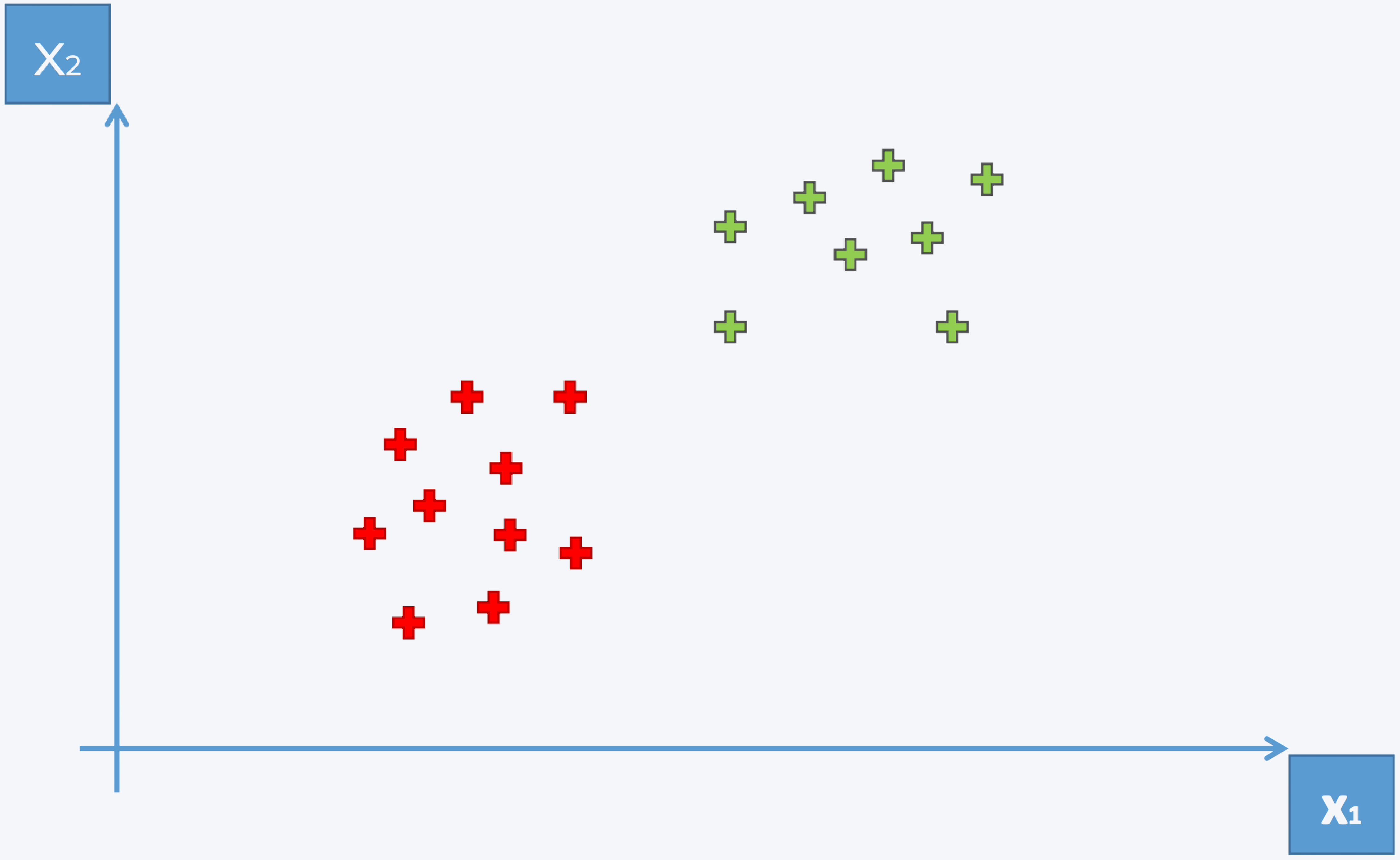
أكاديمية طويق
Tuwaiq Academy

Support Vector Machine (SVM)

Support Vector Machine (SVM)

Again , How to separate these points?



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

How does SVM Work ?

Finding the Best Decision Boundary (Hyperplane)

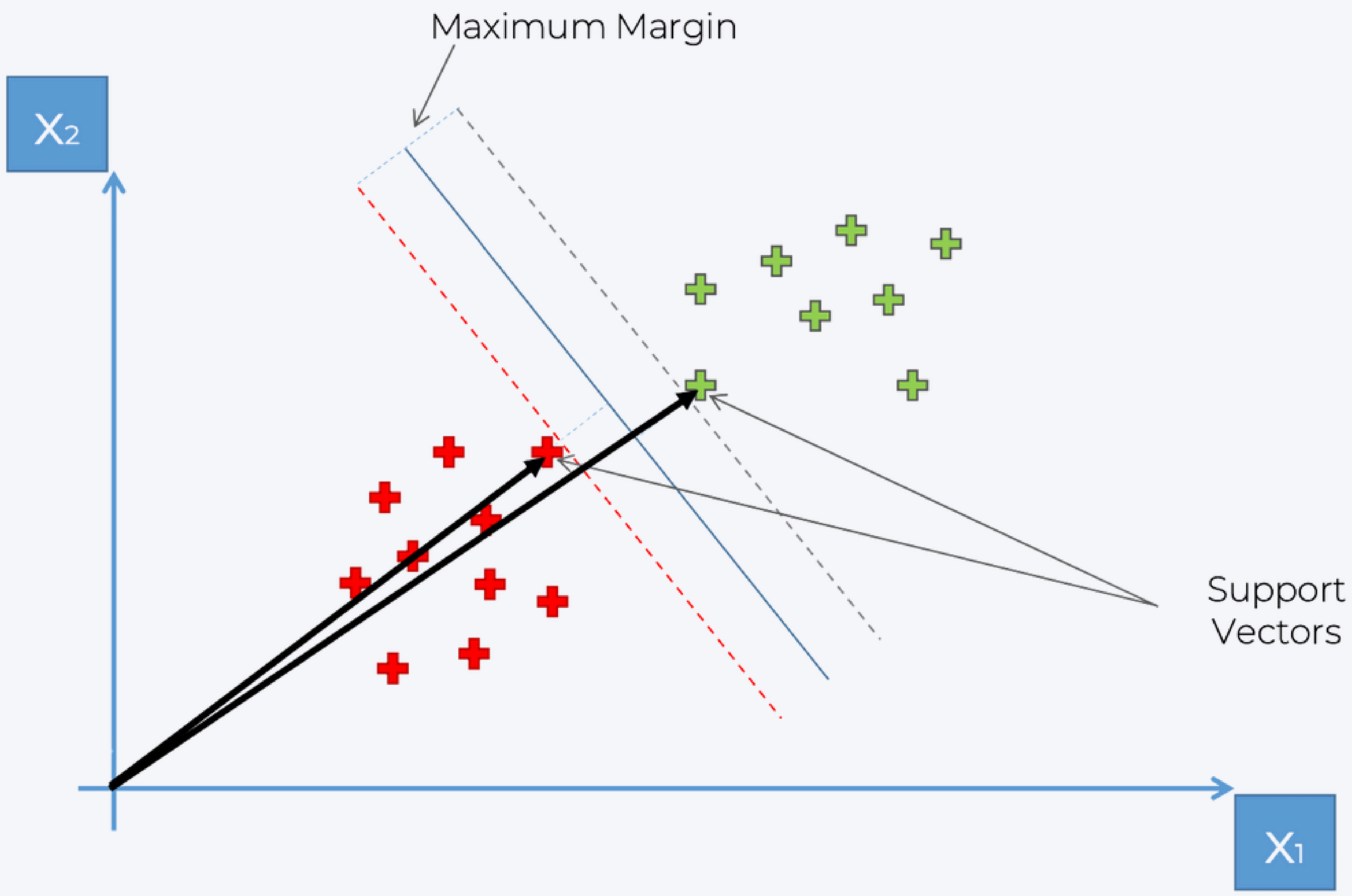
- The black line represents the optimal hyperplane, which separates the two categories (red crosses and green pluses).
- SVM selects this hyperplane to maximize the distance between the closest points from each class.

Support Vectors (Critical Boundary Points)

- The circled points (one from each class) are called support vectors.
- These are the closest points to the decision boundary and directly influence its position.
- Other points do not matter in defining the boundary; only support vectors affect it.

Maximizing the Margin

- The blue lines represent the margin boundaries, showing the maximum distance the decision boundary can maintain from both classes
- The wider the margin, the better the model generalizes to new data



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

How does SVM Work ?

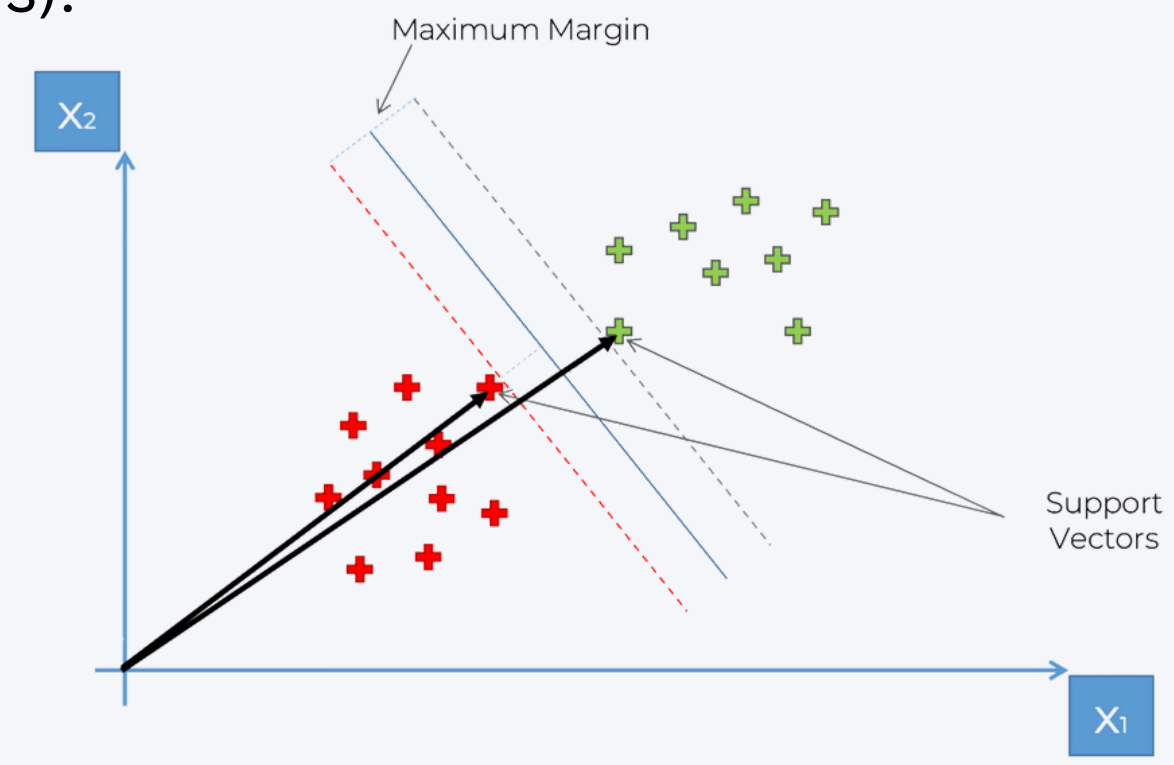
Hard vs. Soft Margin

- If all points can be separated perfectly, it's a hard margin SVM (strict rule).

- If some misclassification is allowed (overlapping points), it's a soft margin SVM (allows flexibility).

How Classification Works:

- New data points are classified based on which side of the hyperplane they fall on.
-
    - If a point falls left of the black line, it is classified as Category 1 (red cross).
    - If a point falls right of the black line, it is classified as Category 2 (green plus).



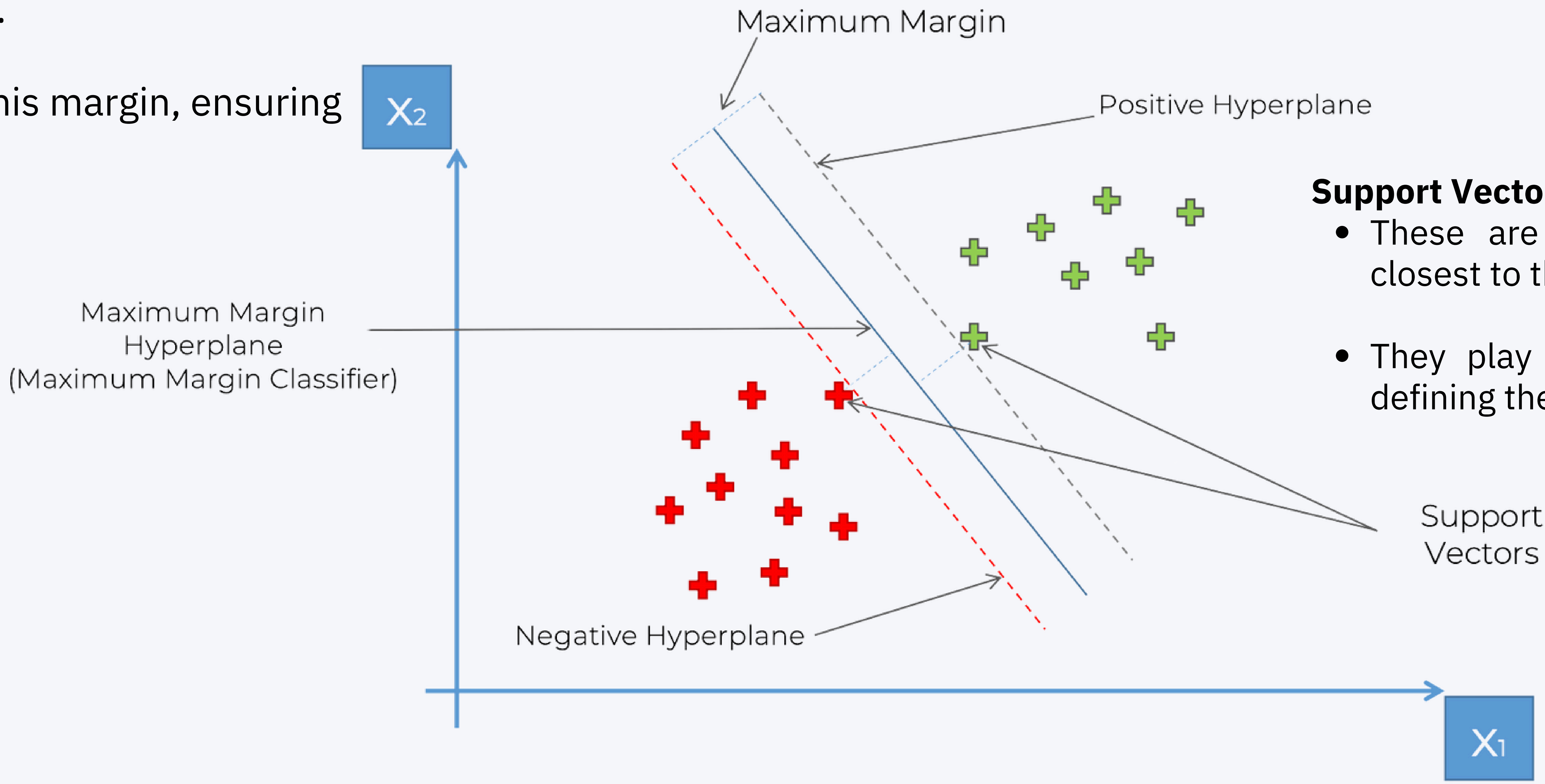
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

How does SVM Work ?

Margin:

- The distance between the hyperplane and the nearest support vectors.
- SVM tries to maximize this margin, ensuring better generalization.



Support Vector:

- These are the data points closest to the hyperplane.
- They play a critical role in defining the boundary.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

What Is So Special About SVMs?

Finds the Optimal Decision Boundary:

- Unlike other classification models that just find any separating boundary**, SVM finds the best one.**
- It maximizes the margin (distance) between the closest data points (support vectors) from each class.
- A larger margin means better generalization to new data.

Why is this important?

- It makes SVM more robust to small changes in data.
- Other models, like logistic regression, may find multiple boundaries, but SVM ensures the best possible separation.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

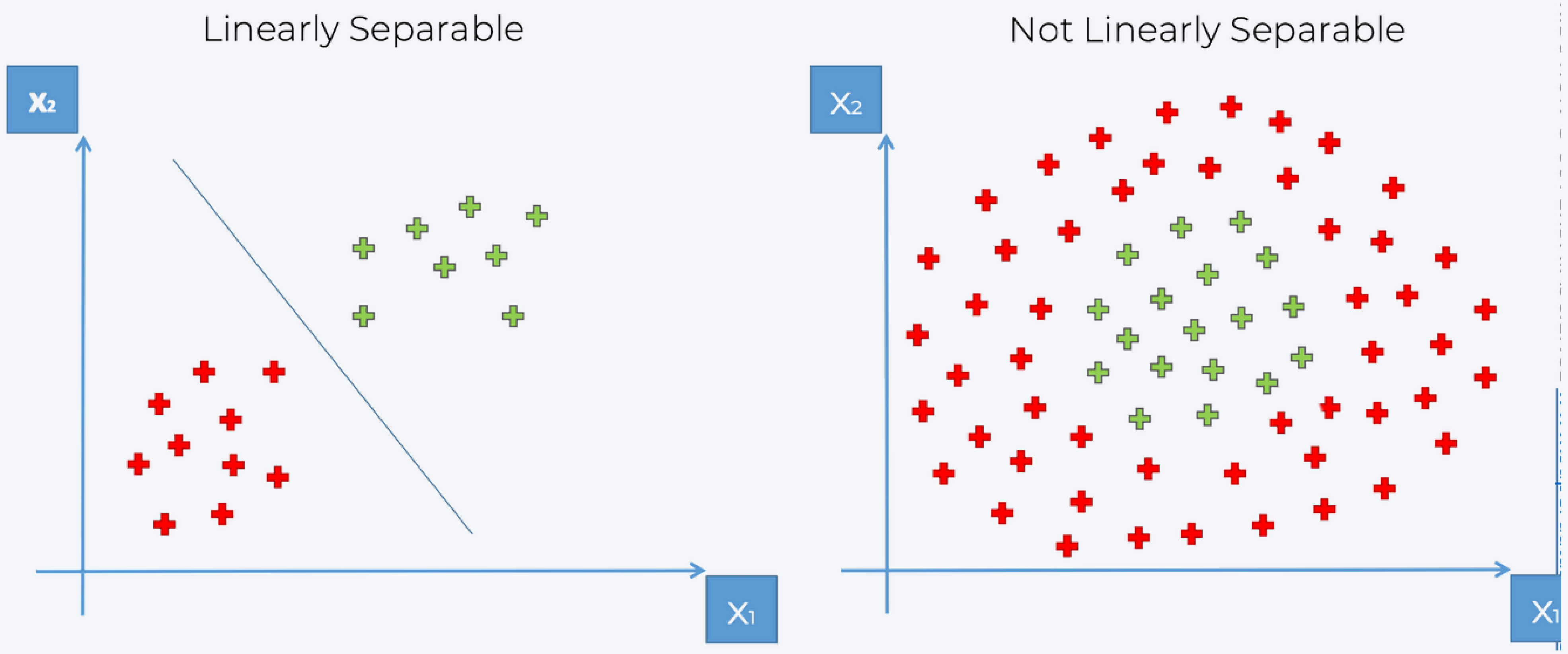
What Is So Special About SVMs?

Handles Non-Linearly Separable Data with the Kernel Trick:

- If the data is not linearly separable, SVM applies the Kernel Trick to map it into a higher-dimensional spacewhere separation is possible.
- This makes SVM highly flexible and capable of solving complex classification problems.

Example:

- Suppose we have data forming a circular pattern that cannot be separated with a straight line.
- Using the Radial Basis Function (RBF) Kernel, SVM maps the data to a higher dimension where a clear linear boundary exists.



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

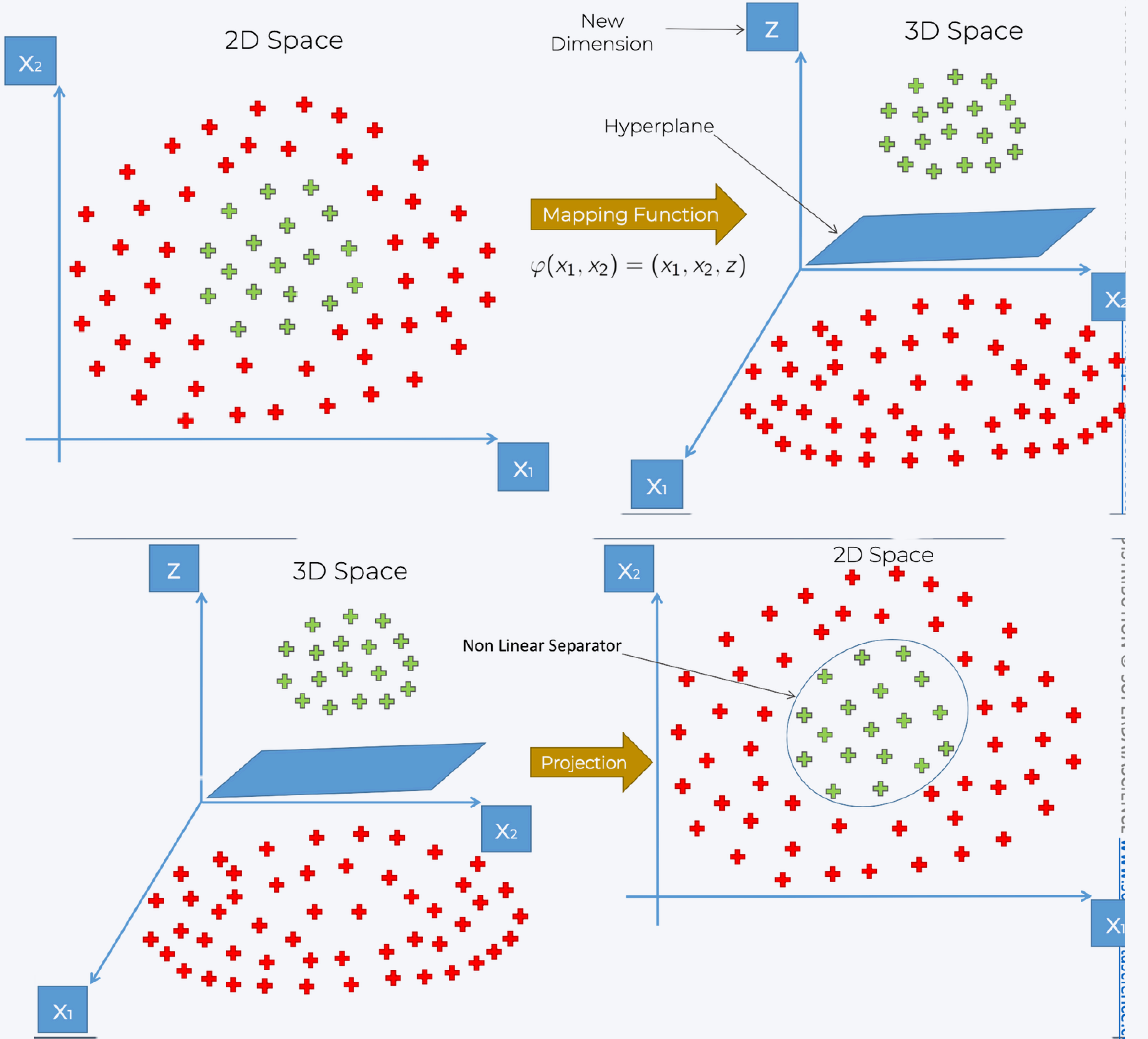
What Is So Special About SVMs?

Works Well in High-Dimensional Spaces

- SVM performs exceptionally well when data has many features (high-dimensional data).
- Many machine learning models struggle with high-dimensional spaces, but SVM remains effective by finding a clear hyperplane.

Examples:

- Text classification (Spam vs. Non-Spam), where each word is a feature.
- Image recognition, where each pixel can be a feature.



Support Vector Machine (SVM)

When To Use / When NOT ?

✅ When Should You Use SVM?

- When data is high-dimensional (e.g., text, images).
- When classes are well-separated.
- When the dataset is small to medium-sized.
- When you need a robust model against outliers.
- When a non-linear decision boundary is required (using kernels).

🚀 When NOT to use SVM?

- When the dataset is huge, as SVM can be slow.
- When there is too much noise, as SVM might struggle to find a clear boundary.
- When the number of features is too large, leading to longer training times.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

Comparison With Other Models

Feature	SVM	Logistic Regression	K-NN
Type	Classification/regression	Classification	Classification
Best for	Complex, high-dimensional data	Simple, linearly separable data	Small datasets, intuitive classification
Computational Cost	High (especially for large datasets)	Low	High (for large datasets)
Handles Non-Linearity?	Yes, with kernel trick	No	Yes (implicitly)

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Support Vector Machine (SVM)

BUT WAIT !!

Mapping to a Higher Dimensional Space can be highly compute-intensive

- **As the number of dimensions increases, the computational complexity grows exponentially.**
  - If a dataset has 10 features, mapping it to a quadratic kernel results in ($10^2 = 100$ dimensions).
  - If mapped to a cubic kernel, it results in ($10^3 = 1,000$ dimensions).
- **The Kernel Trick involves computing a Gram matrix (Kernel matrix) that stores the similarity between every pair of data points.**
  - This requires computing a dot product for every pair, leading to a complexity of $O(n^2)$ or worse.
  - For 1,000 data points, an RBF kernel requires computing a $1,000 \times 1,000$ matrix.
- **SVM's optimization problem (finding the optimal hyperplane) depends on solving quadratic programming problems.**

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Hands-On Code

Support Vector Machine

KERNAL

SVM

KERNAL SVM

The Gaussian RBF Kernel Formula

The Gaussian RBF Kernel transforms data into a higher-dimensional space to make it linearly separable.

$$K\left(\vec{x}, \vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

- x → A data point
- li → A landmark (reference point in the dataset)
- σ (sigma) → A parameter controlling the width of the transformation

- If x is close to **li →K(x,li)** is close to **1** (high similarity).
- If x is far from **li → K(x,li)** is close to **0** (low similarity).

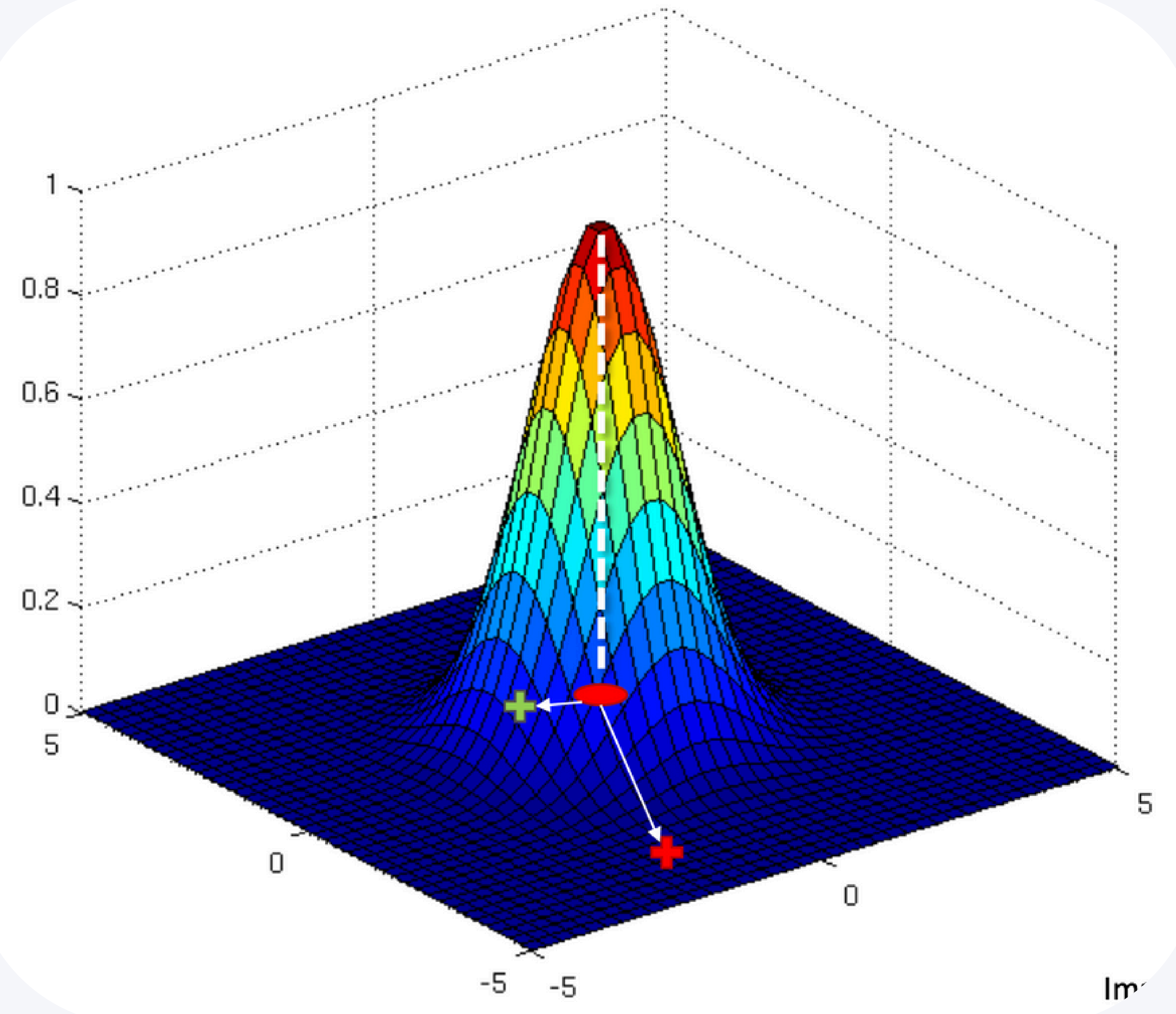
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

KERNAL SVM

The Gaussian RBF Kernel Formula

- The kernel function maps data points into a higher-dimensional space.
- The 3D plot shows a Gaussian bump centered at a landmark point.
- Higher values (closer to 1) mean high similarity to the landmark.
- Lower values (closer to 0) mean low similarity.

$$K(\vec{x},\vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

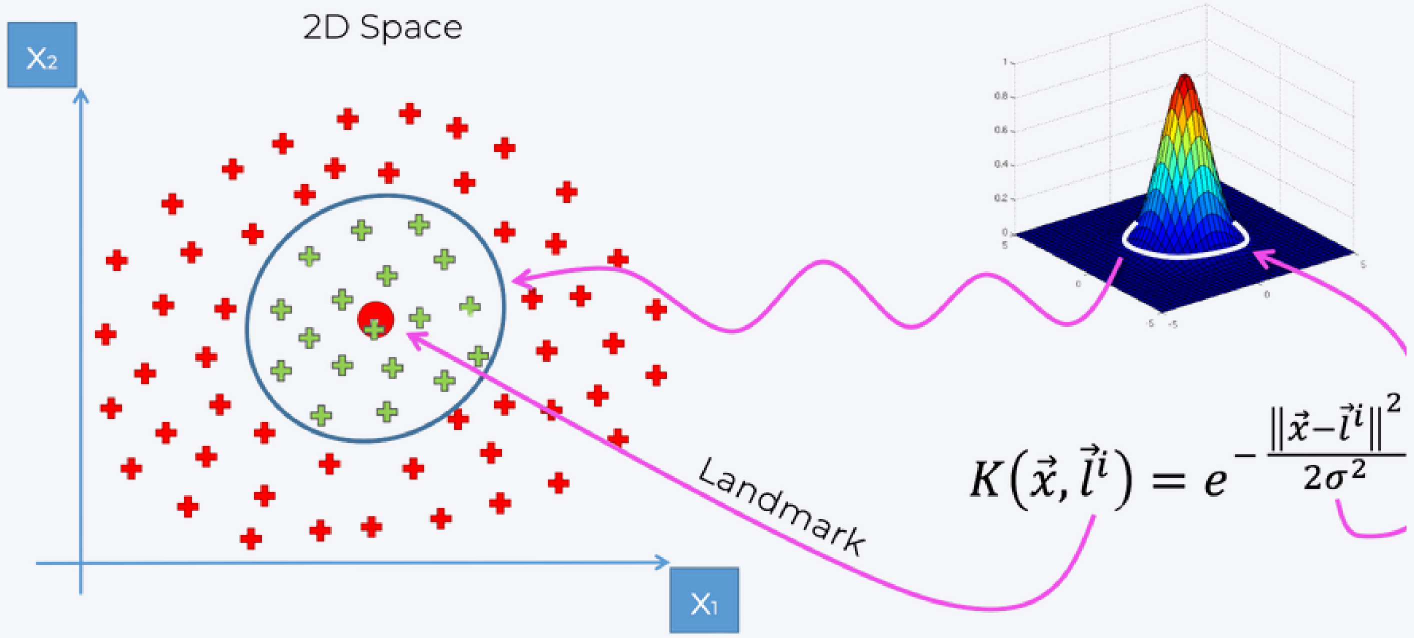
KERNAL SVM

Tuwaiq Academy أكاديمية طويق

How the Gaussian RBF Kernel Maps Data

- The original dataset (left) is in a 2D space, where two classes overlap.

- The Gaussian RBF Kernel transforms the data into a higher-dimensional space, where:
  - Points near the landmark (center green area) have high values.
  - Points farther away (red points) have lower values.



Effect of the Transformation:

- The decision boundary becomes circular, instead of linear.
- This allows SVM to properly separate the two classes.

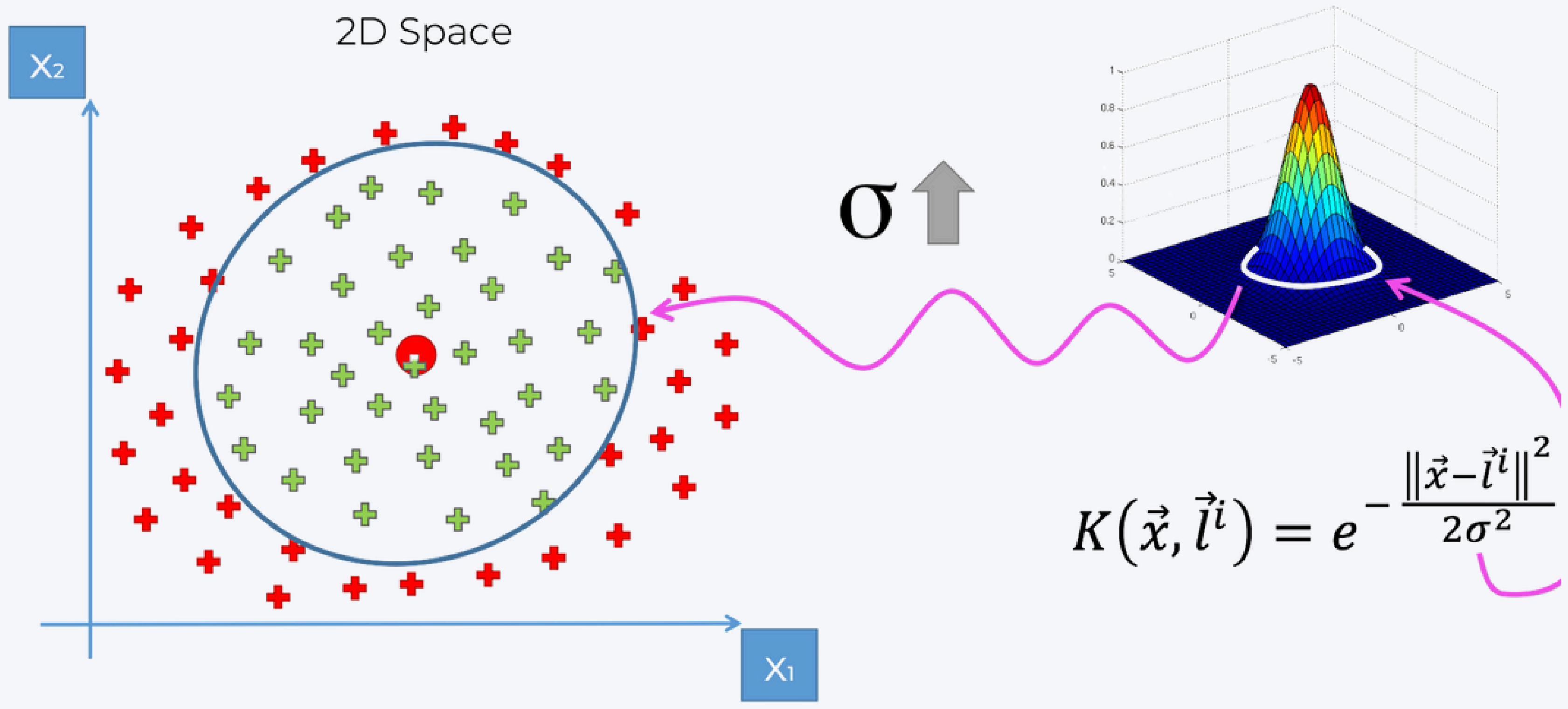
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

KERNAL SVM

Tuwaiq Academy

The Effect of Increasing σ (Sigma)

- The parameter **σ** controls how far the influence of a landmark spreads.
- Increasing **σ** makes the kernel spread out, meaning:
    - More points are considered similar.
    - The decision boundary becomes smoother.



Effect on Classification:

- A larger σ makes decision boundaries wider and smoother.
- Useful when classes are widely spread out.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

KERNAL SVM

The Effect of Decreasing σ (Sigma)

- Reducing **σ** makes the influence of the landmark smaller.

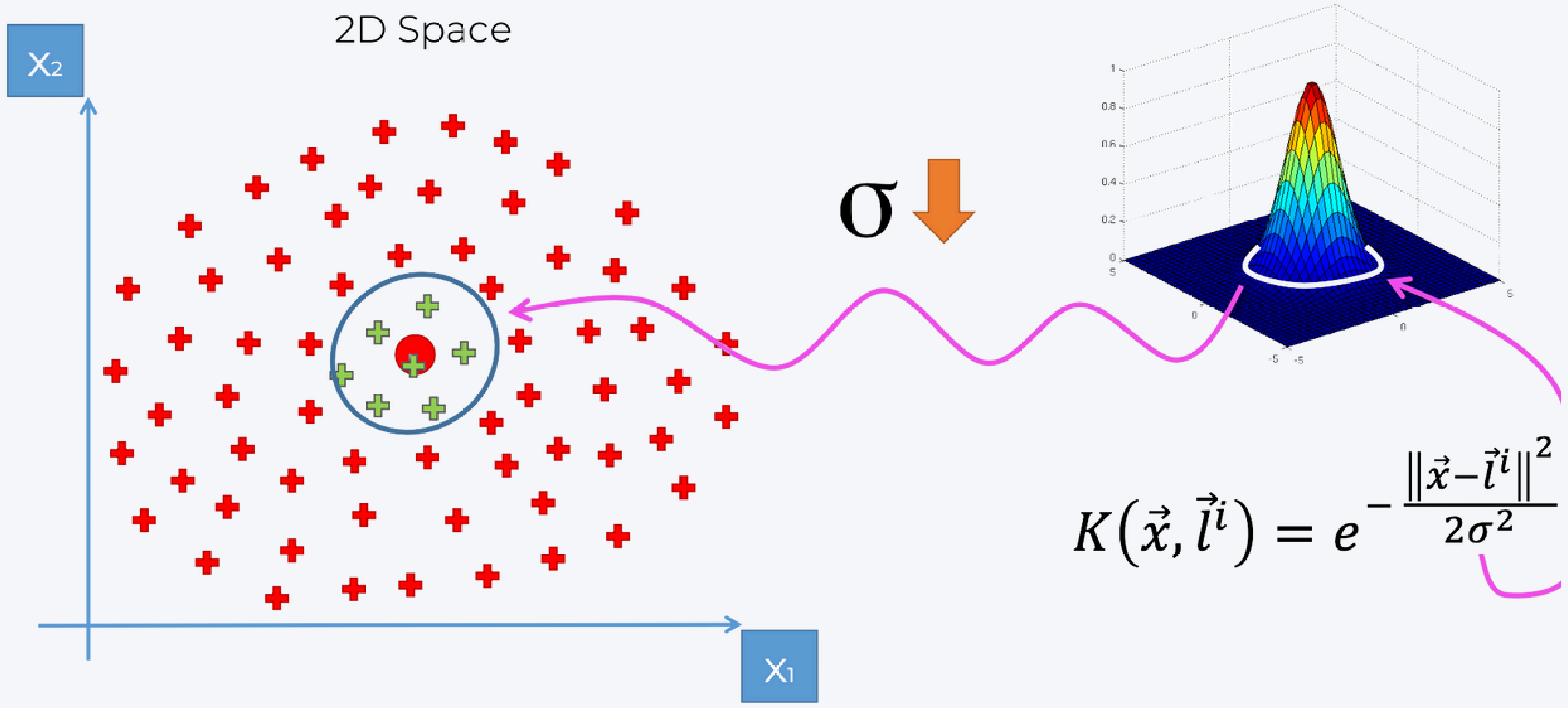
- The decision boundary becomes **tighter** around points.

- The model focuses **only** on very close points.

Risk of Small σ:

- If σ is too small, the decision boundary can become too **complex**.

- This leads to **overfitting** (memorizing training data but performing poorly on new data).



$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

KERNAL SVM

Combining Multiple Gaussian Kernels

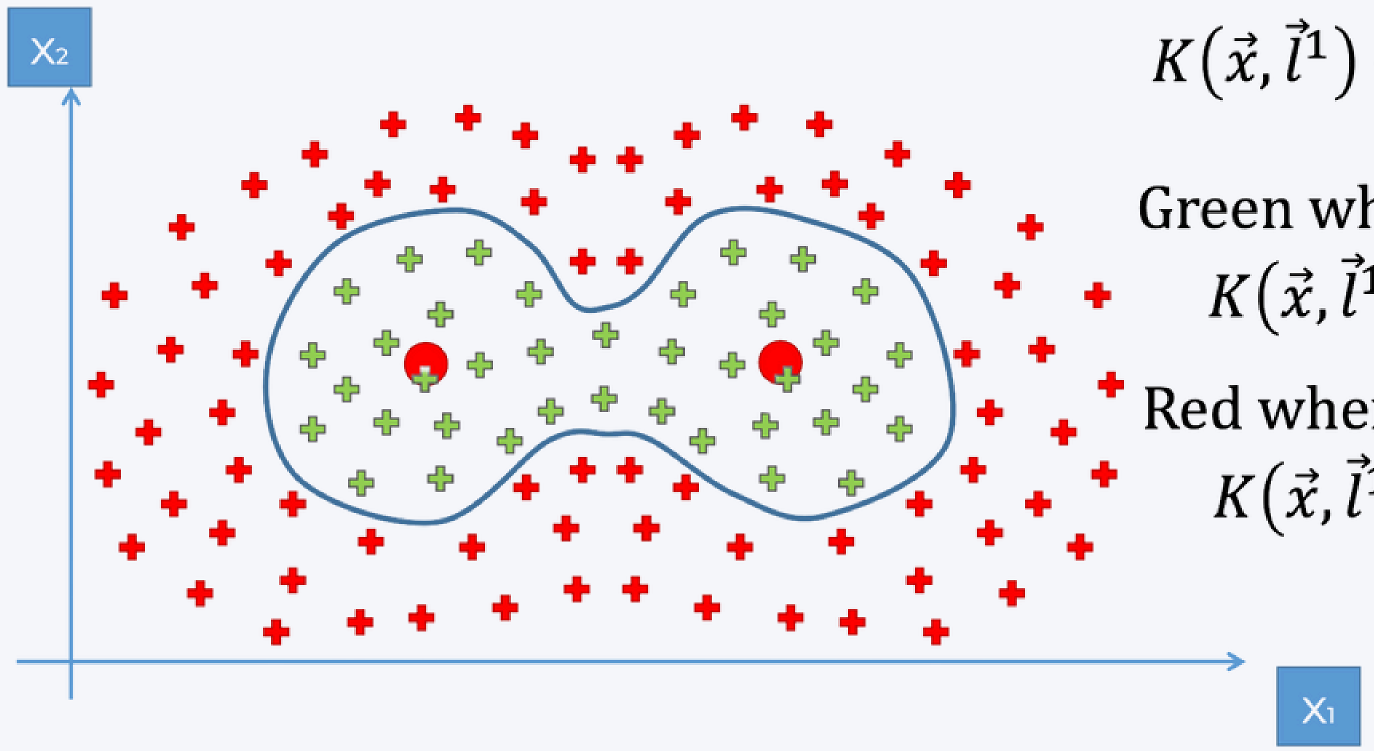
- Instead of one landmark, we use multiple landmarks across the dataset.

- The final transformation is a sum of multiple Gaussian functions.

Effect on the Decision Boundary:

- The model creates a complex but smooth decision boundary.

- The boundary can capture non-linear patterns in the data.



$K(\vec{x}, \vec{l}^1) + K(\vec{x}, \vec{l}^2)$

(Simplified Formula)

Green when:

$K(\vec{x}, \vec{l}^1) + K(\vec{x}, \vec{l}^2) > 0$

Red when:

$K(\vec{x}, \vec{l}^1) + K(\vec{x}, \vec{l}^2) = 0$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

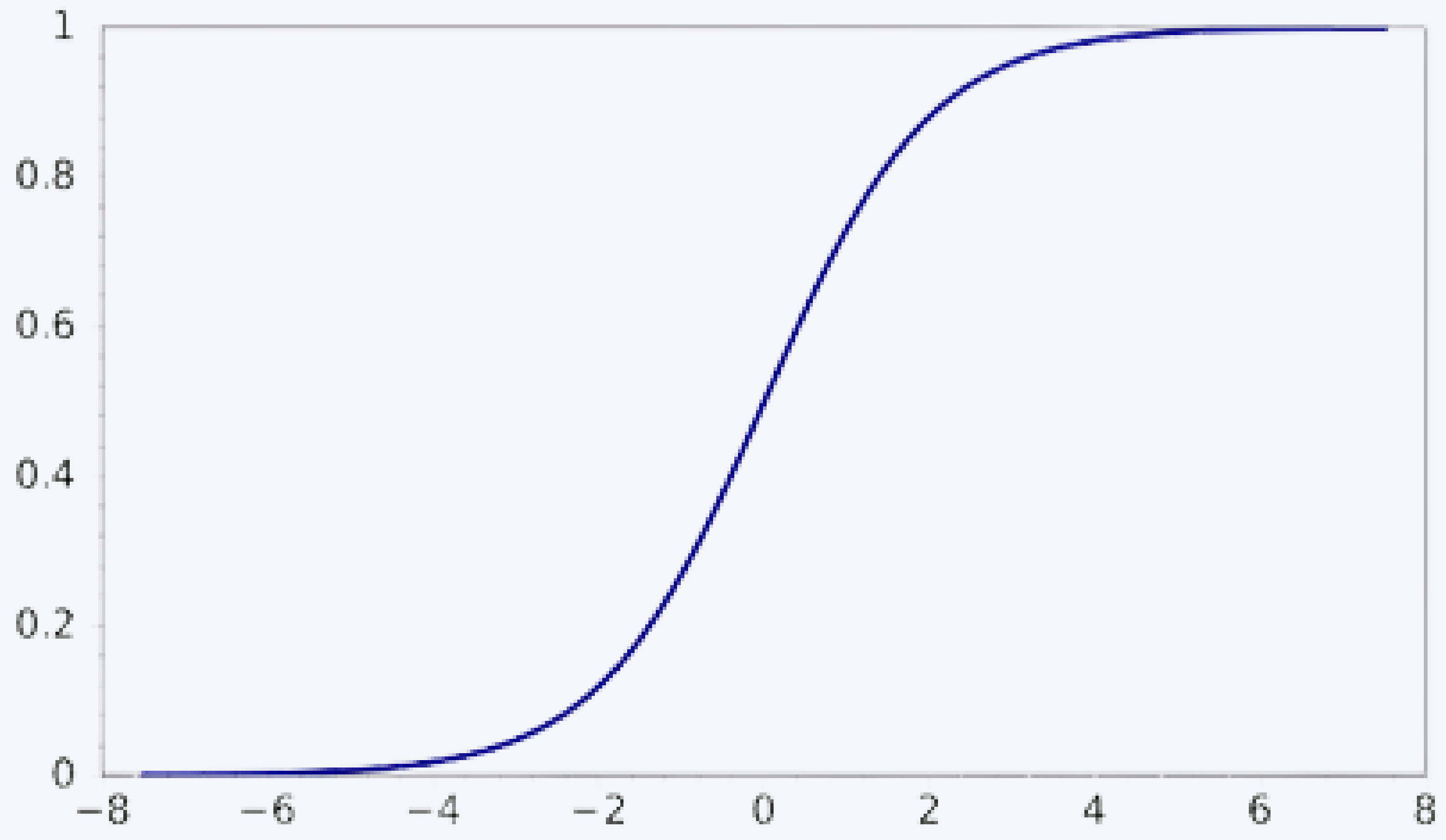
KERNAL SVM

Types of Kernel Functions in SVM

Sigmoid Kernel

$$K(X, Y) = \tanh(\gamma \cdot X^T Y + r)$$

- Based on the hyperbolic tangent (tanh) function, similar to neural networks.
- Transforms the data using a sigmoid-like shape.
- The parameters γ (gamma) and r control the shape of the transformation.



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

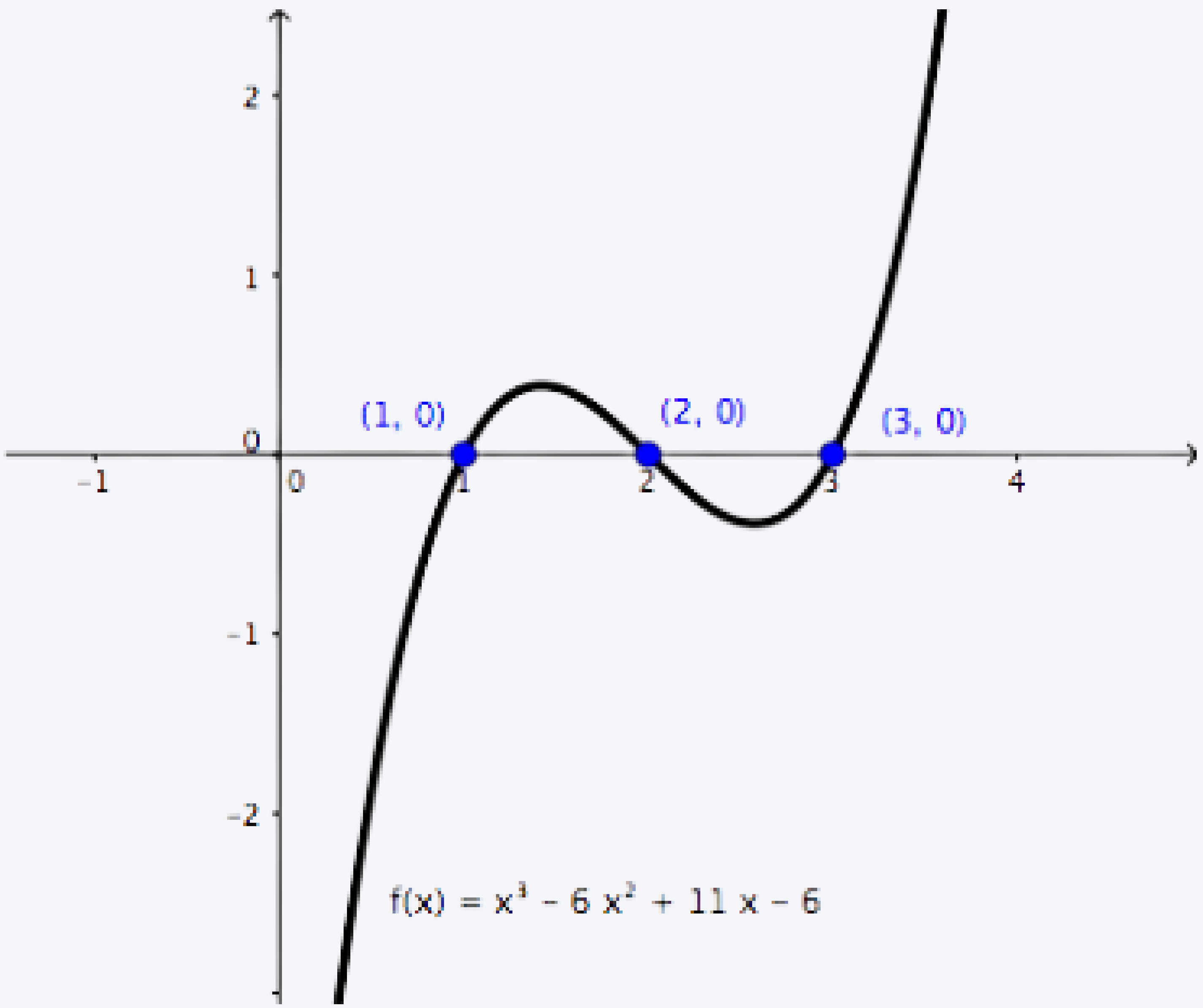
KERNAL SVM

Types of Kernel Functions in SVM

Polynomial Kernel:

$$K(X,Y) = (\gamma \cdot X^T Y + r)^d, \quad \gamma > 0$$

- Maps the input space into a higher polynomial degree space.
- The degree (d) determines the complexity of the decision boundary.



When to Use It:

- When there is a polynomial relationship between features.
- When data is not too high-dimensional, as high degrees can be computationally expensive.

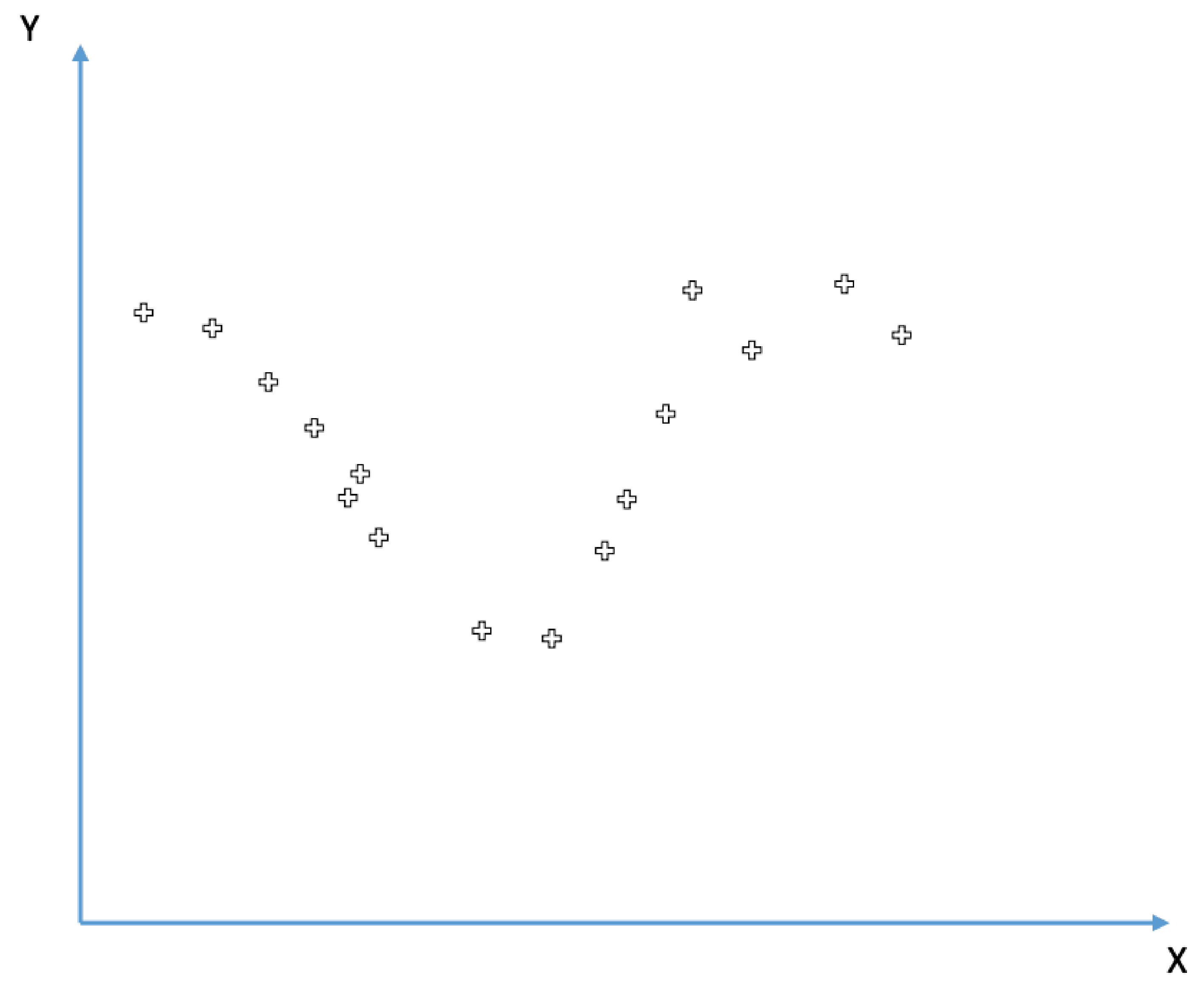
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

NON-Linear SVM (Advanced)

NON-Linear SVM

The Problem – Non-Linear Data

- The scatter plot shows data that does not follow a straight line.
- A simple linear model will not be able to capture this pattern.
- We need a method to handle this non-linearity.

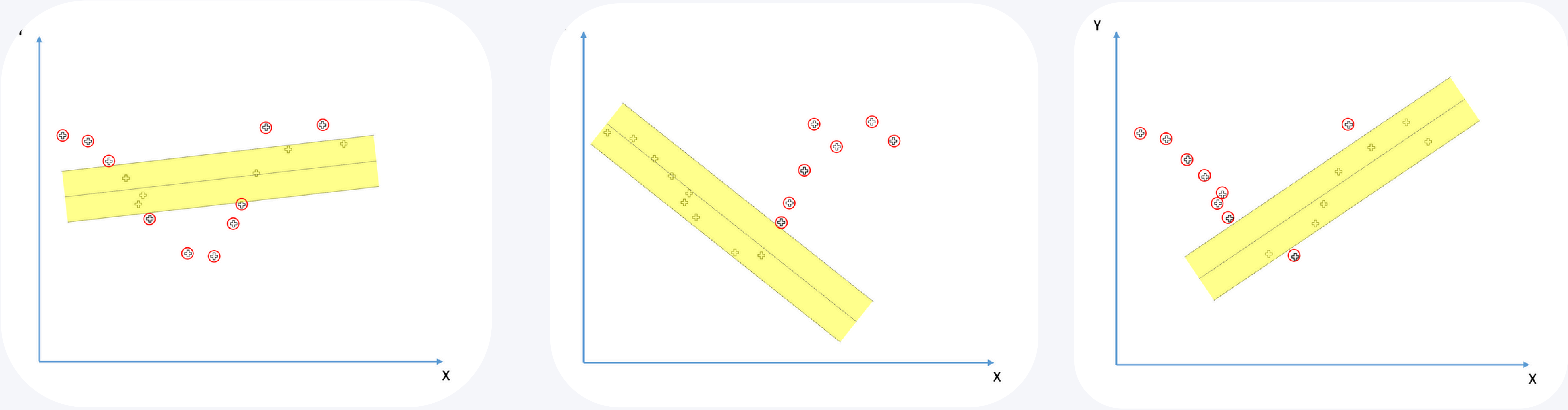


https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

NON-Linear SVM

أكاديمية طويق
Tuwaiq Academy

Attempting a Linear Fit



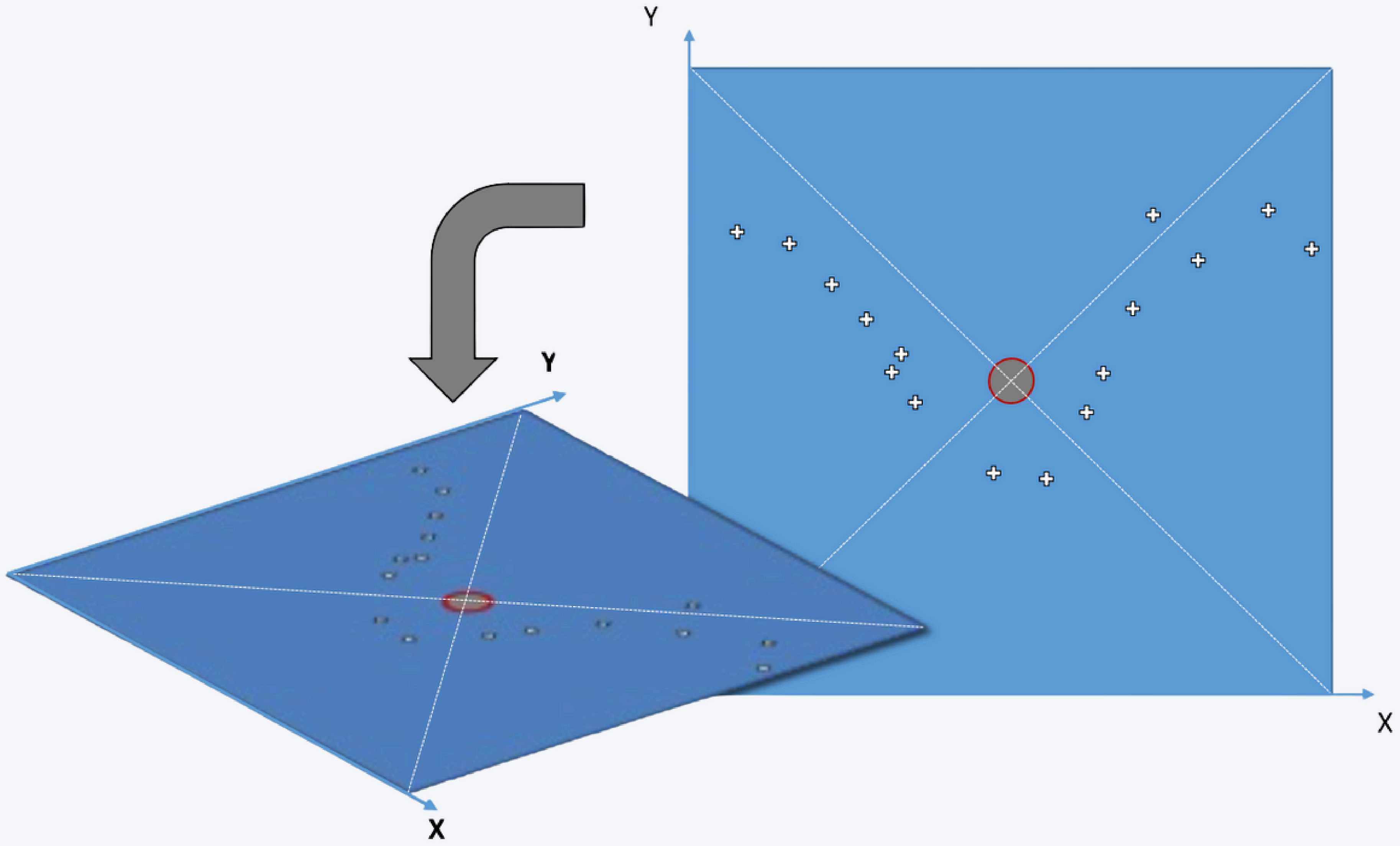
- A linear SVR model tries to fit a straight-line regression with a margin (yellow band).
- The margin represents the acceptable error (epsilon-tube), where points inside it are considered close enough to the prediction.
- However, since the data follows a curved pattern, many points fall outside the margin (red circled points), indicating a poor fit.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

NON-Linear SVM

Moving to a Higher-Dimensional Space

- Instead of forcing a straight-line fit, we use the **Kernel Trick.**
- The idea is to map the data into a **higher dimension**
- where it becomes easier to **fit a curve.**
- In the new space, **the pattern becomes clearer.**

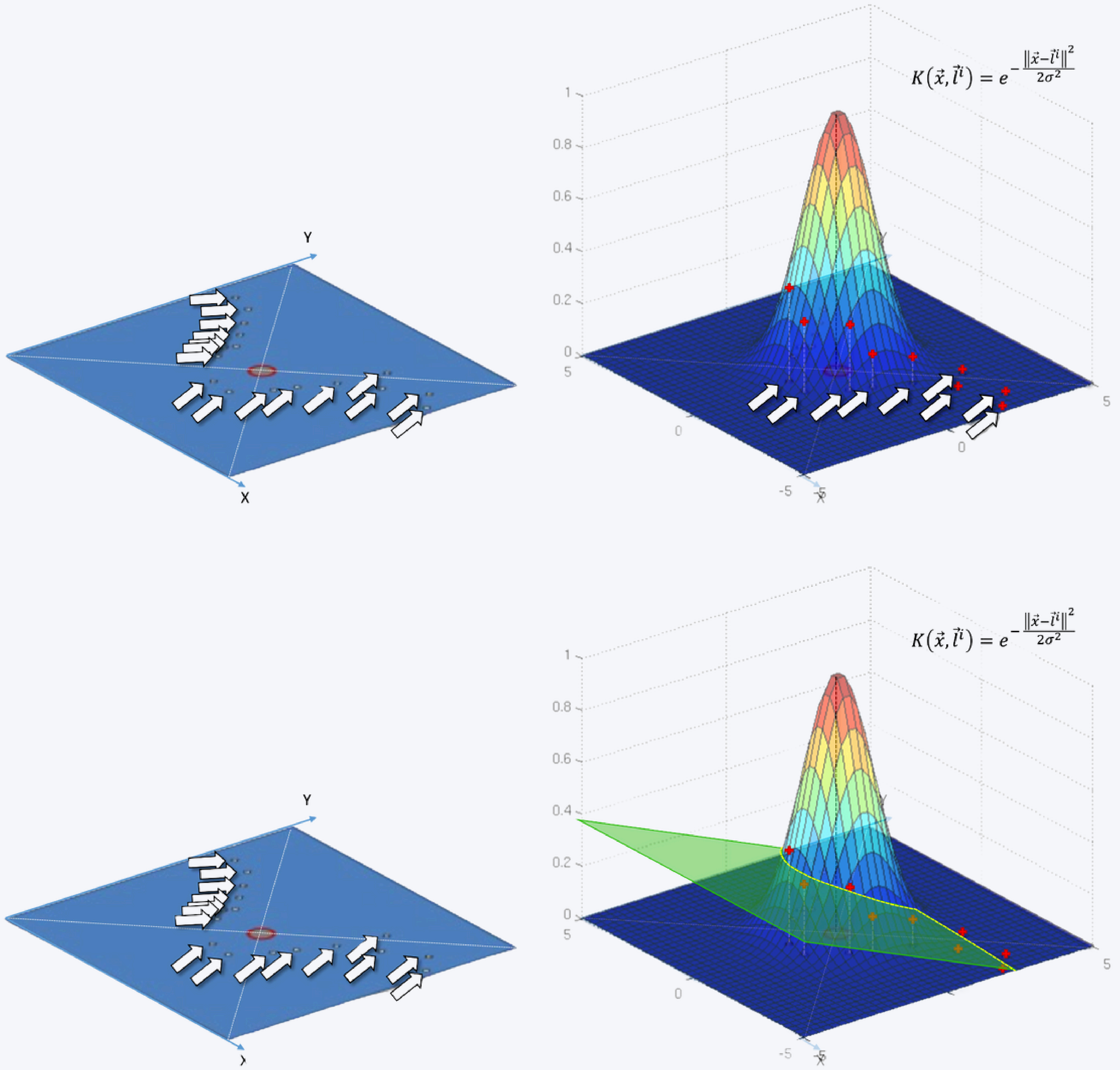


https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

NON-Linear SVM

Moving to a Higher-Dimensional Space

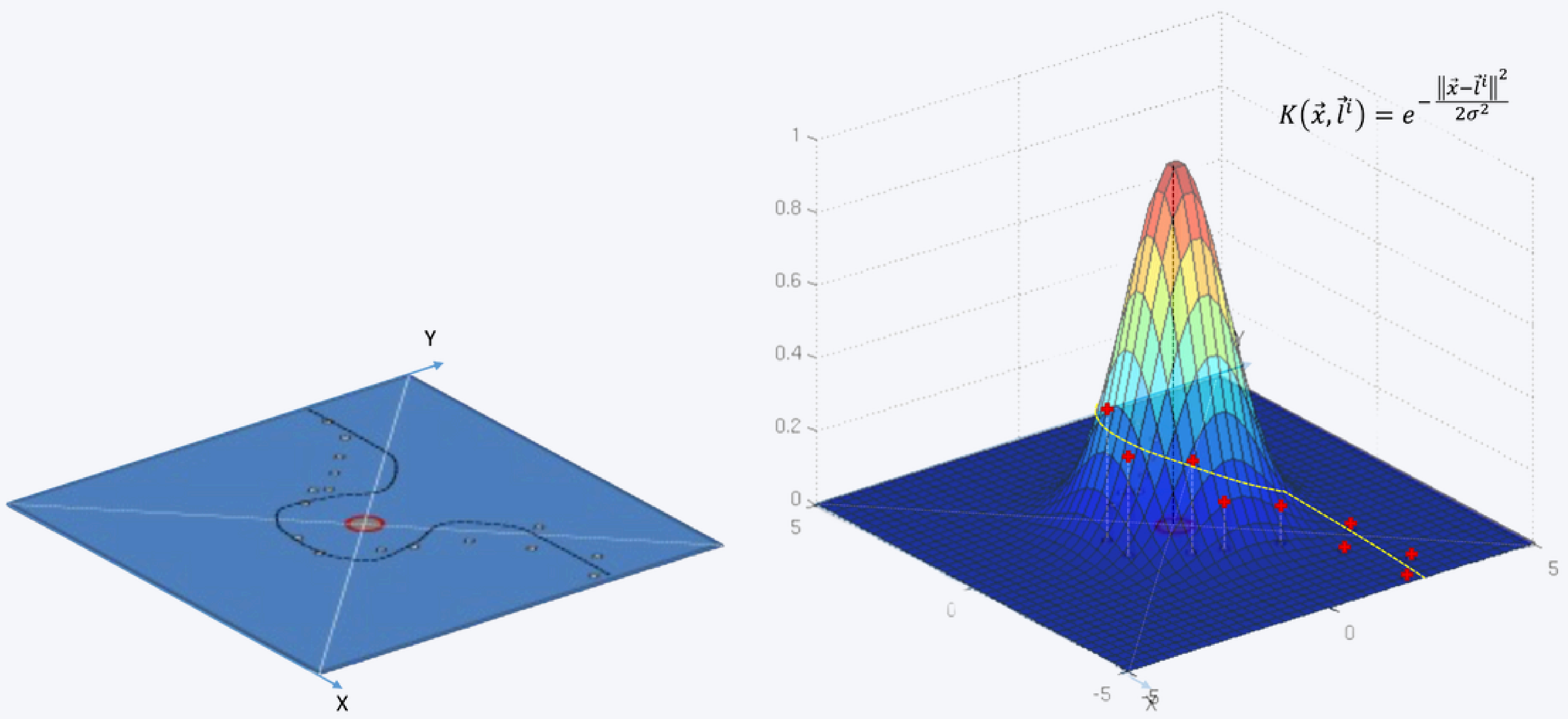
- The Gaussian Radial Basis Function (RBF) Kernel is used to map data into a higher dimension.
- In this new space, the data forms a peak-like structure.
- SVR can now fit a better regression model by drawing a flat plane in the transformed space.



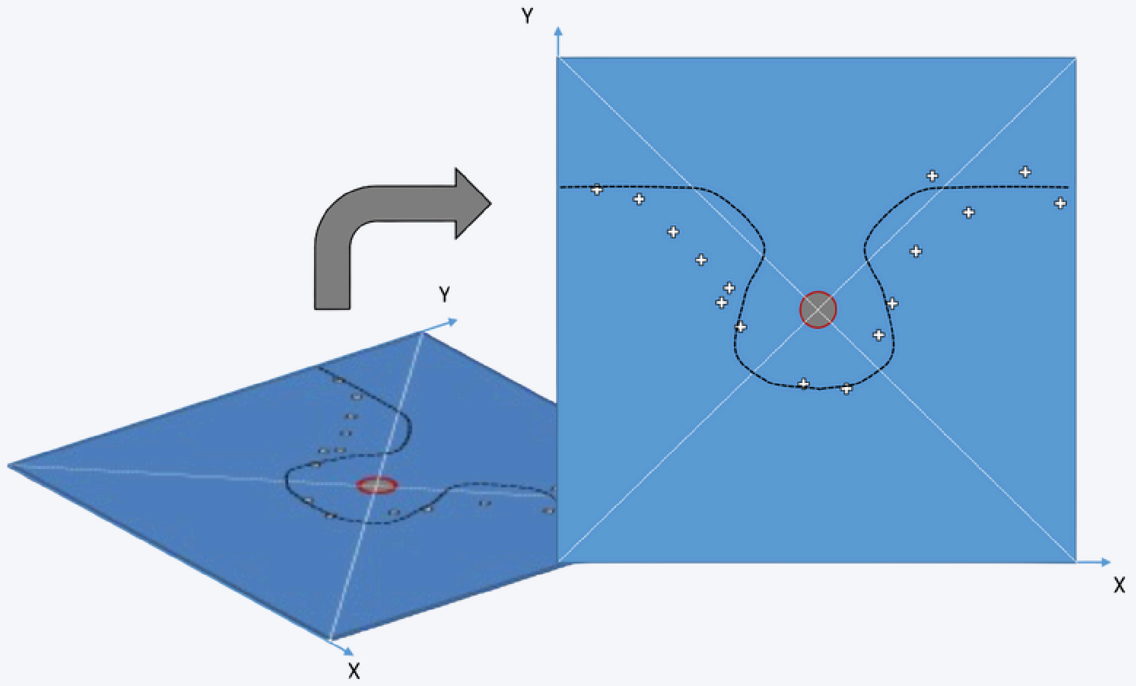
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

NON-Linear SVM

Moving to a Higher-Dimensional Space



- SVR finds the best-fitting surface in the transformed space.
- This ensures that the model captures the curved pattern of the original data.



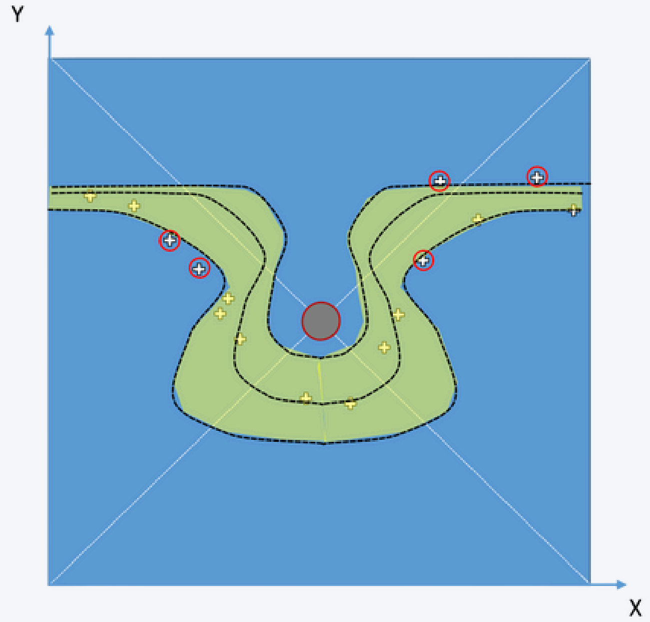
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

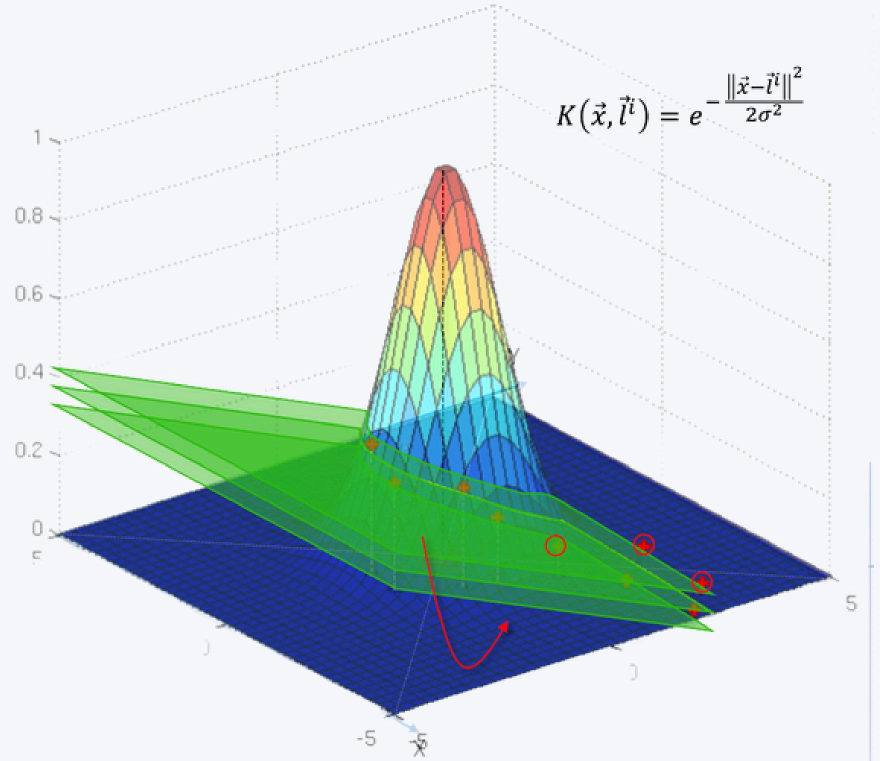
NON-Linear SVM

Moving to a Higher-Dimensional Space

- Once the model is trained in the higher-dimensional space, we map it back to the original space.

- The result is a smooth, non-linear regression curve that properly fits the data.





https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

NON-Linear SVM

Summary of SVMs

- Once the model is trained in the higher-dimensional space, we map it back to the original space.

- The result is a smooth, non-linear regression curve that properly fits the data.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Hands-On Code

KERNAL SVM