Unsupervised learning → clustering
→ density estimation

- Parametric Ulearning
  1. construct gaussian mixture models
  2. Use expectation maximization algo
- non-parametric

Clustering as a mixture of gaussians
"parametric (vs) non-parametric distribution".

Different statistical distributions : (next page)

*If data is ordinal / interval based, only non-parametric statistics can be used

- Data
  - quantitative
    - discrete
    - continuous
      - Interval scale
      - Ration scale
  - qualitative
    - Non-numerical
      - Nominal
      - Ordinal (order present)

Sequential learning
mapping input seq to output seq using state machines. Hidden state seq present.

Active learning.

Theory of rational agency: (action selection theories)

★ Density estimation ((how using deep generative NNs?))
→ Estimating probability density function of random variable in a population from sample's help.

q: Difference between probability density fn & probability distribution?
fn

q: what is maximum likelihood estimation?
→ finding the values of parameters that result in best fit curve.

* likelihood & loglikelihood

$$L(\mu, \sigma; \text{data}) = P(\text{data}; \mu, \sigma)$$

q when is least squares minimization same as max likelihood estimation? why does it happen in that case?
& How

Bayesian Inference for parameter estimation:

Bayes theorem: $P(\theta/\text{data}) = \frac{P(\text{data}/\theta) \times P(\theta)}{P(\text{data})}$

★ Parameter estimation using Bayes theorem
"prior distribution".
θ → set of parameters ($\theta = \{\mu, \sigma\}$ for gaussian distribution)
$P(\theta/\text{data})$ → posterior distribution
$P(\theta)$ → prior distribution
$P(\text{data})$ → evidence & data = $\{y_1, y_2, \cdots, y_n\}$

Can we use bayesian inference for classification problems? How? Is it used for discrete data / continuous or both?