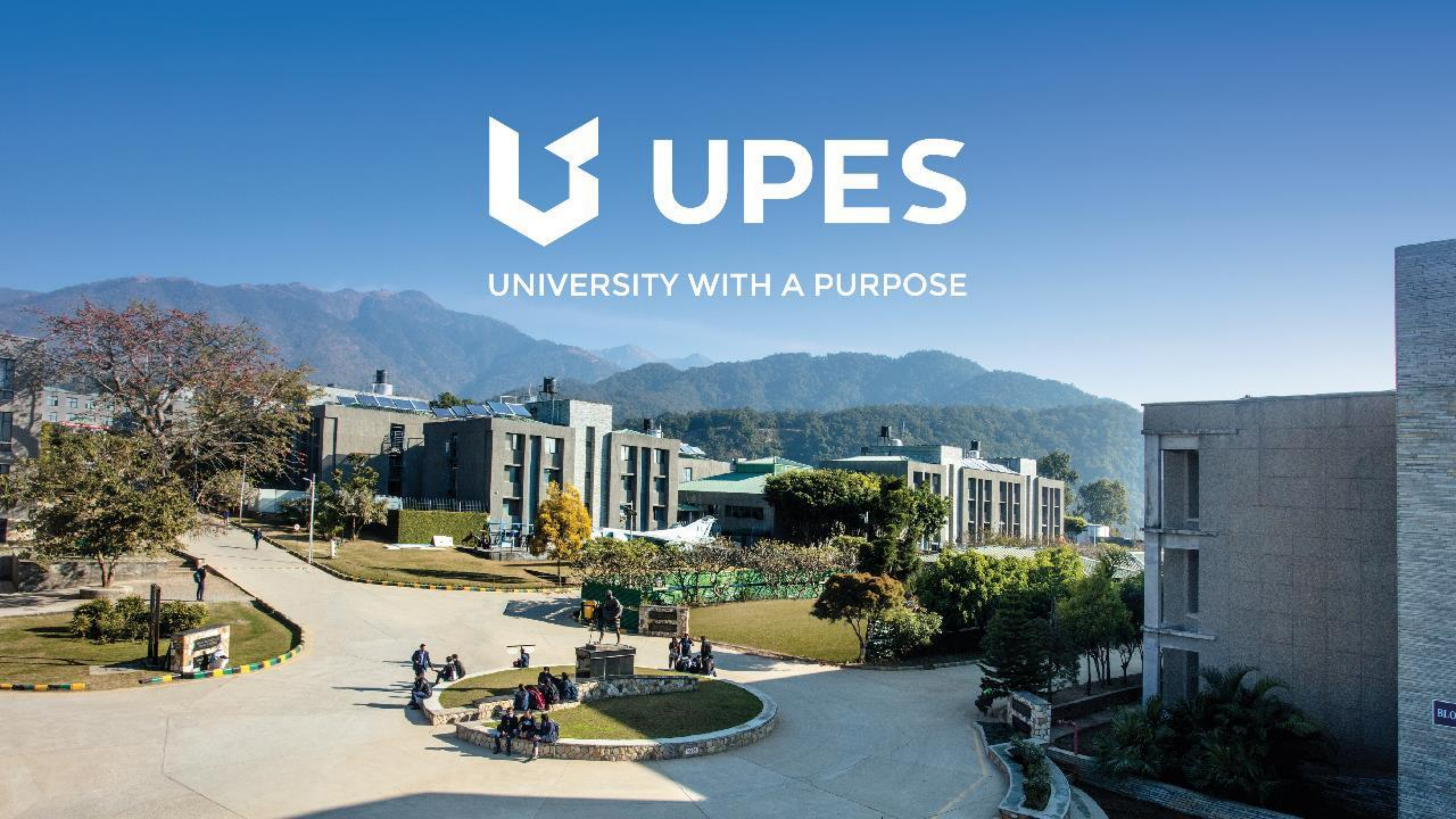
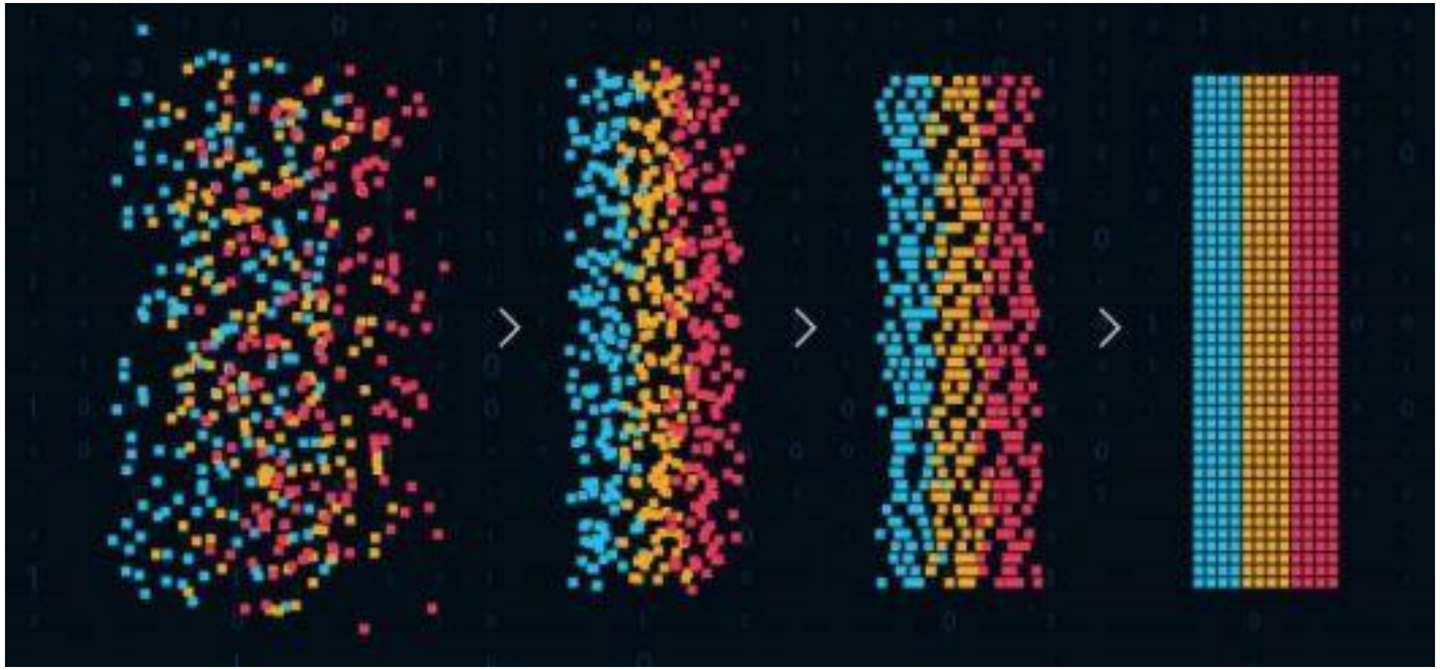

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML
Dr Gopal Singh Phartiyal
26/08/2021

Decision Theory

- Need: Decision theory establishes the fundamentals on what decision to take once ***inference*** from **data** is mode.

- Inference step
- Determine either $p(t|\mathbf{x})$ or $p(\mathbf{x}, t)$.

- Decision step
- For given X, determine optimal t.

Decision Theory

- Example:
- Goal: To tell whether a person has cancer or not.
- Input: Medical image (X-ray), a vector or matrix
- Target: Two Classes
  - Cancer/No-cancer (C_1/C_2 or True/False or 1/0)
- Inference problem
  - To compute $p(\mathbf{x},C_1)$ and further $p(\mathbf{x}, t)$
- Decision problem
  - This inferred value will decide whether to diagnose or not. Therefore, we want the decision to be to be optimal
- **In a nutshell:** how to make optimal decisions based on appropriate probabilities

Images Source: url: https://www.shutterstock.com/search/lung+cancer+x+ray

Decision Theory

- In the previous example, if we want to know the class of new patient based on his/her X-ray image. We need $p(C_k|x)$.
- According to Bayes theorem

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

$$p(C_k|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_k)p(C_k)}{p(\boldsymbol{x})}$$

- All quantities on the right can be computed from $p(\boldsymbol{x}, C_k)$ i.e. Joint probability
- Remember sum and product rule to relate Joint probability with marginal/prior probability and conditional probability

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j|X = x_i)p(X = x_i) \end{aligned}$$

Continued…

- If we wish to minimize the chance of assigning x to wrong class, then **intuitively** we will choose class having higer posterior probability.

- New goal: To make as few misclassifications as possible
- How?
- Distribute the input space into **decision regions** (R_k) such that allo x on R_k are assigned to C_k.
  - Here we have two classes, therefore two decision regions
- The boundaries between these regions is called ***decision boundaries*** or ***decision surface***.

Continued…

- Optimal decision rule via example of cancer patient
- We will try to minimize misclassification.
- Miscalssification: **x** in R_1 assigned to C_2 or **x** in R_2 assigned to C_1

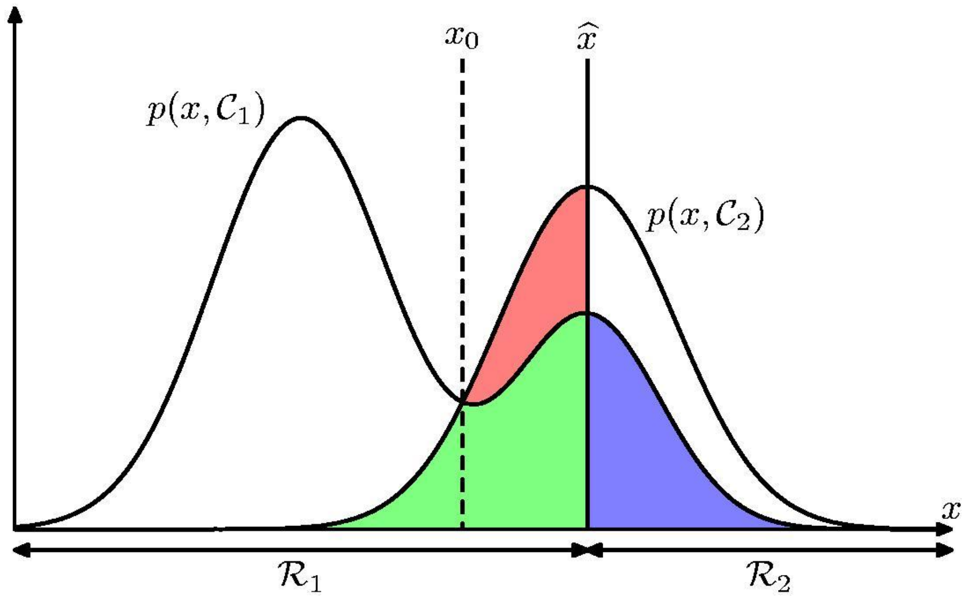
$$\begin{aligned} p(\text{mistake}) &= p(\mathbf{x} \in \mathcal{R}_1, \mathcal{C}_2) + p(\mathbf{x} \in \mathcal{R}_2, \mathcal{C}_1) \\ &= \int_{\mathcal{R}_1} p(\mathbf{x}, \mathcal{C}_2)\,\mathrm{d}\mathbf{x} + \int_{\mathcal{R}_2} p(\mathbf{x}, \mathcal{C}_1)\,\mathrm{d}\mathbf{x}. \end{aligned}$$

$$\begin{aligned} p(\text{correct}) &= \sum_{k=1}^{K} p(\mathbf{x} \in \mathcal{R}_k, \mathcal{C}_k) \\ &= \sum_{k=1}^{K} \int_{\mathcal{R}_k} p(\mathbf{x}, \mathcal{C}_k)\,\mathrm{d}\mathbf{x} \end{aligned}$$

- The p(mistake) is minimum when x is assigned to class which has the highest posterior probability
- General case: For more number of classes, it is easier to maximize the probability of correct classification minimizing the incorrect classification.

Graphically: Two classes, single variable

Schematic illustration of the joint probabilities $p(x, \mathcal{C}_k)$ for each of two classes plotted against x, together with the decision boundary $x = \widehat{x}$. Values of $x \geqslant \widehat{x}$ are classified as class $\mathcal{C}_2$ and hence belong to decision region $\mathcal{R}_2$, whereas points $x < \widehat{x}$ are classified as $\mathcal{C}_1$ and belong to $\mathcal{R}_1$. Errors arise from the blue, green, and red regions, so that for $x < \widehat{x}$ the errors are due to points from class $\mathcal{C}_2$ being misclassified as $\mathcal{C}_1$ (represented by the sum of the red and green regions), and conversely for points in the region $x \geqslant \widehat{x}$ the errors are due to points from class $\mathcal{C}_1$ being misclassified as $\mathcal{C}_2$ (represented by the blue region). As we vary the location $\widehat{x}$ of the decision boundary, the combined areas of the blue and green regions remains constant, whereas the size of the red region varies. The optimal choice for $\widehat{x}$ is where the curves for $p(x, \mathcal{C}_1)$ and $p(x, \mathcal{C}_2)$ cross, corresponding to $\widehat{x} = x_0$, because in this case the red region disappears. This is equivalent to the minimum misclassification rate decision rule, which assigns each value of x to the class having the higher posterior probability $p(\mathcal{C}_k|x)$.



Minimizing the Loss

- This time, simply minimizing the number of misclassification will not work.
- We have to prioritize misclassifications.
- Example: Diagnosing cancer patient
- Misclassifications
  - No cancer actually but diagnosed with cancer treatment as per model
  - Had cancer actually but diagnosed as healthy
- Are these two misclassifications of equal importance?????
- We need other parameter to measure the impact of misclassifications and then minimize it. ***Loss function***

Minimizing the Loss

- ***Expected loss***

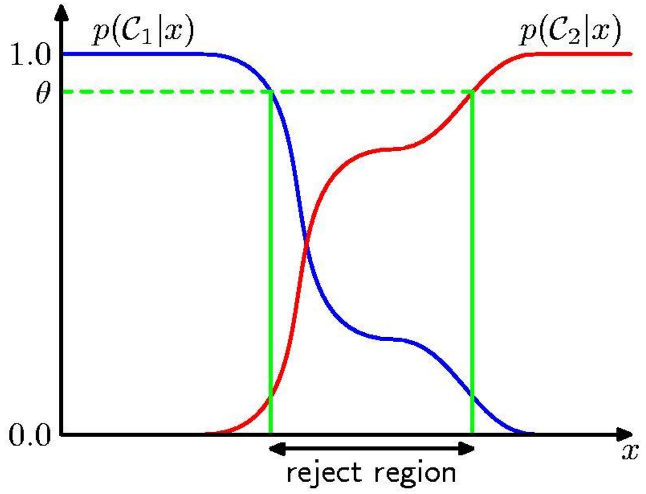
$$\mathbb{E}[L] = \sum_k \sum_j \int_{\mathcal{R}_j} L_{kj} p(\mathbf{x}, \mathcal{C}_k)\,\mathrm{d}\mathbf{x}$$

$$\mathbb{E}[L] = \sum_k L_{kj} p(\mathcal{C}_k|\mathbf{x})$$

Decision

$$\begin{array}{ll} & \begin{array}{cc} \text{cancer} & \text{normal} \end{array} \\ \text{Truth} \begin{array}{l} \text{cancer} \\ \text{normal} \end{array} & \begin{pmatrix} 0 & 1000 \\ 1 & 0 \end{pmatrix} \end{array}$$

Reject option



Like to combine Inference and Decision ?

- ***Inference stage:*** *Computing posterior probabilities via models learning*
- ***Decision stage**: Making decisions or class assignments*

- *Why not do it in one go (combine the stages): one function doing both the stages*
  - *Discriminant function does that*

- *The advantage of keeping inference and decision stages separate are*
  - Minimizing risk (loss matrix may change over time)
  - Reject option
  - Unbalanced class priors
  - Combining models

Types of approaches to solve decision problems

- ***Generative models:*** *First model joint distributions, then decision*
- ***Discriminative models:*** *First model class posterior probabilities then decision*
- ***Discriminant functions:*** *Directly map input to output (probabilities play no role)*

Next time:

Thank You

UPES

UNIVERSITY WITH A PURPOSE