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: $P(\theta/\text{data}) = \dfrac{P(\text{data}/\theta) \times P(\theta)}{P(\text{data})}$

* 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 $= \{y_1, y_2, \cdots, y_n\}$
↳ normalizing const (helps making $\Sigma P(\theta/\text{data}) = 1$)

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

Different statistics from the posterior distribution & their physical significance!
→ expected value = mean
→ variance ⇒ uncertainity
→ mode ≠ MAP estimate.

"gaussian distribution is conjugate to itself wrt gaussian likelihood function."

(Latent Dirichlet Allocation algo) ✱

Markov Chain Monte Carlo methods → to Calculate posterior distribution

✱✱ Updating beliefs iteratively in real time. using bayesian inference → kalman filter.

Prior acts as a regularizer ~~here~~ in bayesian inference.

q when does MAP estimate equals MLE?
Overfitting due to Bayesian priors ⟹ Pending

Marginalization ↴

$$P(X) = \int_y P(X, Y=y)\, dy.$$

What is discriminant analysis?
When is it used?
↓
when dependant variable is categorical & predictor/independant variable is interval.
↓
develop discriminant fn as a linear combination of independant variables to descriminate between categories of dependant variable.

Discriminant analysis (vs) Analysis of variance.
(vs) regression analysis
Correlation is not causation ✱

probability distribution

continuous
eg: Normal
Standard Normal

discrete
eg: Binomial
Poisson

MCMC methods

→ monte carlo simulations
→ markov chains (are memoryless)
* bell curve, Law of large nos.
Markov → Non independant events may also conform to patterns
(In long run, dist settle to pattern)
* if events are subj to fixed probs.,
interdependant events conform to average.

q. How can bayesian inference be used to quantify uncertainity in predictions?

MCMC → Random samplig of parameters in probabilistic space to approximate the posterior distribution in Bayesian inference.

~~Where~~ can we use these posterior distributions?
~~How~~
→ quantifying uncertainity
→ comparing models
→ generating predictions

Central limit theorem & law of large nos
Covariance vs correlation vs causation

Statistical distributions

q: data discrete/continuous?
q. symmetry of data & outliers scenario.
q. upper & lower limits of data
q. likelihood of occurence of extreme values

Discrete distributions:	Continuous
→ Binomial	→ Normal
→ Poisson	→ Exponential
→ Negative binomial	→ logistic cauchy
→ geometric	→ gamma
→ Bernoulli	→ chisquared

q. lets say we have a distribution which is generated by combining multiple commonly known distributions. How do we find those & seperate those distributions?

Joint distributions:

Discriminant analysis:

Exponential regression: $y = \alpha e^{\beta x}$ → Can be converted to linear
power regression: $y = \alpha x^{\beta}$

confidence interval (vs) prediction interval (vs) tolerance interval

q. Finding confidence interval ~~using~~ not using different ~~means~~ distributions.
q. How to get confidence interval for mean given we have a sample?