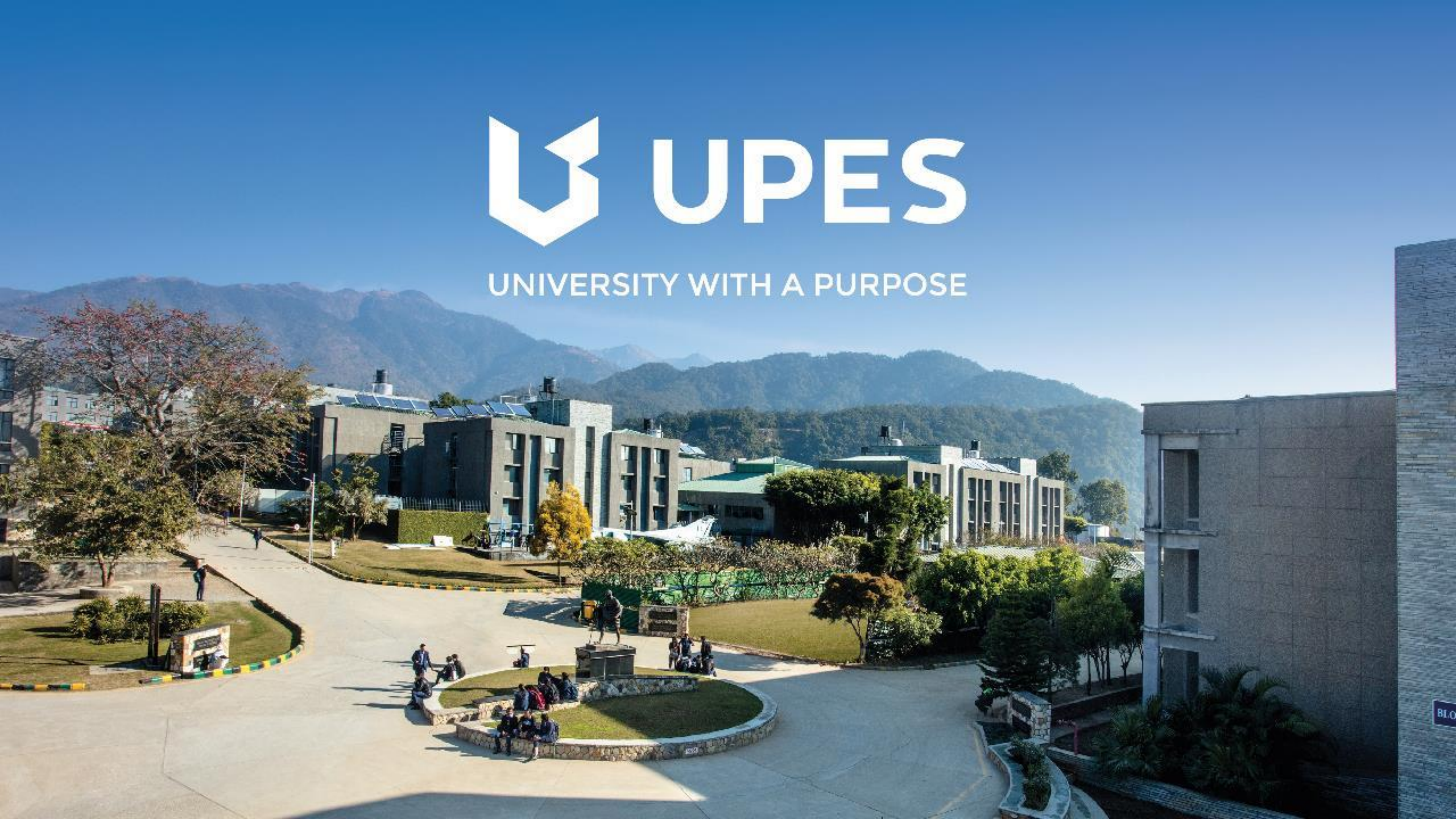
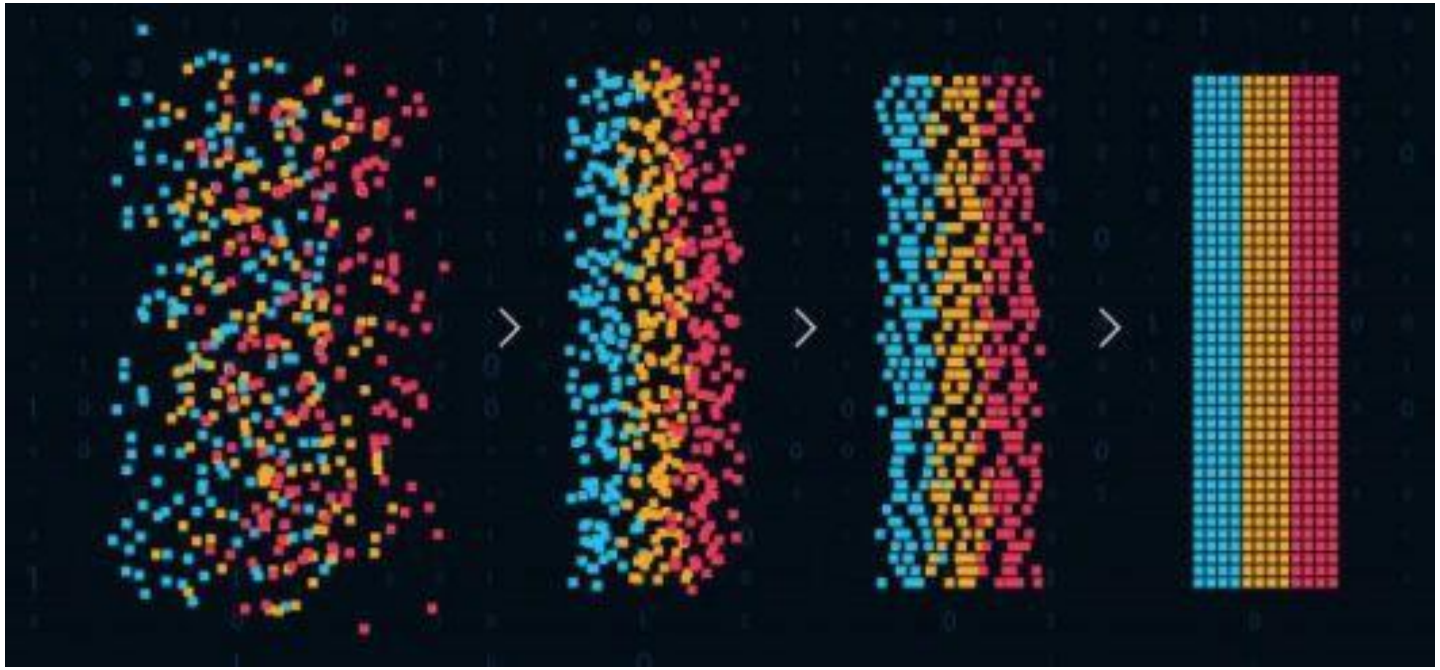

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML

Dr Gopal Singh Phartiyal

26/08/2021

Information Theory

- Consider a discrete random variable *x.*
- ***Question****: How much* information is received when we observe a specific value for this variable x.
- Degree of ***surprise:*** On learning the value of x.

- If we are told that a highly improbable event has just occurred, we will have received more information than if we were told that some very likely event has just occurred, and if we knew that the event was certain to happen we would receive no information.

- measure of information content will therefore depend on the probability distribution $p(x)$, and we therefore look for a quantity $h(x)$ that is a monotonic function of the probability $p(x)$ and that expresses the information content

Mathematical formulation

- Consider 2 events x and y that are unrelated

- Net information gain in overserving the two together = sum of information gained on observing them separately

$$h(x, y) = h(x) + h(y).$$

- The two events are statistically independent as well.

$$p(x, y) = p(x)p(y)$$

- Negative sign
- Base 2
- Unit of information = **bits**

$$\boldsymbol{h(x) = -\log_2 p(x)}$$

Mathematical formulation

- The average amount of information transferred between sender and receiver is = expectation of

$$h(x) = -\log_2 p(x)$$

$$\mathrm{H}[x] = -\sum_x p(x) \log_2 p(x)$$ ***Entropy !!!***

- Important quantity in
  - coding theory
  - statistical physics
  - machine learning

Example

- Consider a random variable x having 8 possible states, each of which is equally likely.

- In order to communicate the value of x to a receiver, we would need to transmit a message of length 3.

$$H[x] = ??$$

- Change probabilities

$$(1/2 , 1/4 , 1/8 , 1/16 , 1/64 , 1/64 , 1/64 , 1/64)$$

- New $H[x]$??

Example

x	a	b	c	d	e	f	g	h
$p(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$
code	0	10	110	1110	111100	111101	111110	111111

- New H[*x*] ??

$$\begin{aligned} \mathrm{H}[x] &= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{16}\log_2\frac{1}{16} - \frac{4}{64}\log_2\frac{1}{64} \\ &= 2 \text{ bits} \end{aligned}$$

$$\begin{aligned} \text{average code length} &= \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{16}\times 4 + 4\times\frac{1}{64}\times 6 \\ &= 2 \text{ bits} \end{aligned}$$

Another Example

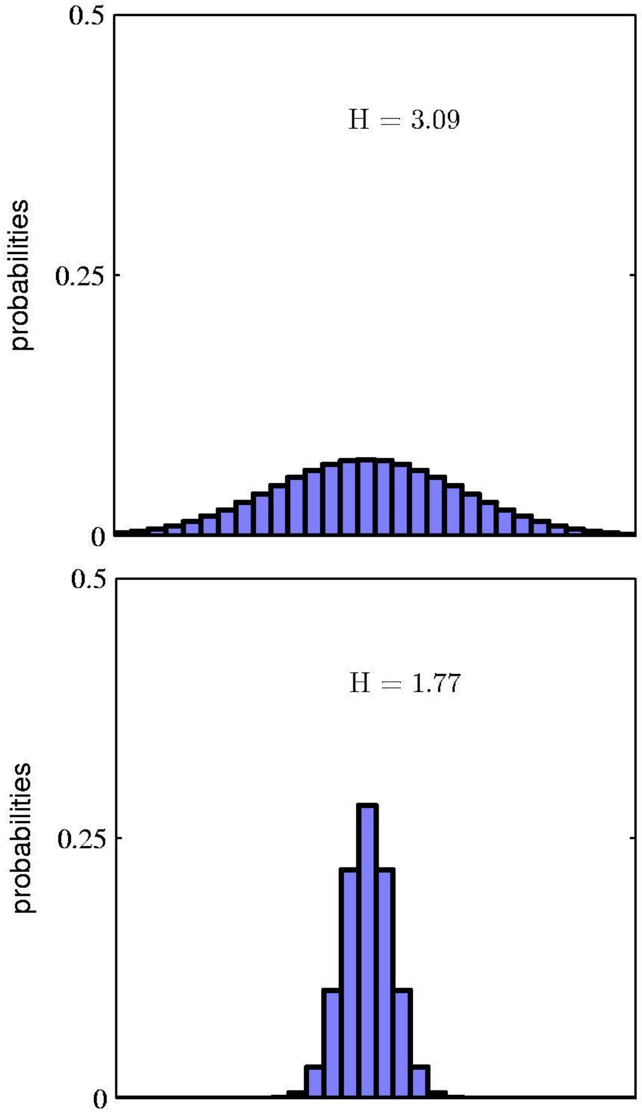
- In how many ways can N identical objects be allocated M bins?

$$W = \frac{N!}{\prod_i n_i!}$$

$$\mathrm{H} = \frac{1}{N}\ln W \simeq -\lim_{N\to\infty}\sum_i \left(\frac{n_i}{N}\right)\ln\left(\frac{n_i}{N}\right) = -\sum_i p_i \ln p_i$$

- Entropy maximized when

$$\forall i : p_i = \frac{1}{M}$$



Important formulas

- Differential Entropy

$$\mathrm{H}[x] = \frac{1}{2}\left\{1 + \ln(2\pi\sigma^2)\right\}.$$

- Conditional Entropy

$$\mathrm{H}[\mathbf{y}|\mathbf{x}] = -\iint p(\mathbf{y},\mathbf{x}) \ln p(\mathbf{y}|\mathbf{x})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x}$$

$$\mathrm{H}[\mathbf{x},\mathbf{y}] = \mathrm{H}[\mathbf{y}|\mathbf{x}] + \mathrm{H}[\mathbf{x}]$$

- Kullback-Leibler Divergence

$$\mathrm{KL}(p\|q) \simeq \frac{1}{N}\sum_{n=1}^{N}\left\{-\ln q(\mathbf{x}_n|\boldsymbol{\theta}) + \ln p(\mathbf{x}_n)\right\}$$

$$\mathrm{KL}(p\|q) \geqslant 0 \qquad \mathrm{KL}(p\|q) \not\equiv \mathrm{KL}(q\|p)$$

- Mutual Information

$$\mathrm{I}[\mathbf{x},\mathbf{y}] = \mathrm{H}[\mathbf{x}] - \mathrm{H}[\mathbf{x}|\mathbf{y}] = \mathrm{H}[\mathbf{y}] - \mathrm{H}[\mathbf{y}|\mathbf{x}]$$

Next time:

Thank You

UPES

UNIVERSITY WITH A PURPOSE