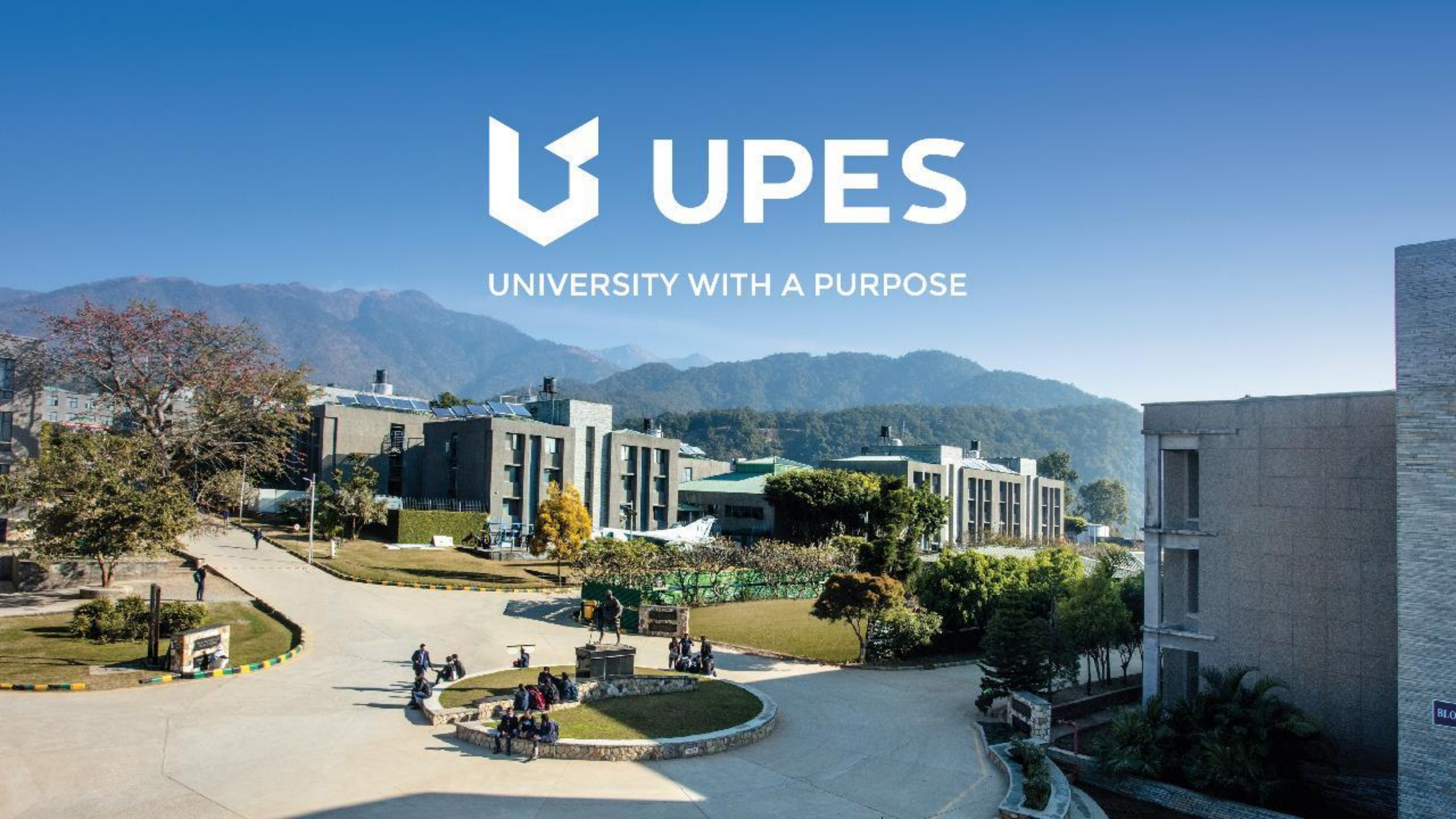
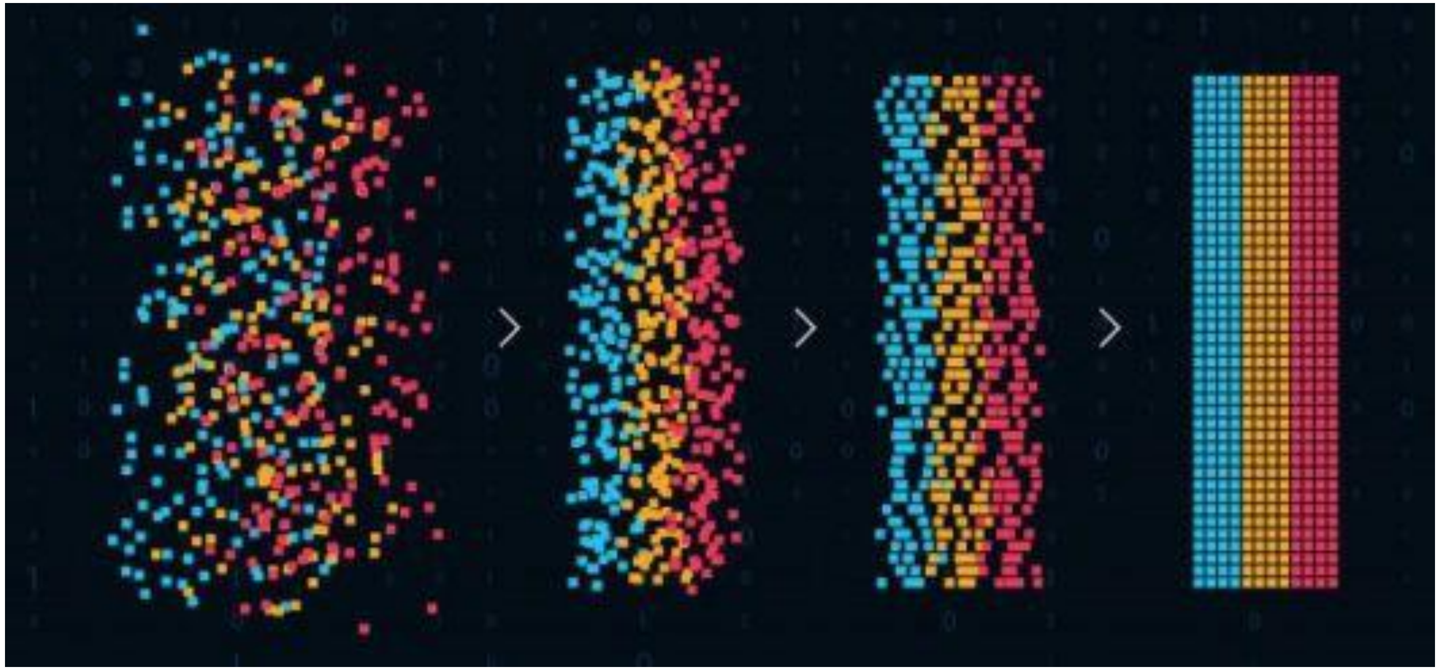

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML
Dr Gopal Singh Phartiyal
25/10/2021

Recap: Linear Models for Classification

- Goal of classification: Take input (let say x) and assign it to one of the K discrete classes C_k a classes where k = 1, 2, 3, …….K.
- Generic assumption: Classes are disjoint (an input can be assigned to one and only one class, no more no less)
- Models analogous to regression models but for classification problems
- The input space is divided into decision regions whose boundaries are termed as **decision boundaries** or **decision surfaces**.
- At first we will discuss linear models for classification? Decision surface.
- (D-1) dimensional Hyperplane is a linear function of D dimensional input .
- Datasets whose classes can be separated by linear decision surfaces are called linearly separable.

Recap: Linear Models for Classification

- For regression problems, the target variable t was simply the vector of real numbers whose values we wish to predict
- In the case of classification, there are various ways of using target values to represent class labels
- **Example**: Two-class problem solved by probabilistic models
- Most convenient is the binary representation

$$t \in \{0, 1\}$$

- Where, t = 1 represents class C_1 and t = 0 represents class C_2. Interpret the value of t as probability of class C_1.

Recap: Linear Models for Classification

- For more than two class: one hot encoding or one-of-K coding is used.

$$\mathbf{t} = (0, 1, 0, 0, 0)^{\mathrm{T}}$$

- For non-probabilistic models, alternative choices of target variable representation can be opted.
- Categories:
  - Discriminant functions
  - Generative
  - Deterministic

$$y(\mathbf{x}) = f\left(\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0\right)$$

Linear Discriminant Functions

Example: Two-class classification problem

- Input x is assigned to class C_1. if $y(\mathbf{x}) \geqslant 0$ otherwise to class C_2
- The decision boundary therefore is $y(\mathbf{x}) = 0$
- Consider x_A and x_B. Both are on the decision surface. This means

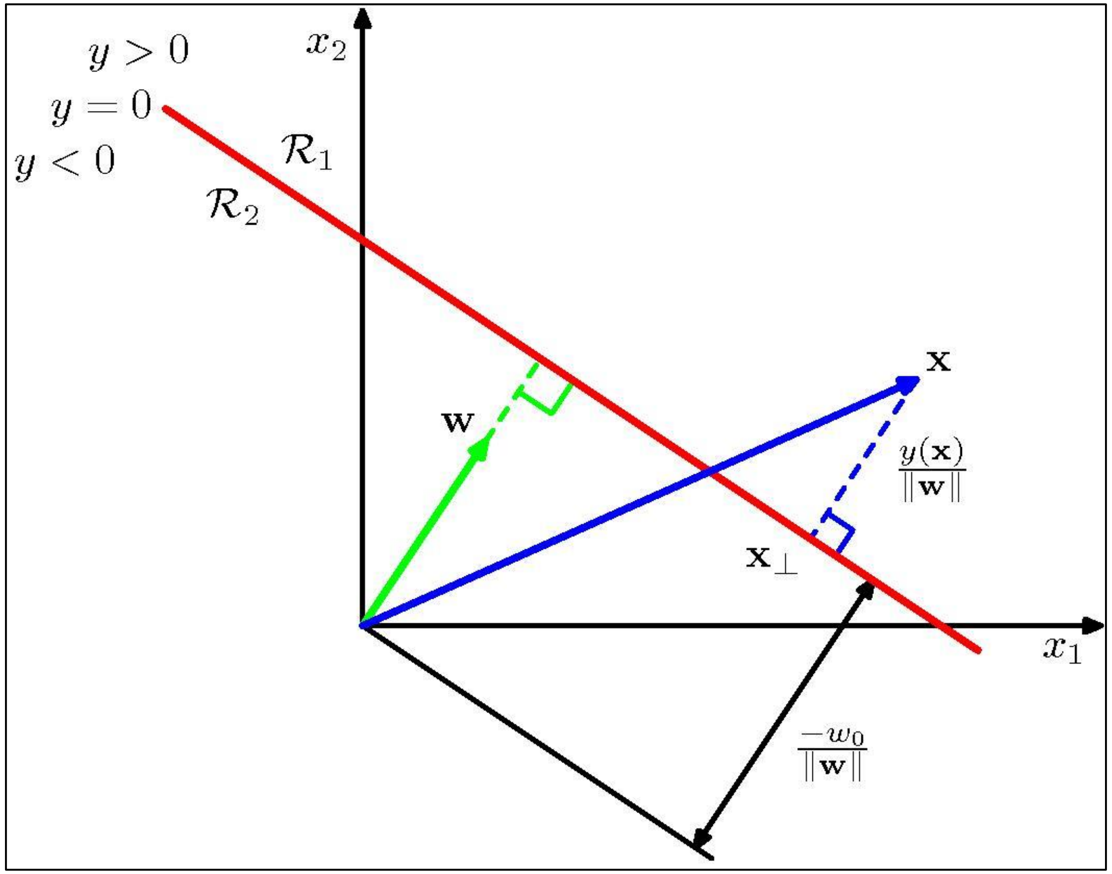
$$\mathbf{w}^{\mathrm{T}}(\mathbf{x}_{\mathrm{A}} - \mathbf{x}_{\mathrm{B}}) = 0$$

- Which in turn implies w is perpendicular to every point x which lies on the decision surface.
- Therefore w can determine the orientation of the decision surface.

Linear Discriminant Functions

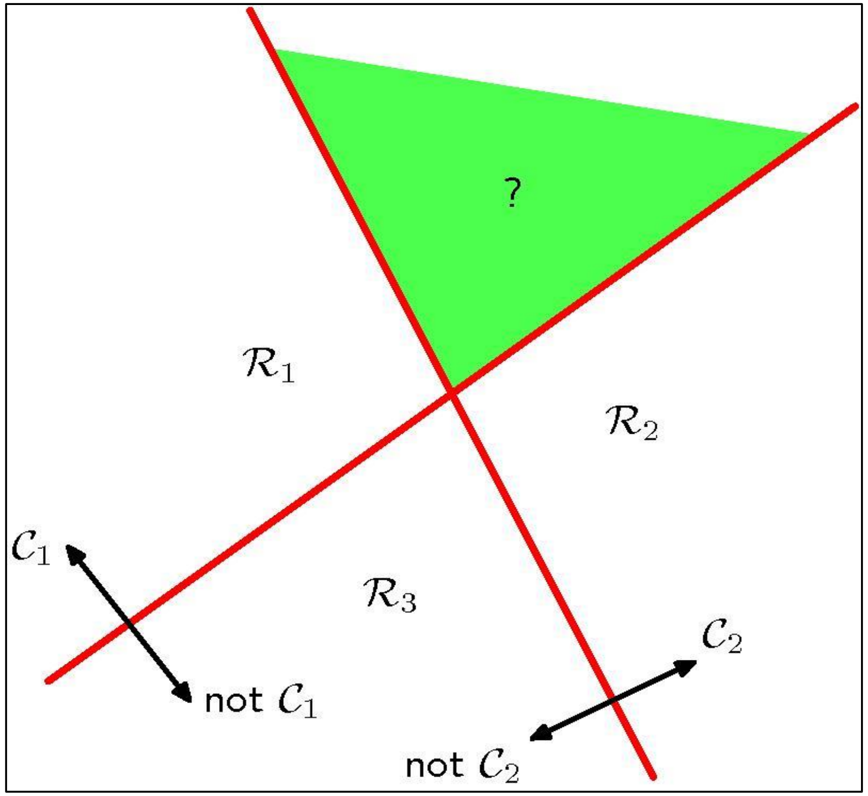
- Also the following holds true if x lies on the decision surface.

$$\frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}}{\|\mathbf{w}\|} = -\frac{w_0}{\|\mathbf{w}\|}.$$



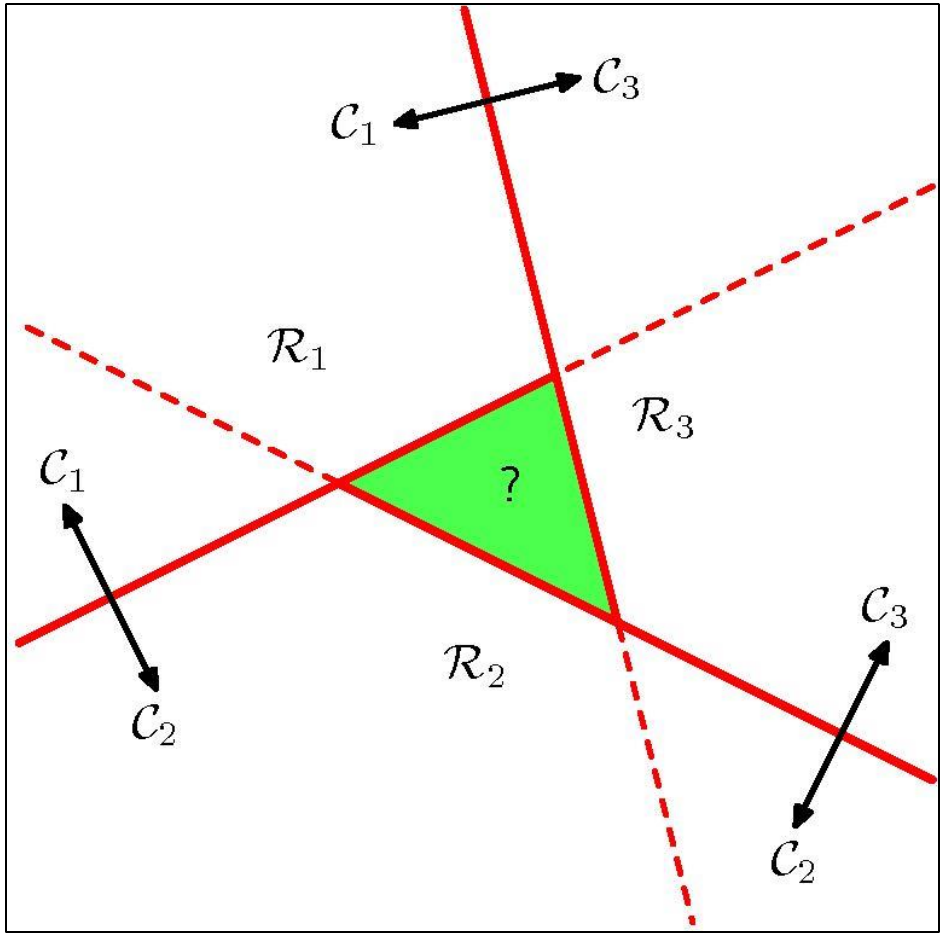
Linear Discriminant Functions

- More than two classes (say K classes)
- Multiple classes: Same approach (K-1 classifiers)
  - Each classifier acts as "One-versus-rest" classifier
  - Ambiguous regions



Linear Discriminant Functions

- Multiple classes
- Alternate approach: K(K-1)/2 classifiers
  - Binary classifiers for each pair of classes.
  - One-versus-one classifier
  - X assigned according to majority vote
  - Ambiguous regions are still present



Linear Discriminant Functions

- Multiple class problem: Alternate approach
- K class discriminant function or classifier defined a

$$y_k(\mathbf{x}) = \mathbf{w}_k^{\mathrm{T}}\mathbf{x} + w_{k0}$$

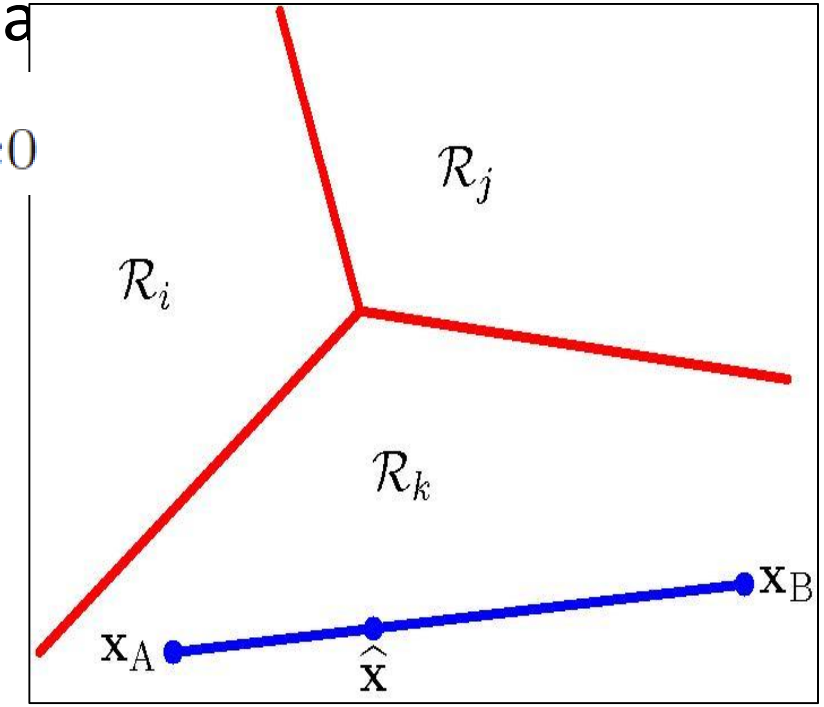
- Assign point x to class C_k if

$$y_k(\mathbf{x}) > y_j(\mathbf{x}) \text{ for all } j \neq k$$

- Decision boundary between class k and class j is

$$y_k(\mathbf{x}) = y_j(\mathbf{x})$$

$$(\mathbf{w}_k - \mathbf{w}_j)^{\mathrm{T}}\mathbf{x} + (w_{k0} - w_{j0}) = 0$$



- Singly connected and convex discriminant function

Thank You

UPES

UNIVERSITY WITH A PURPOSE