⭐ CHAPTER — KERNEL TRICK (Full With All Types)

❗ Problem:

Data is not linearly separable.

❤️ Kernel Trick:

Map data to a higher dimension **implicitly** using a kernel function → separation becomes possible.

🔥 4 Most Important Kernel Types

1️⃣ LINEAR KERNEL

$K(x,x') = x \cdot x'$K(x, x') = x \cdot x'K(x,x′)=x⋅x′

📌 What it does:

Computes **dot product** between two points.

📌 Meaning:

Measures whether two points are pointing in the same direction.

📌 When to use:

- Data is already **linearly separable**
- Features are **high-dimensional**
  (text data, NLP, TF-IDF vectors, gene expression)

📌 Decision boundary:

Straight line / hyperplane.

📌 Benefits:

- Fast
- No extra parameters
- Works great when no non-linearity required

2️⃣ POLYNOMIAL KERNEL

K(x,x′)=(x⋅x′+c)dK(x, x') = (x \cdot x' + c)^dK(x,x′)=(x⋅x′+c)d

📌 Purpose:

Allows **non-linear curves** of degree d.

📌 Intuition:

Adds interactions like:

- x12x_1^2x12
- x1x2x_1 x_2x1x2
- x23x_2^3x23
  … etc.

📌 When useful:

- When relationship is not linear but **smooth & polynomial**
- Example: price prediction, some image boundaries

📌 Shape of boundary:

Curved boundary with polynomial complexity.

📌 Notes:

- Degree (d) increases complexity
- High degree = overfitting possible

3️⃣⭐ RBF (Gaussian) Kernel — MOST IMPORTANT

$K(x,x') = e^{-\gamma ||x - x'||^2}$

📌 What it measures:

Distance-based similarity.

- Points close → similarity ≈ 1
- Points far → similarity ≈ 0

📌 Intuition:

Creates **radial / circular** zones around points.

📌 Real power:

It maps data to an **infinite-dimensional space**.

📌 When to use:

- Data is **highly non-linear**
- Circles, spirals, moons
- Complex ML tasks

- When you don’t know the shape of boundary

📌 Decision boundary:

Highly flexible, curved, smooth.

📌 Parameter:

- γ controls how tight the curve is

4️⃣ SIGMOID KERNEL

$K(x,x') = \tanh(\alpha (x \cdot x') + c)$

📌 Based on:

Neural networks activation function (tanh).

📌 Intuition:

Acts like a **two-layer neural network**.

📌 When used:

- Rarely used now
- Used in early SVM research
- Works well when data resembles NN behavior

📌 Notes:

- Can behave unpredictably
- Sensitive to parameters
- Not as stable as RBF