# Kernel Method

CS 3780/5780

PollEv.com/karthiksridharan488

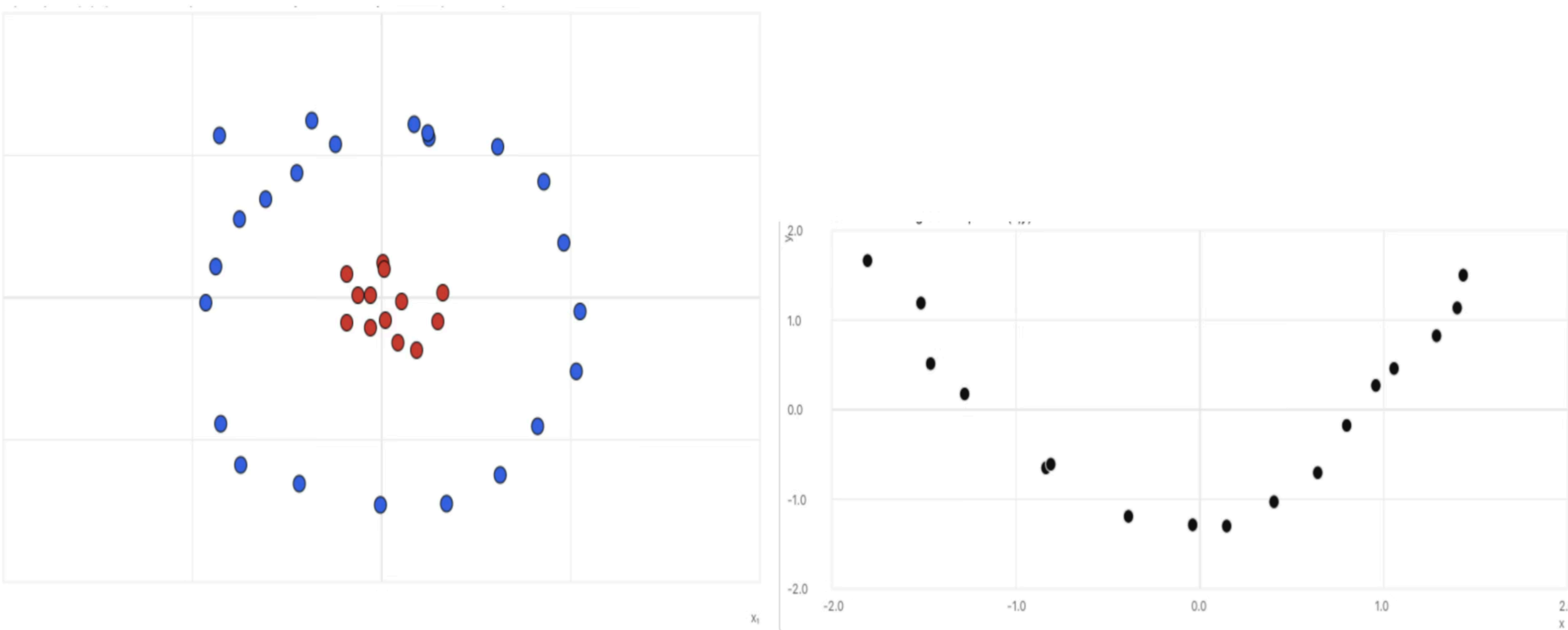
# Kernel Method

- Linear classification methods
  - perceptron
  - Logistic regression
  - Support Vector Machines
- Linear regression methods
  - Ordinary Least squares
  - Ridge regression

- Generic form:

$$\mathbf{w}, b = \arg\min_{\mathbf{w}\in\mathbb{R}^d, b\in\mathbb{R}} \sum_{i=1}^{n} \ell(\mathbf{w}^\top \mathbf{x}_i + b, y_i) + \lambda \mathbf{w}^\top \mathbf{w}$$

# Need Non-linearity



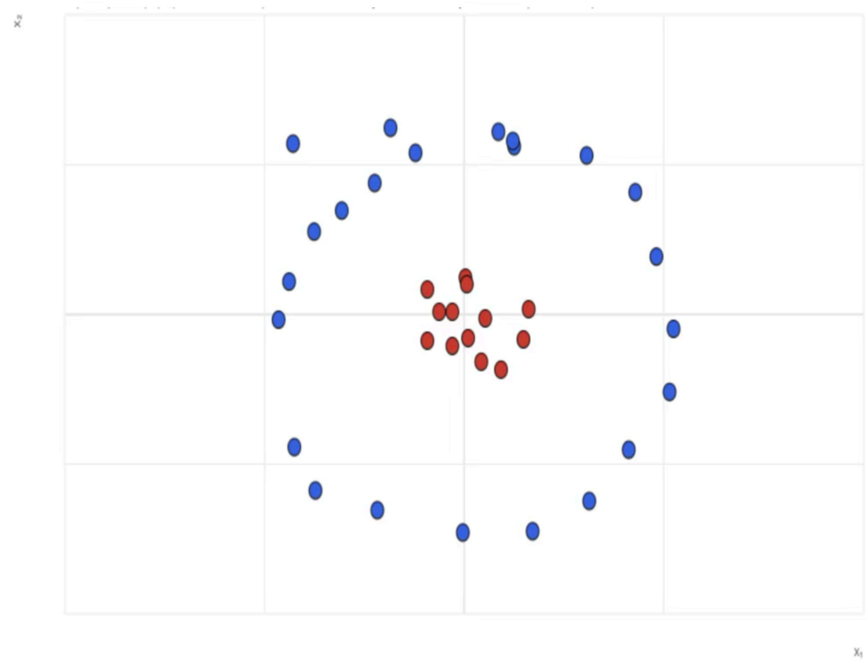

(a) Classification (b) regression

Figure 1: Need Non-linearity

# Magic 1: Feature space

- Map x to feature space: x -> $\Phi(x)$ (high dimensional)

- Use linear method in feature space



(a) Classification

Quiz 1: Suggest a feature mapping for the classification example above so that once we apply linear method on feature space, like SVM for instance, we get good classification error on the example.

**Magic 1: Linear in $\phi(x)$ leads to non-linearity in x space.**

# Quiz 2

Quiz 2: If $\mathbf{x}$ were $d$ dimensional, and:

1. we wanted to capture all functions of $\mathbf{x}$ that are polynomials of up to degree 2, how many dimensions would $\phi(\mathbf{x})$ be?
2. we wanted to capture all functions of $\mathbf{x}$ that are polynomials of up to degree $p$, how many dimensions would $\phi(\mathbf{x})$ be?

# Magic 2

- We can even have Φ(x) to be infinite dimensional !!!!
- This Magic: never explicitly enumerate Φ(x), implicitly compute k(x,x') = <Φ(x), Φ(x')>

Quiz 3: What is $\phi(\mathbf{x})$ for the following kernel function: $k(x, x') = (1 + x \cdot x')^p$?

# Examples of Kernels

- Linear kernel: $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$
- Polynomial kernel: $k(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^\top \mathbf{x}')^p$
- Radial Basis Fucntion (RBF) kernel: $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{x}'\|^2\right)$

# Magic 2

- We can even have Φ(x) to be infinite dimensional !!!!
- This Magic needs 2 steps:
  - Step 1: linear methods produce solutions of the form

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)$$

  - Step 2: Using k(x,x') = <Φ(x), Φ(x')>:

$$h(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}) + b = \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b$$

# Representer Theorem

**Claim: Any w that is solution to optimization problem**

$$\mathbf{w}, b = \arg\min_{\mathbf{w},b} \sum_{i=1}^{n} \ell(\mathbf{w}^\top \Phi(\mathbf{x}_i) + b, y_i) + \lambda \mathbf{w}^\top \mathbf{w}$$

**takes the form** $\mathbf{w} = \sum_{i=1}^{n} \alpha_i \Phi(\mathbf{x}_i)$

# How to find the α's

$$\begin{aligned}
\alpha_1, \ldots, \alpha_n, b &= \underset{\alpha_1,\ldots,\alpha_n \in \mathbb{R}, b \in \mathbb{R}}{\arg\min} \sum_{j=1}^{n} \ell\Big(\sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)^\top \phi(x_j) + b, y_j\Big) + \lambda \left(\sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)\right)^\top \left(\sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)\right) \\
&= \underset{\alpha_1,\ldots,\alpha_n \in \mathbb{R}, b \in \mathbb{R}}{\arg\min} \sum_{j=1}^{n} \ell\Big(\sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)^\top \phi(x_j) + b, y_j\Big) + \lambda \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) \\
&= \underset{\alpha_1,\ldots,\alpha_n \in \mathbb{R}, b \in \mathbb{R}}{\arg\min} \sum_{j=1}^{n} \ell\Big(\sum_{i=1}^{n} \alpha_i k(x_i, x_j) + b, y_j\Big) + \lambda \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j)
\end{aligned}$$

# Kernel Matrix and the Soln.

- Kernel matrix: K $$K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$$
- Alternate form for solution to alpha's

$$\alpha_1, \ldots, \alpha_n, b = \arg\min_{\alpha \in \mathbb{R}^n, b \in \mathbb{R}} \sum_{j=1}^{n} \ell(K\alpha + b, y_j) + \lambda \alpha K \alpha$$

# Are Functions Kernels?

**A function $k : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ corresponds to inner product w.r.t. some feature map $\phi$ if and only if the function is positive semi-definite. That is, for any set of points, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and any $n$, the kernel matrix given by $K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$ is positive semi-definite.**

- Complicated to verify, here are a set of simple rules.

# Algebra of Kernels

1. $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$
2. $k(\mathbf{x}, \mathbf{x}') = c k_1(\mathbf{x}, \mathbf{x}')$
3. $k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}')$
4. $k(\mathbf{x}, \mathbf{x}') = g(k(\mathbf{x}, \mathbf{x}'))$
5. $k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') k_2(\mathbf{x}, \mathbf{x}')$
6. $k(\mathbf{x}, \mathbf{x}') = f(\mathbf{x}) k_1(\mathbf{x}, \mathbf{x}') f(\mathbf{x}')$
7. $k(\mathbf{x}, \mathbf{x}') = \exp(k_1(\mathbf{x}, \mathbf{x}'))$
8. $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top A \mathbf{x}'$

where $k_1, k_2$ are valid kernels, $c > 0$, $g$ is a polynomial function with positive coefficients, $f$ is any function and $A$ is a positive definite matrix.

# Quiz 4

Quiz 4: Using the properties above show that the following functions are valid kernels:

$$k(x, x') = \exp\left(-\frac{(x - x')^2}{2\sigma^2}\right) \qquad \text{and} \qquad k(S, S') = \exp(|S \cap S'|)$$

where in the above $S_1, S_2 \subset \Omega$ where you can think of $\Omega$ as a finite set and for a set $A$, $|A|$ denotes the cardinality of $A$.