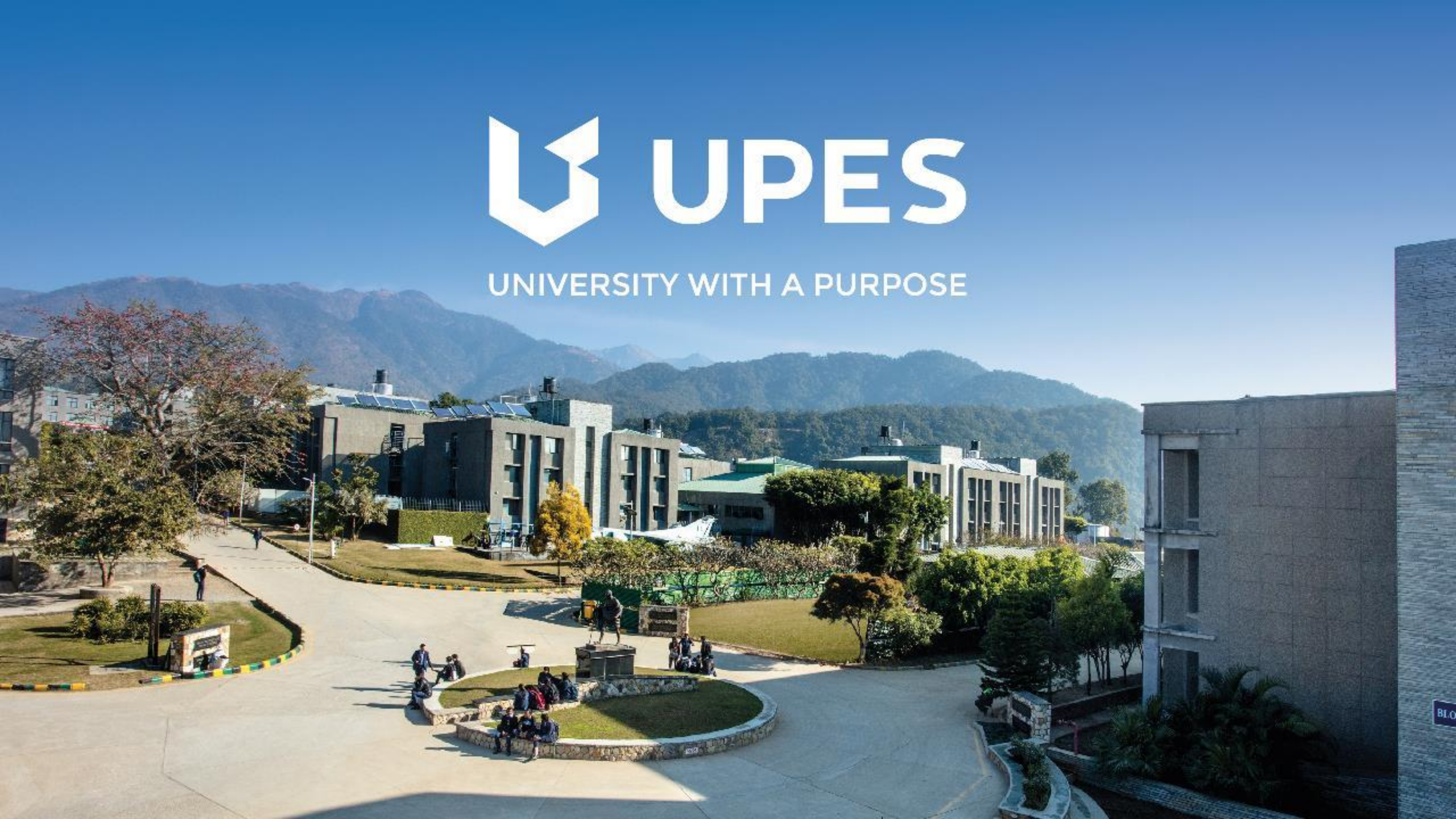
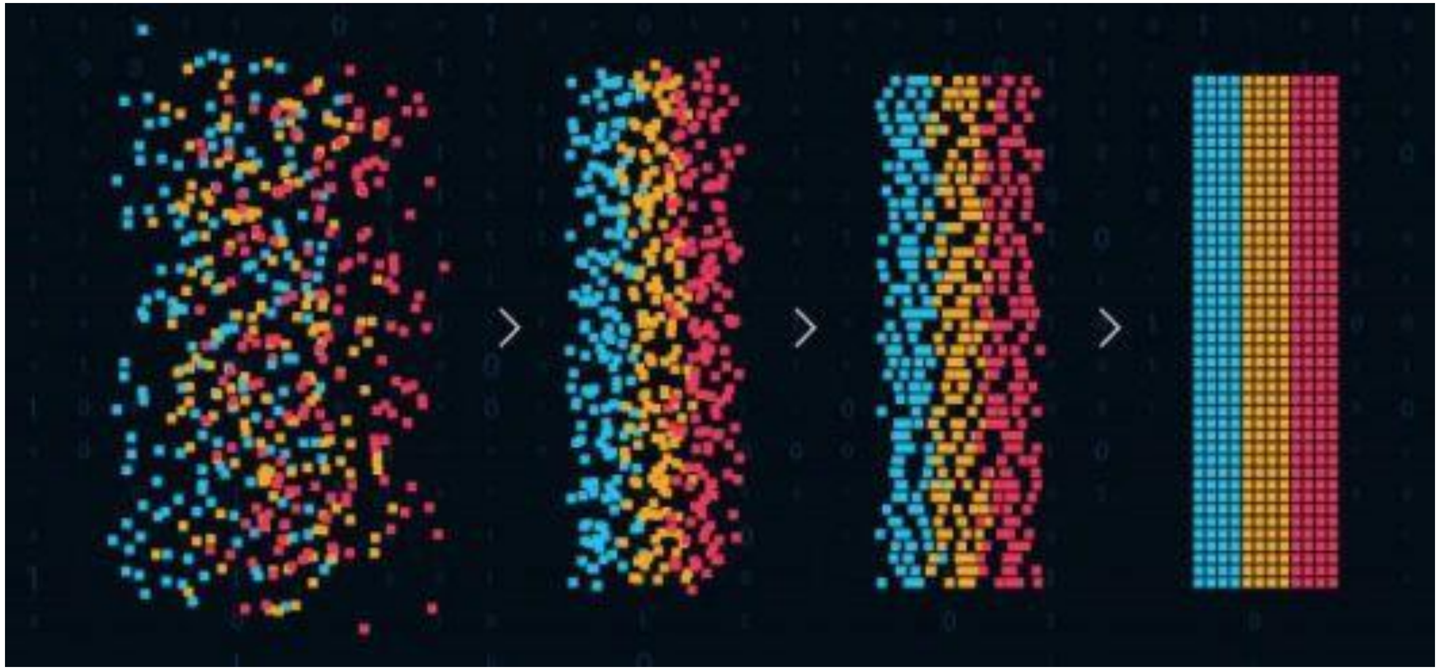

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML

Dr Gopal Singh Phartiyal

16/09/2021

Recap: Types

- Discrete random variables: a countable number of possible values.
- continuous random variables: Takes all values in a given interval of numbers.
- Parametric distributions: Governed by adaptive parameters (mean, variance etc.).
  - Binomial and Multinomial distribution for discrete random variables
  - Gaussian distribution for continuous random variables
- In order to apply these distributions for density estimation, need to determine suitable value of these parameters given an observed data set
  - If frequenist treatment: determine parameters by optimizing likelihood function
  - In Bayesian treatment: posterior probabilities of parameters from prior probabilities of parameters conditioned to observed data

Binary Variables: Beta Distribution

- Overfitting with Binomial (frequist treatment)
- Let's go for Bayesian treatment
- We need prior distribution in the form proportional to μ and $(1-\mu)$,
- So that posterior should be a product of factors of form $\mu^x(1-\mu)^{1-x}$.

- Therefore we choose Beta distribution as prior distribution
- It will assure ***conjugacy***

Binary Variables: Beta Distribution

- Distribution over

$$\mu \in [0, 1]$$

$$\int_0^1 \mathrm{Beta}(\mu|a, b)\, \mathrm{d}\mu = 1.$$

$$\begin{aligned}
\mathrm{Beta}(\mu|a,b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1} \\
\mathbb{E}[\mu] &= \frac{a}{a+b} \\
\mathrm{var}[\mu] &= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

$$\Gamma(x) \equiv \int_0^\infty u^{x-1} e^{-u}\, \mathrm{d}u.$$

Binary Variables: Posterior

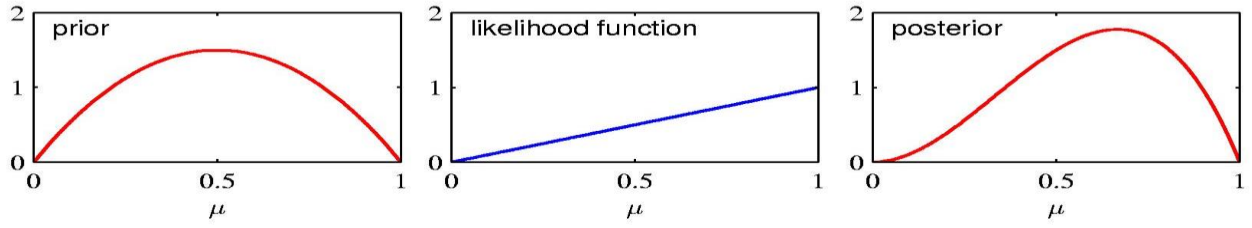
Posterior = likelihood*prior

$$\begin{aligned} p(\mu|a_0,b_0,\mathcal{D}) &\propto p(\mathcal{D}|\mu)p(\mu|a_0,b_0) \\ &= \left(\prod_{n=1}^{N}\mu^{x_n}(1-\mu)^{1-x_n}\right)\mathrm{Beta}(\mu|a_0,b_0) \\ &\propto \mu^{m+a_0-1}(1-\mu)^{(N-m)+b_0-1} \\ &\propto \mathrm{Beta}(\mu|a_N,b_N) \end{aligned}$$

- Proportionality analogous to prior
- Posterior

$$\boxed{p(\mu|m,l,a,b) = \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)}\mu^{m+a-1}(1-\mu)^{l+b-1}.}$$

- Effective number of observations
- Sequential method



Binary Variables: Posterior: Properties

As the size of the data set, N, increase

$$
\begin{aligned}
a_N &\rightarrow m \\
b_N &\rightarrow N - m \\
\mathbb{E}[\mu] &= \frac{a_N}{a_N + b_N} \rightarrow \frac{m}{N} = \mu_{\mathrm{ML}} \\
\operatorname{var}[\mu] &= \frac{a_N b_N}{(a_N + b_N)^2 (a_N + b_N + 1)} \rightarrow 0
\end{aligned}
$$

Posterior varies between prior and maximum likelihood

Binary Variables:Prediction under the Posterior

What is the probability that the next coin toss will land heads up?

$$
\begin{aligned}
p(x=1|a_0,b_0,\mathcal{D}) &= \int_0^1 p(x=1|\mu)p(\mu|a_0,b_0,\mathcal{D})\,\mathrm{d}\mu \\
&= \int_0^1 \mu p(\mu|a_0,b_0,\mathcal{D})\,\mathrm{d}\mu \\
&= \mathbb{E}[\mu|a_0,b_0,\mathcal{D}] = \frac{a_N}{b_N}
\end{aligned}
$$

Next time: Multinomial Distribution

Thank You

UPES
UNIVERSITY WITH A PURPOSE

Multinomial Variables

Variable with K states

- 1-of-K coding scheme:

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

$$p(\mathbf{x}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

$$\forall k : \mu_k \geqslant 0 \quad \text{and} \quad \sum_{k=1}^{K} \mu_k = 1$$

$$\mathbb{E}[\mathbf{x}|\boldsymbol{\mu}] = \sum_{\mathbf{x}} p(\mathbf{x}|\boldsymbol{\mu})\mathbf{x} = (\mu_1, \ldots, \mu_K)^{\mathrm{T}} = \boldsymbol{\mu}$$

$$\sum_{\mathbf{x}} p(\mathbf{x}|\boldsymbol{\mu}) = \sum_{k=1}^{K} \mu_k = 1$$

Multinomial Variables