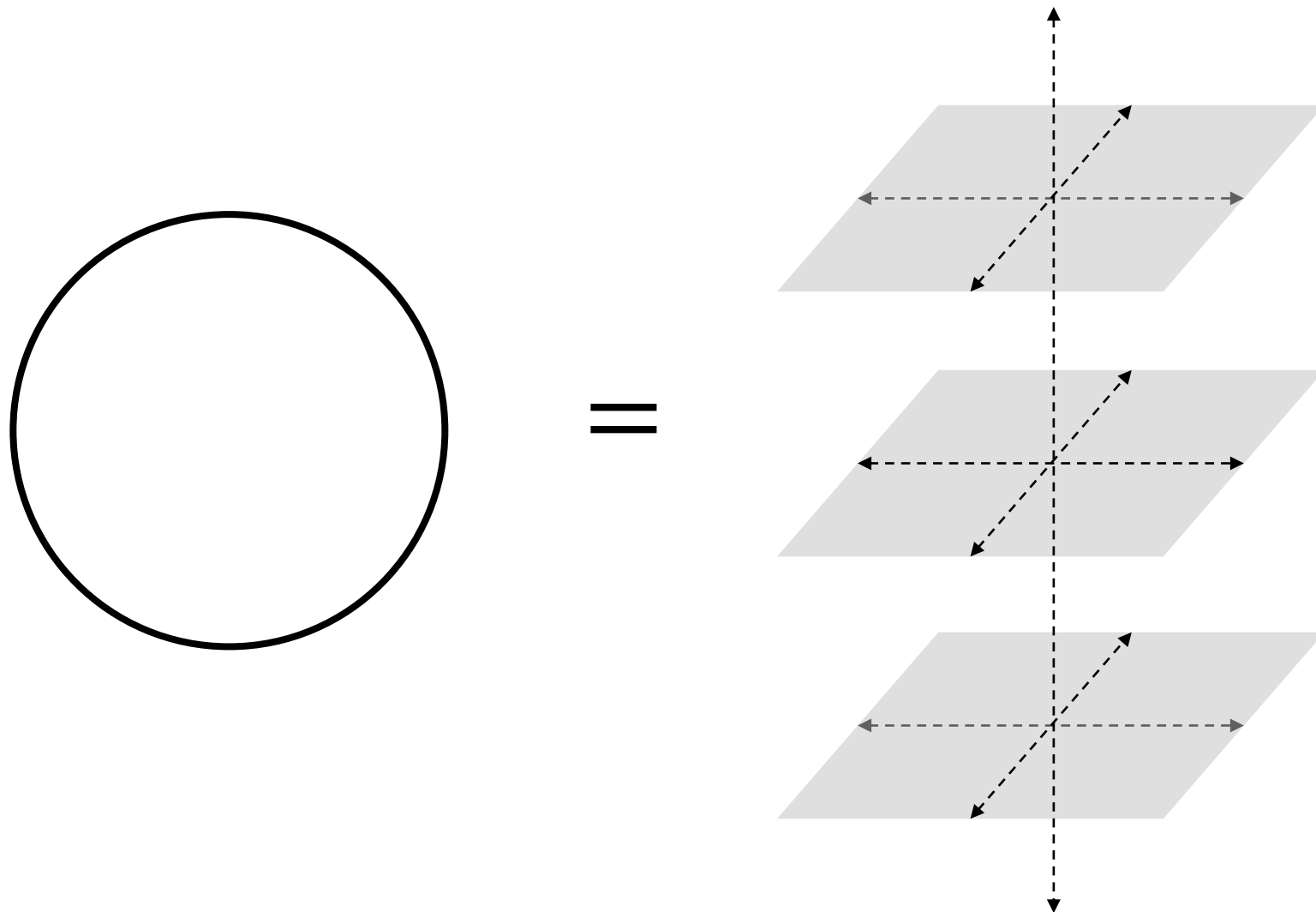
Nonlinear classifier



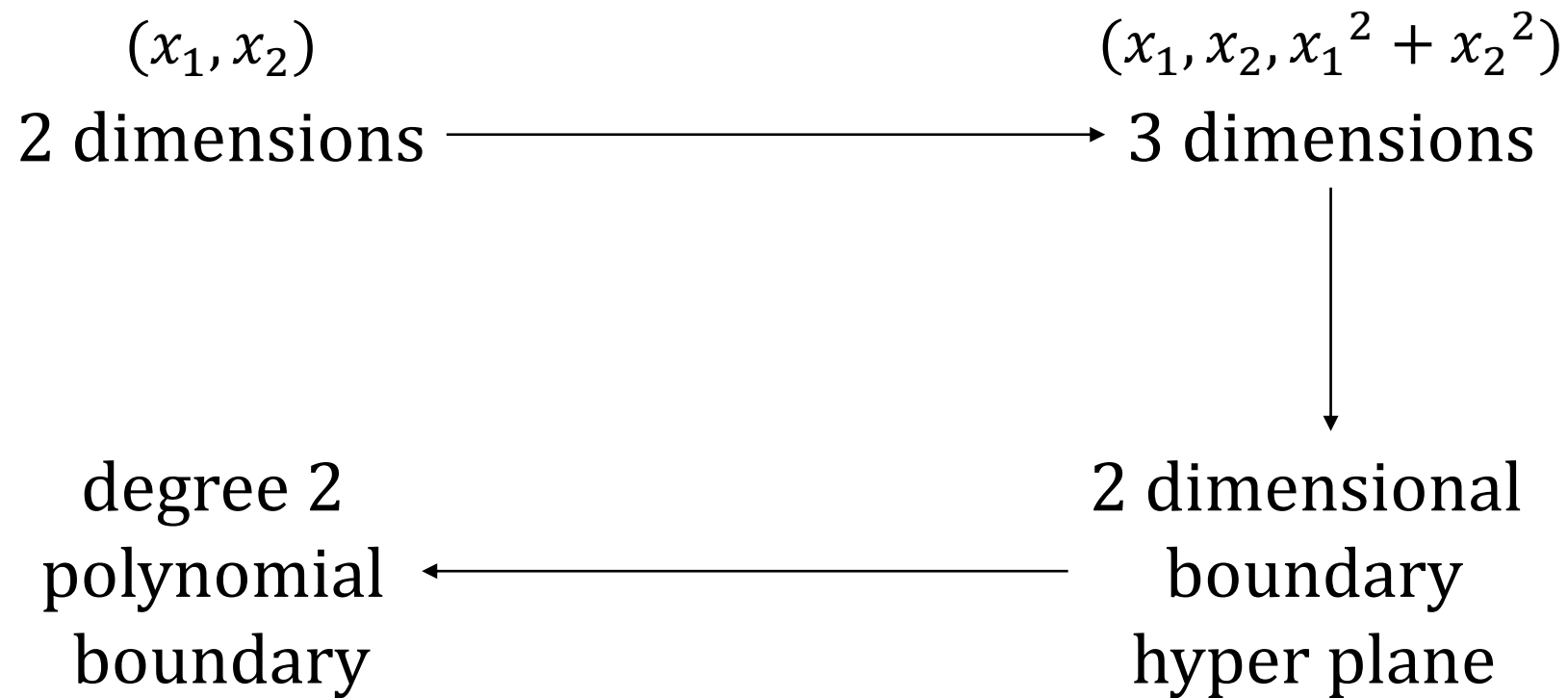
Nonlinear classifier vs. High dimensional space



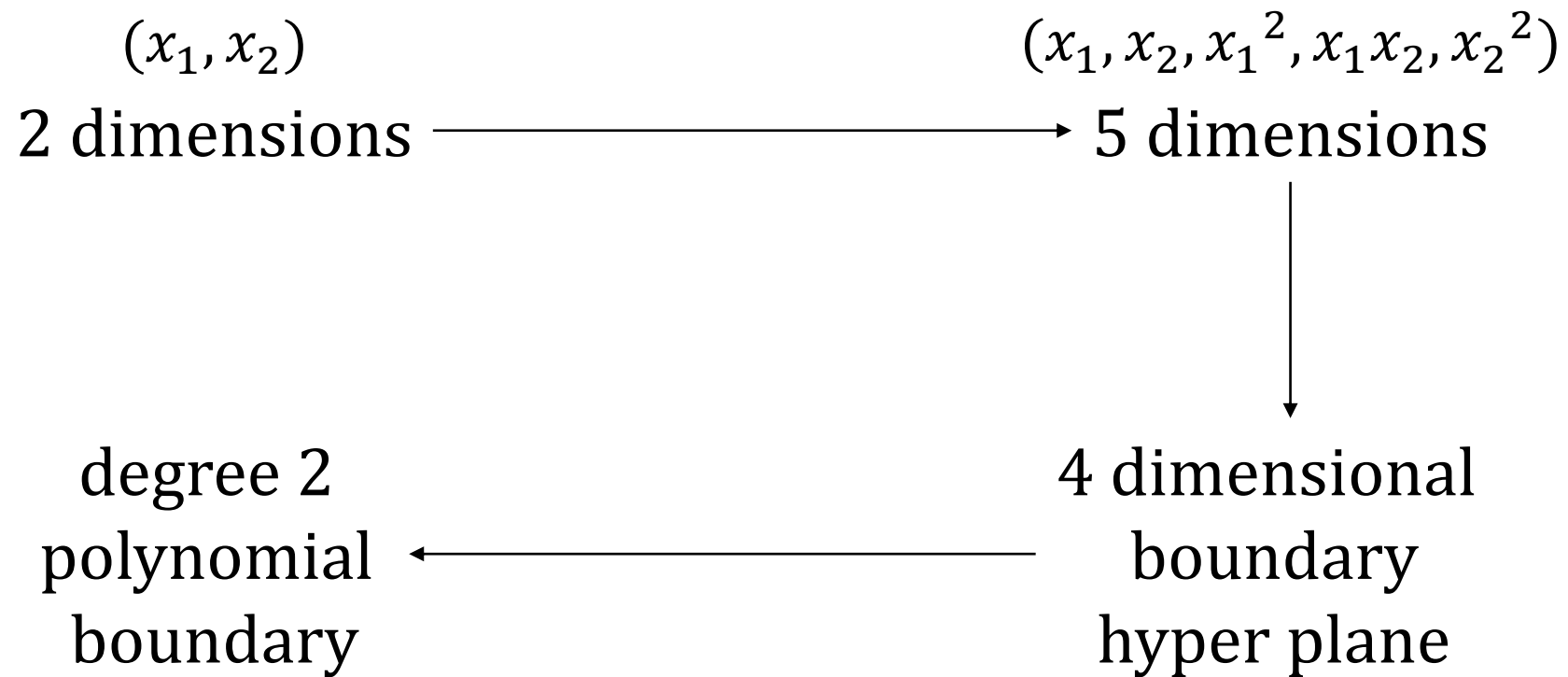
Nonlinear classifier vs. High dimensional space



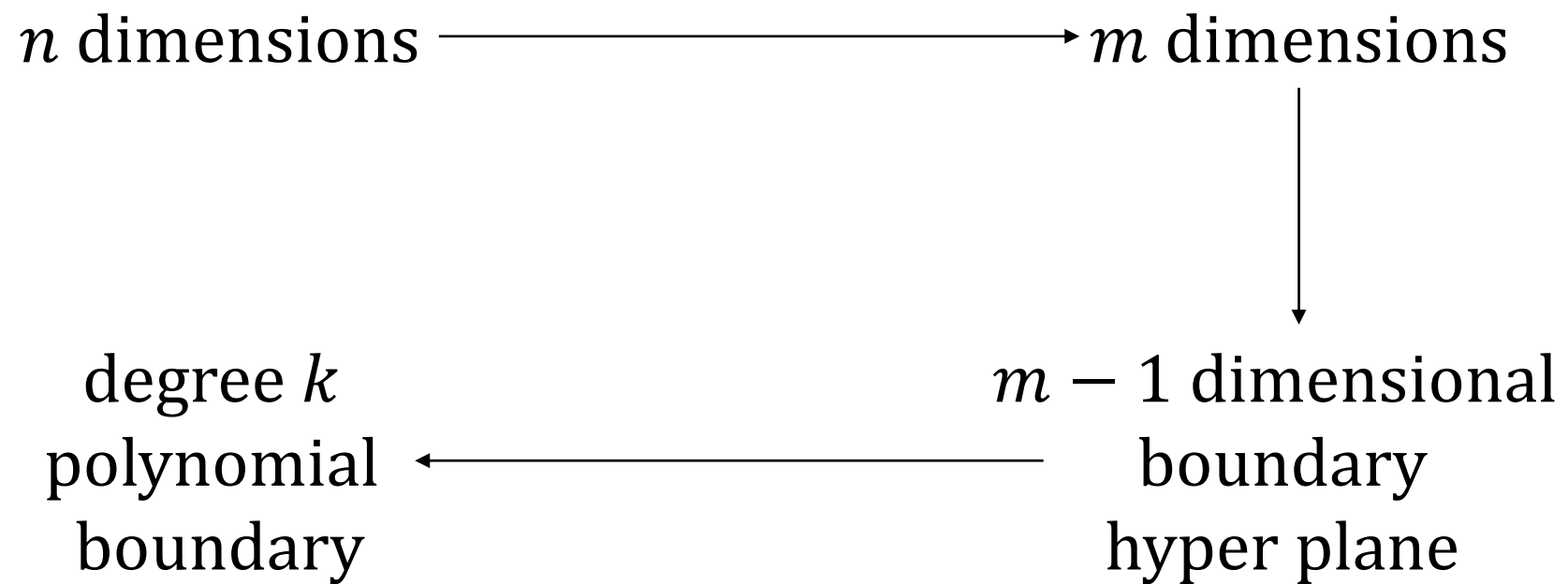
Polynomial kernel



Polynomial kernel



Polynomial kernel



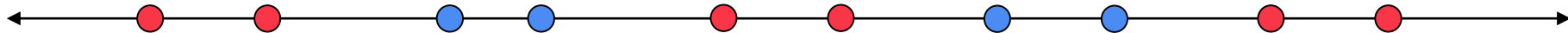
Polynomial kernel

Original dimension: x_1, x_2

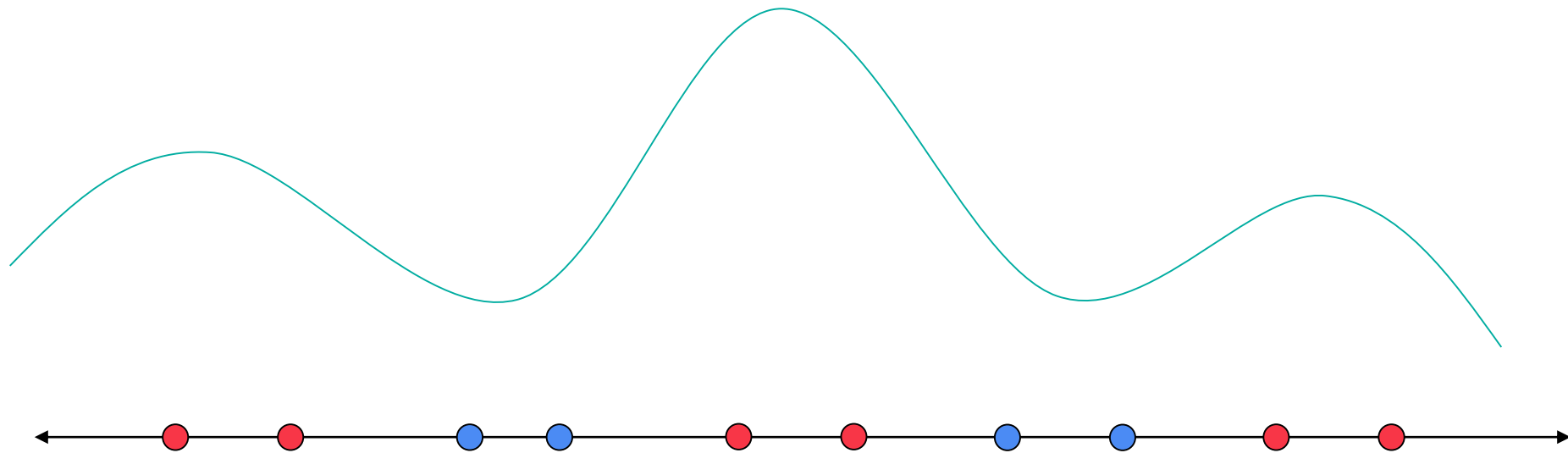
Degree	1	2	3	...
Tools	x_1, x_2	$x_1, x_2, x_1^2, x_1x_2, x_2^2$	$x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3$...
E.g.	$2x_1 - x_2 = 1,$ $5x_1 + 4x_2 = 3$	$x_1x_2 = 1,$ $3x_1^2 - x_2^2 = 7$	$x_1^3 + 2x_1^2 - x_1 - x_2 = 2$...

Concept of radial basis function (RBF)

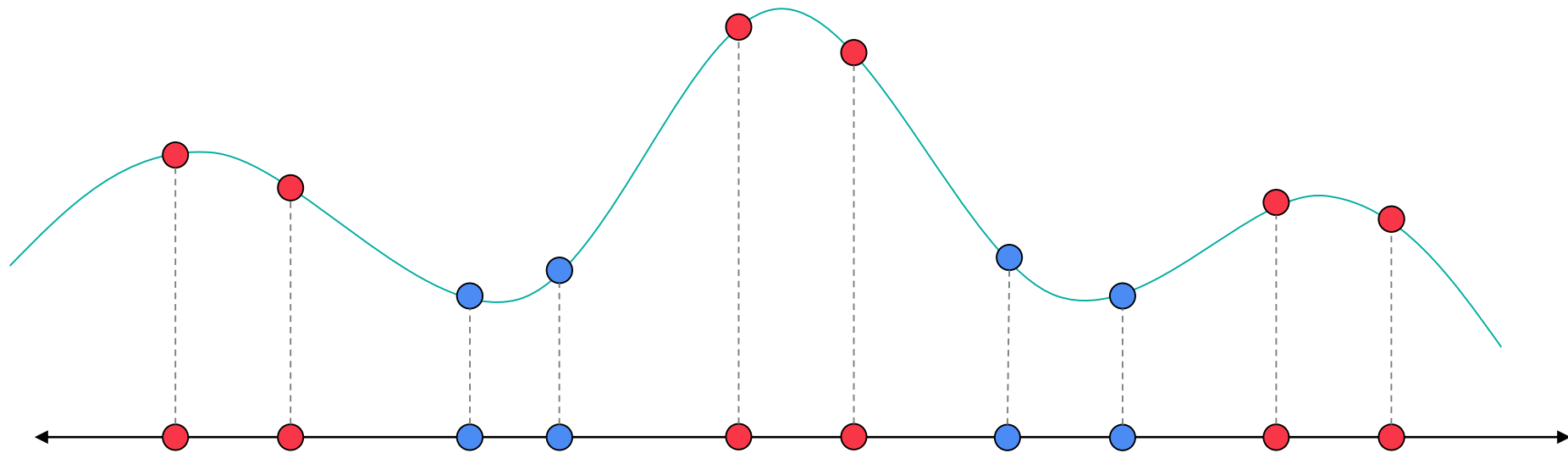
Again, nonlinearly separable data



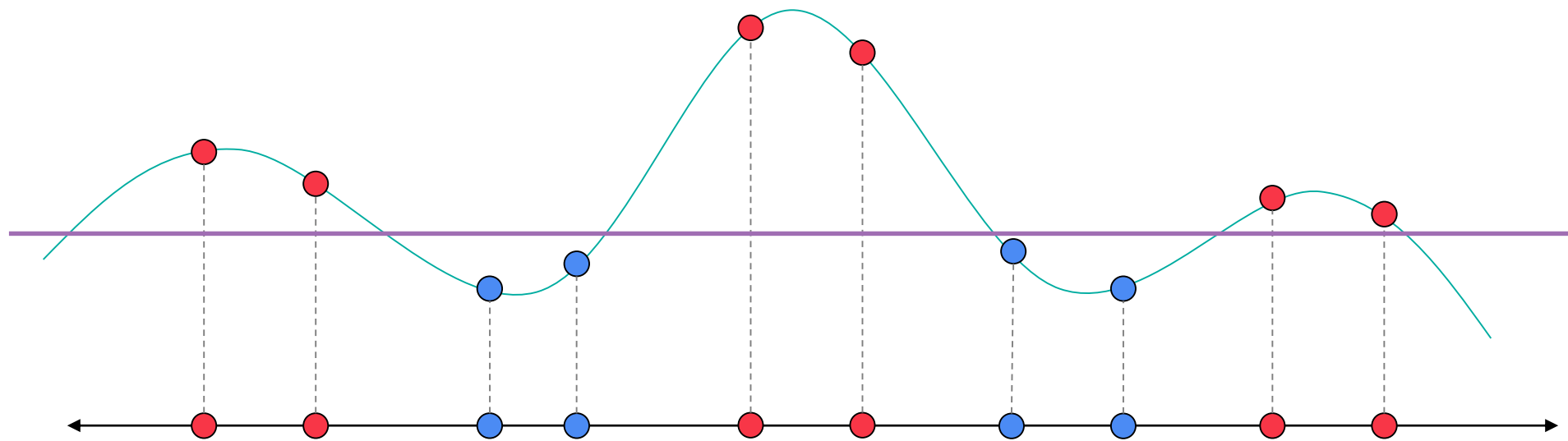
Again, nonlinearly separable data



Again, nonlinearly separable data



Again, nonlinearly separable data

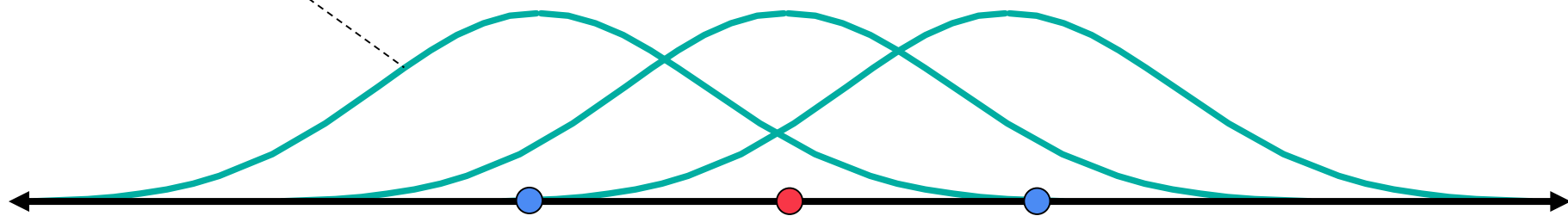


RBF kernel

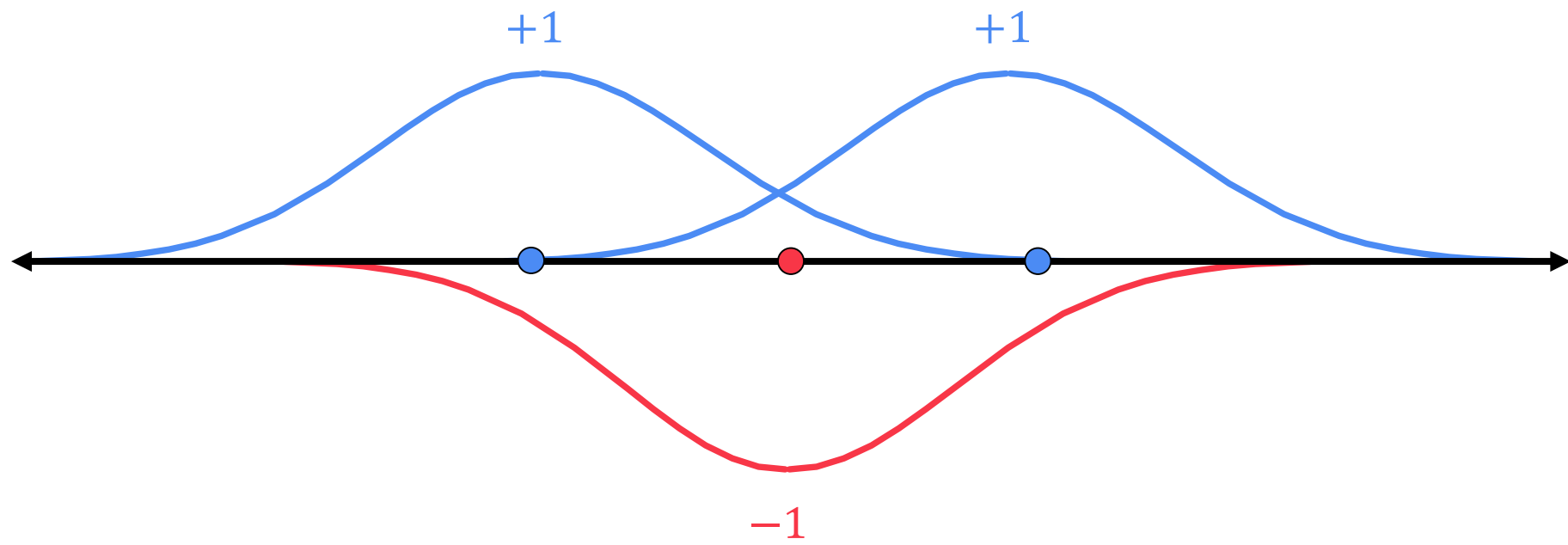


RBF kernel

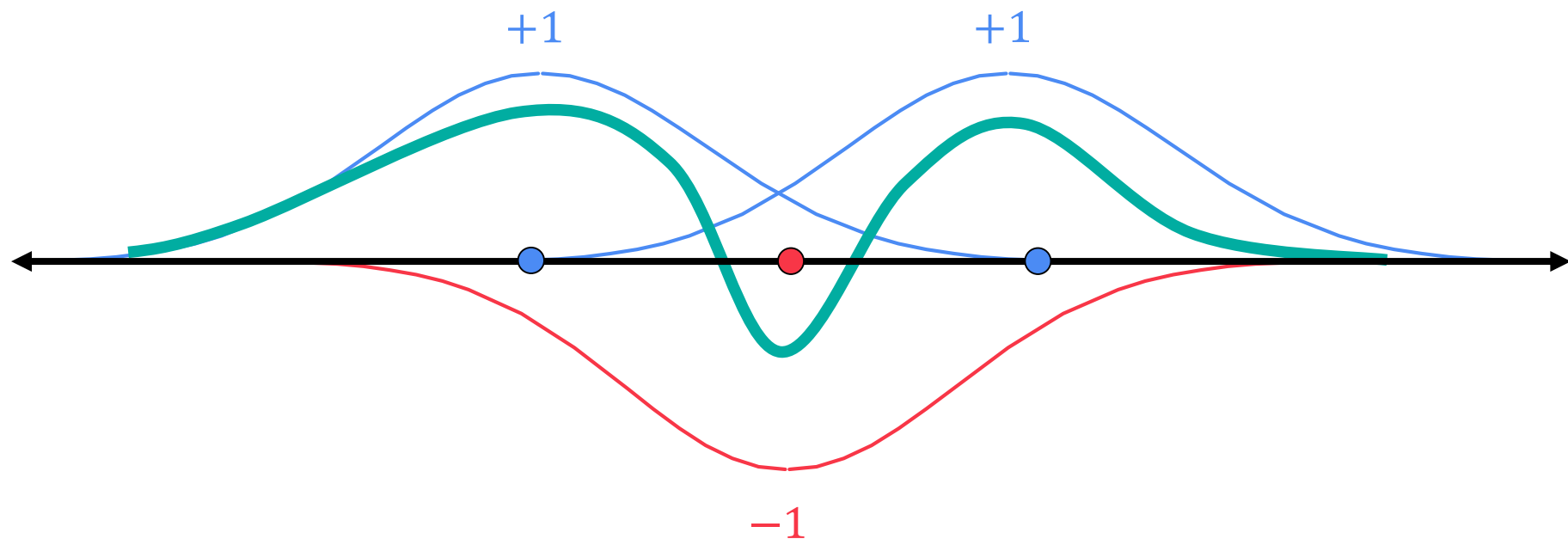
Radial basis function



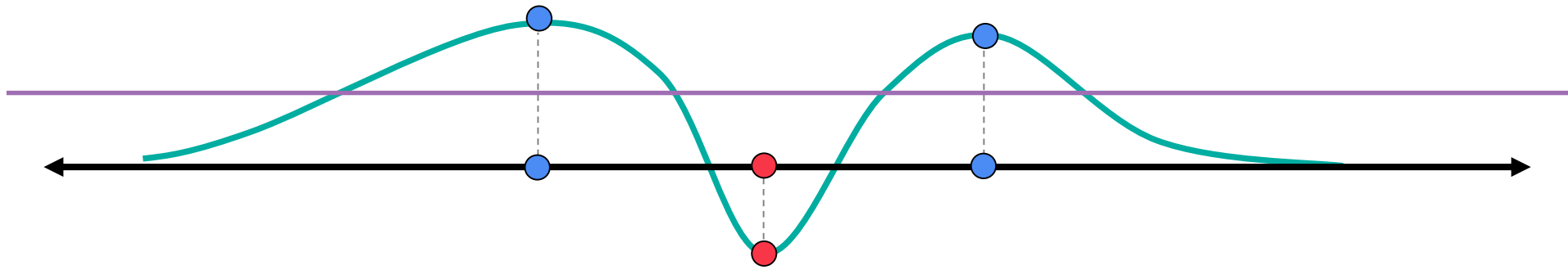
RBF kernel



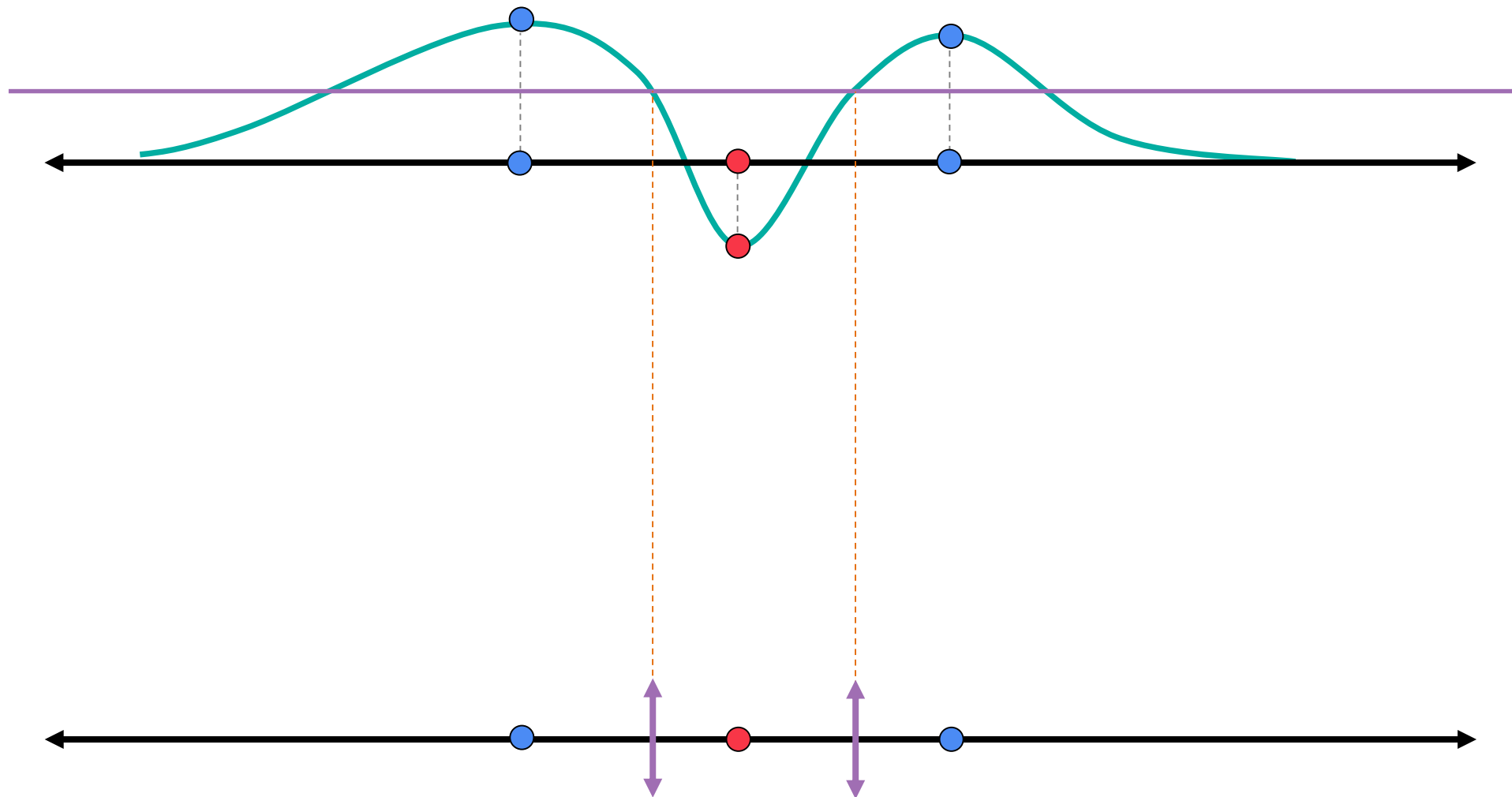
RBF kernel



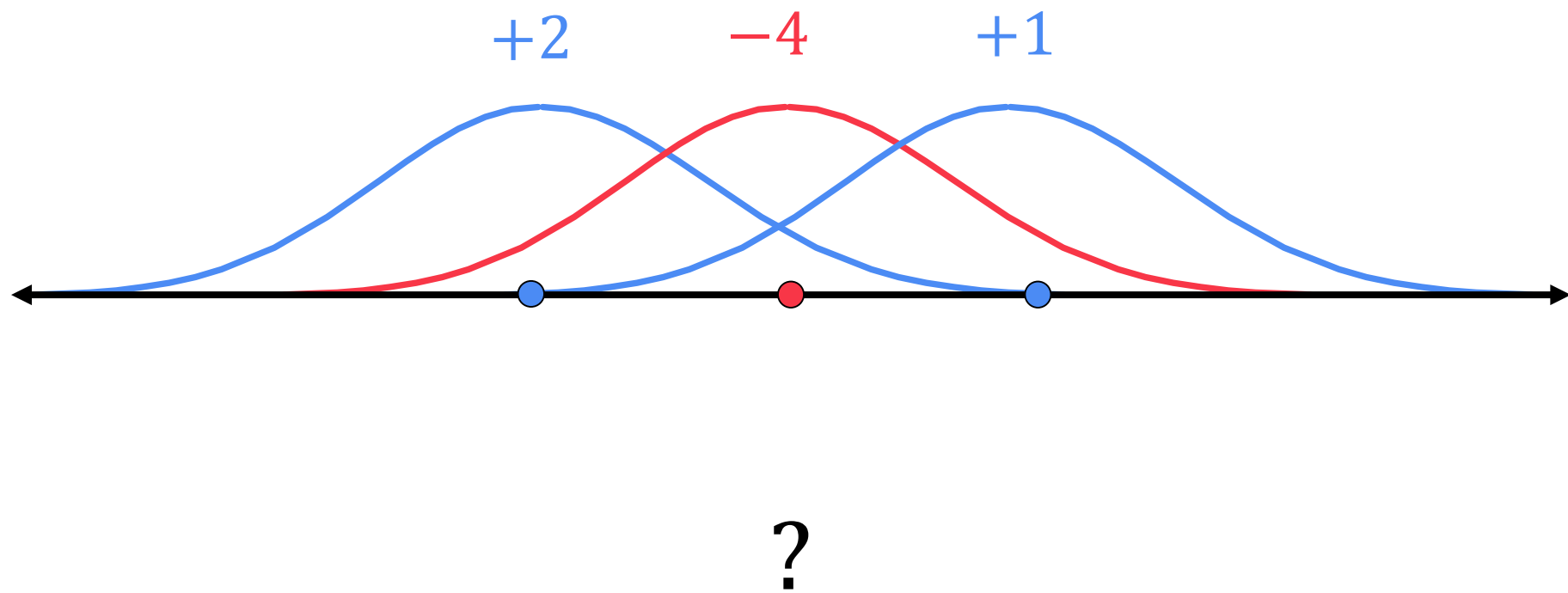
RBF kernel



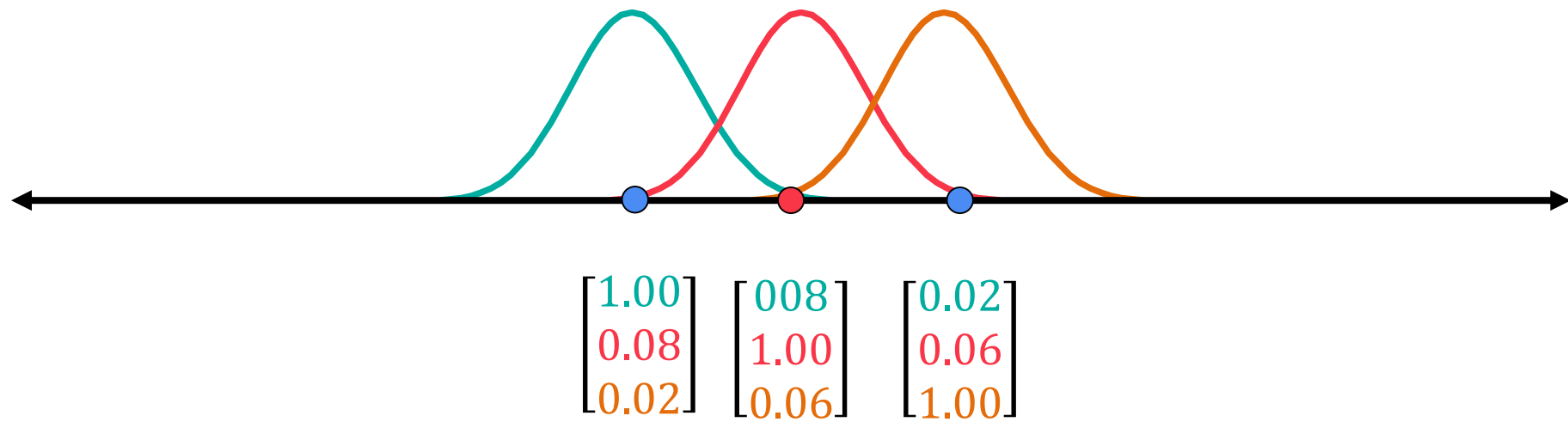
RBF kernel



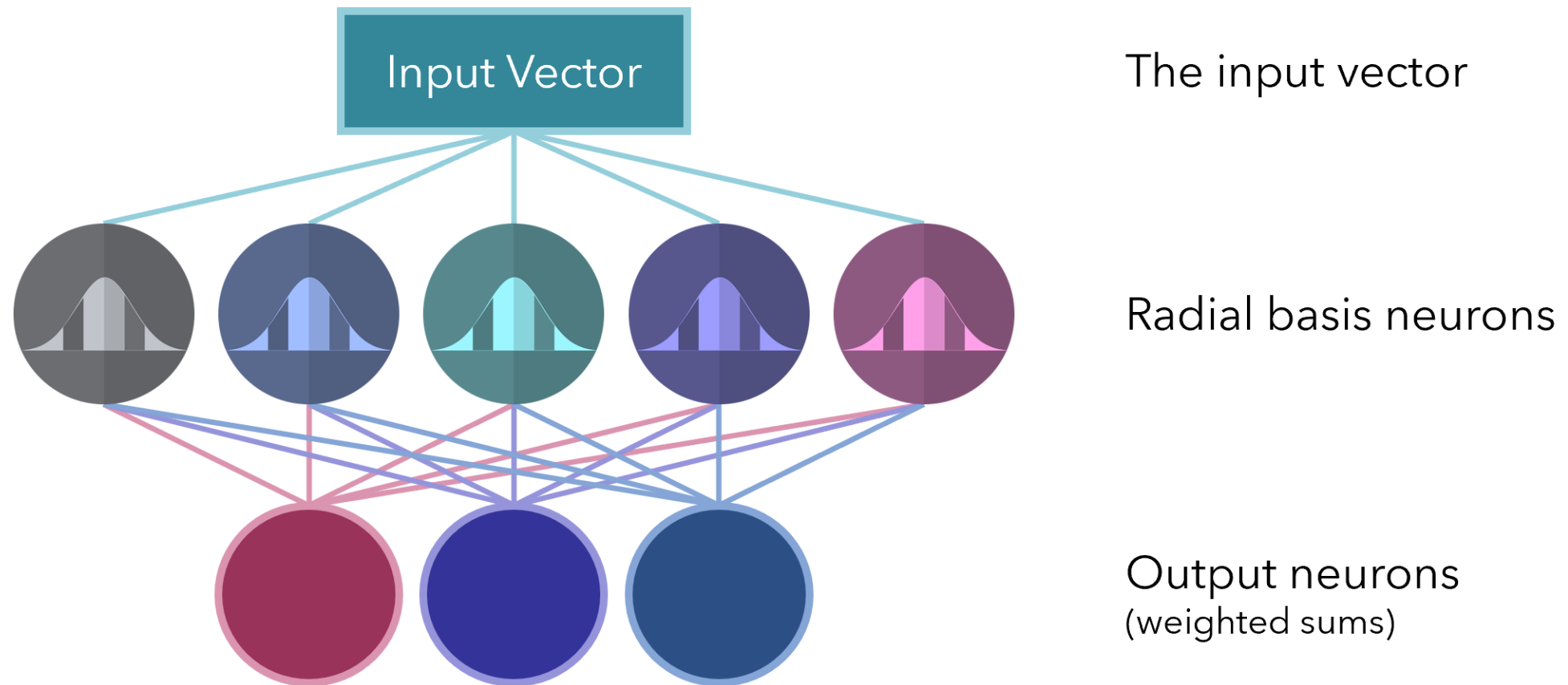
RBF kernel



How to find the coefficient of radial basis



How to find the coefficient of radial basis



중간고사 공고

- 2020.11.04(수) & 11.06(금)

순번	팀명	발표 날짜
1	YOLO	11.04
2	AKI	11.04
3	Harmony	11.04
4	안시성	11.04
5	H:J	11.04/11.06
6	Ajsoftware	11.06
7	제니리아	11.06

발표 시간	질의응답 시간	1팀 당 소요 시간	전체 소요 시간
15분	5분	20분	140분

발표 자료(PPT 파일) 제출	
기한	2020.11.03(화) 오후 11:59
제출 방식	e-강의동 > 데이터사이언스응용 > 과제 > 중간고사
Note #1	팀원 전원 공동 발표(담당 부분)
Note #2	계획서 발표(Upgrade) + 진행 상황(Progress) + 장애요소(Obstacle) + 향후 계획(Plan)

Thank you