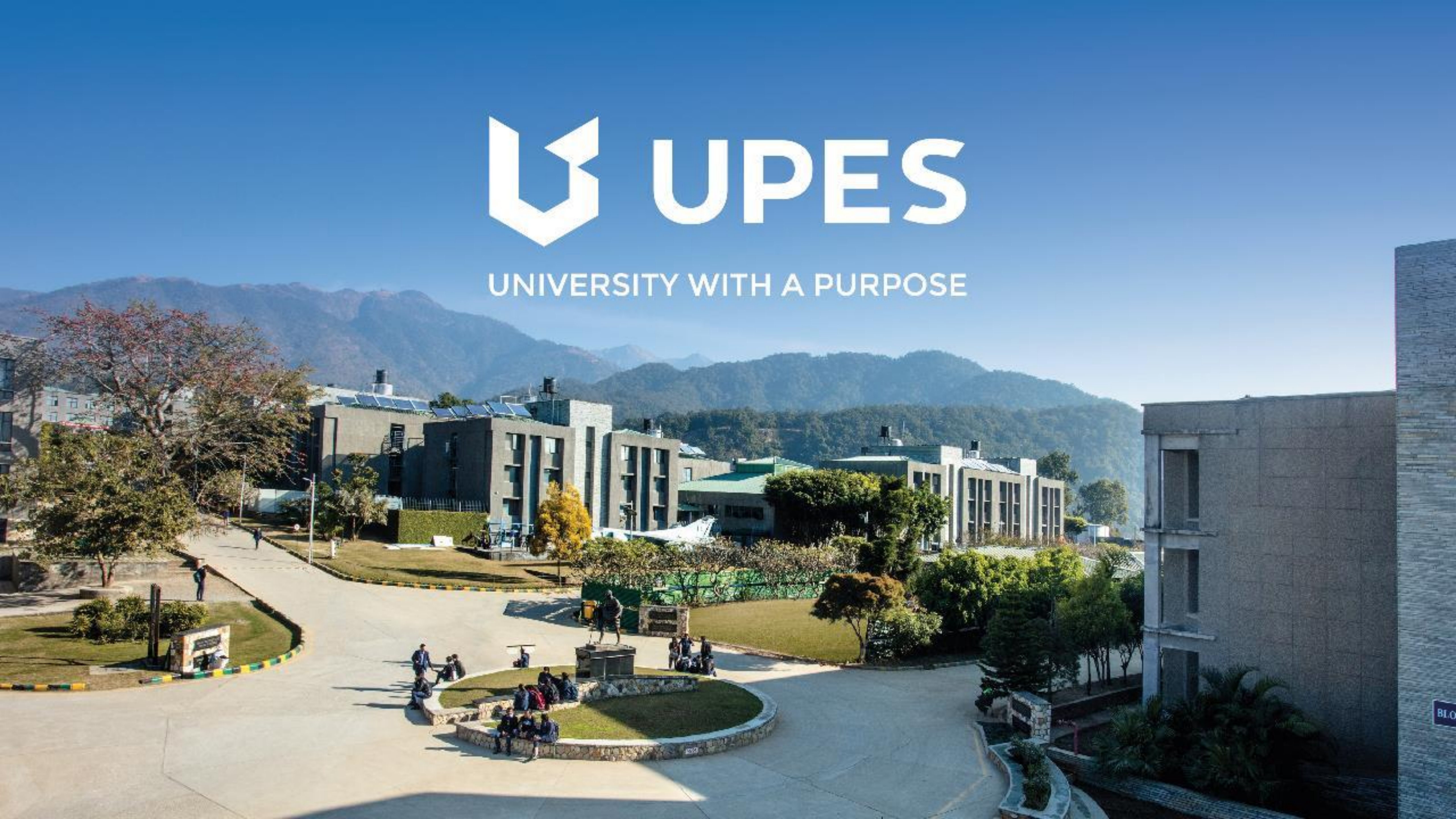
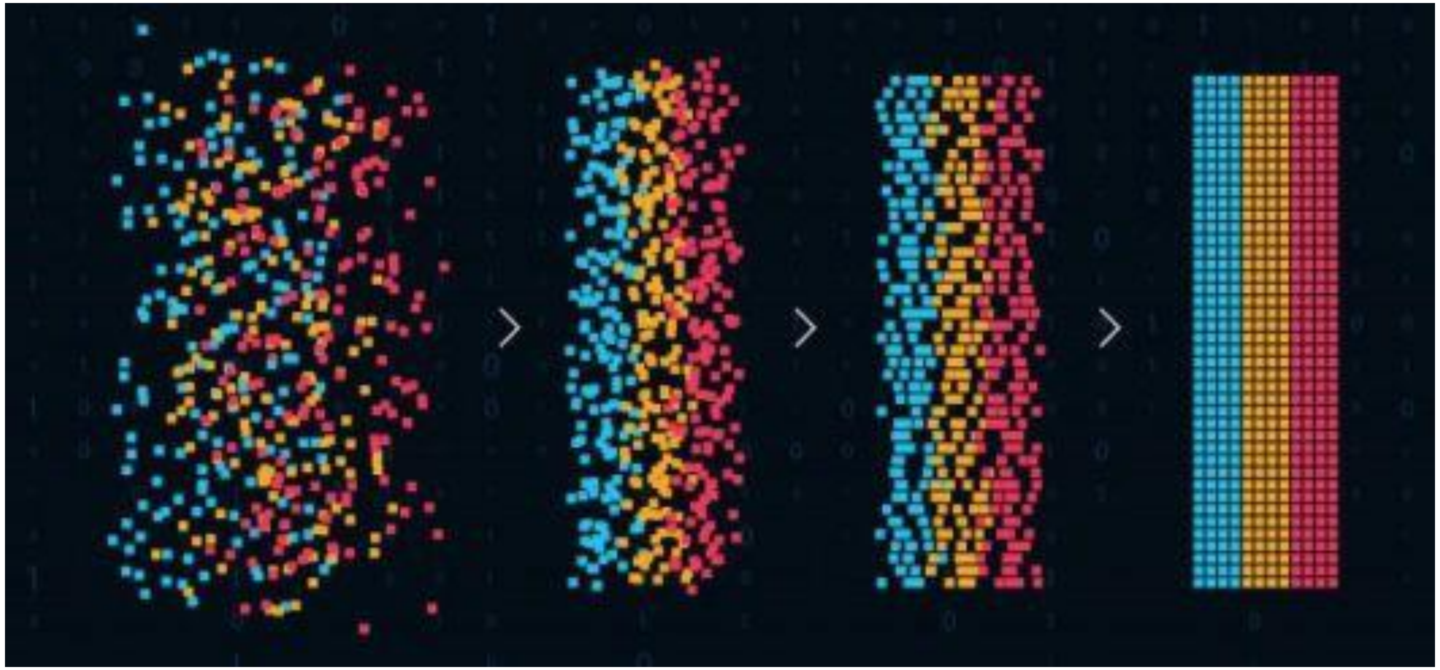

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML

Dr Gopal Singh Phartiyal

9/11/2021

Recap: Linear Models for Classification

- **Classification:** Assign input x to one of the class labels
- Why we call it linear?
- Linearly separarble dataset and Hyperplanes
- **Linear Discriminant Functions**
  - **Two class**
  - **Multiclass**
    - **3 approaches:** one-versus-the-rest, one-versus-one, and k classifiers
    - Decision region ambiguities
- **Fisher's Linear Discriminant**
  - Projection algorithm

Li-Mod for Classification: Perceptron Algorithm

- The perceptron of Rosenblatt corresponds to a two-class model in which the input vector **x** is first transformed using a fixed nonlinear transformation to give a feature vector $\phi(\mathbf{x})$
- Then construct a generalized linear model of the form

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

- Where the nonlinear activation function *f*(.) is a step function of the form

$$f(a) = \begin{cases} +1, & a \geqslant 0 \\ -1, & a < 0 \end{cases}$$

- Non-probabilistic model

Li-Mod for Classification: Perceptron Algorithm

- How to compute w?
- Standard error function doesn't work
- Alternate error function: **Perceptron criterion**
- Goal: if xn belongs to class C1 then $\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) > 0,$

iElse f xm belongs to class C2, then $\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) < 0$

- Also, $t \in \{-1, +1\}$
- In both conditions $\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)t_n > 0$ holds true.
- Therefore minimizing the error of the form would be beneficial $E_{\mathrm{P}}(\mathbf{w}) = -\sum_{n\in\mathcal{M}} \mathbf{w}^{\mathrm{T}}\phi_n t_n$

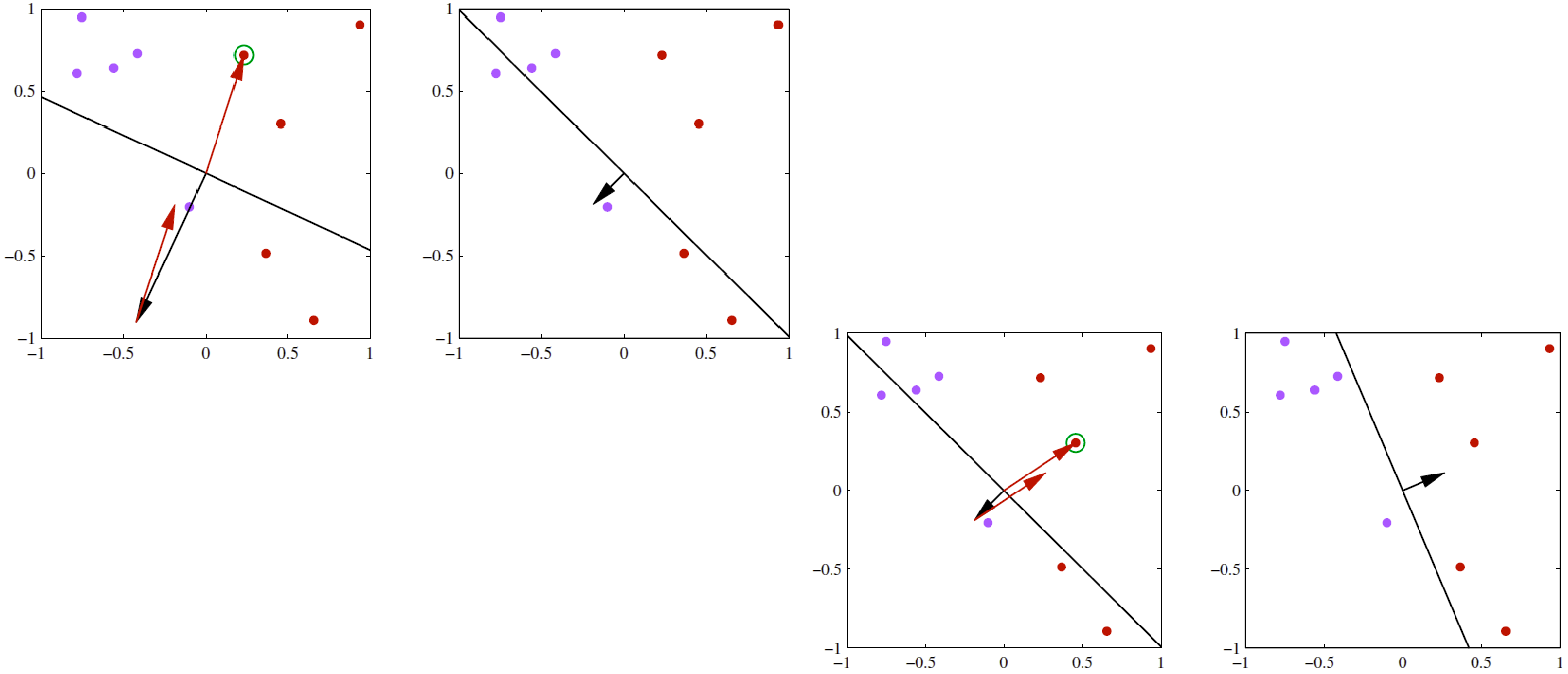
Li-Mod for Classification: Perceptron Algorithm

- How to compute w?
- Therefore minimizing the error of the form $E_{\mathrm{P}}(\mathbf{w}) = -\sum_{n \in \mathcal{M}} \mathbf{w}^{\mathrm{T}} \phi_n t_n$ to obtain w would be beneficial
- Applying stochastic gradient descent w takes

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E_{\mathrm{P}}(\mathbf{w}) = \mathbf{w}^{(\tau)} + \eta \phi_n t_n$$

- This is the perceptron learning algorithm.

Li-Mod for Classification: Perceptron Algorithm



Thank You

UPES

UNIVERSITY WITH A PURPOSE