- Unsupervised learning
  - Clustering
  - density estimation

- Unsupervised learning
  - 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~~-numerial
    - 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 NN's?))
  → 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

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

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

Bayesian Inference for parameter estimation:

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

Parameter estimation using Bayes theorm

"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 $= \{y_1, y_2, \dots, y_n\}$
  ↳ 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 $\Rightarrow$ mean
- → variance $\Rightarrow$ uncertainity
- → mode $\neq$ 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 discriminate 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 prob.,
interdependant events conform to average.

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

MCMC → Random sampling of parameters in probabilistic space to approximate the posterior distribution in bayesian inference.

~~Where~~ How can we use these posterior distributions?

→ 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:
→ Binomial
→ Poisson
→ Negative binomial
→ geometric
→ Bernoulli

Continuous
→ Normal
→ Exponential
→ logistic cauchy
→ gamma
→ 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

quantifies uncertainity of estimated skill of model

quantifies uncertainity in single forecast

q. Finding confidence interval ~~using~~ using different ~~means~~. distribution

q. How to get confidence interval for mean given we have a sample?

Methods to get prediction intervals:
→ bootstrap resampling
→ delta method
→ bayesian method
→ mean-variance estimation method

Resampling methods:
- → cross validation → cluster based oversampling
- → bootstraping → random under sampling
- Random oversampling
- SMOTE

Probability mass fn (vs) probability density fn

Hypothesis testing:
- → null hypothesis
- → p value & critical value
- → traditional testing (vs) bayesian testing

Marginal distributions

X, Y are jointly distributed random variables

Entropy → list of generalized entropies
most widely used → shannon's entropy

$$\phi_1(P) = -\sum_{i=1}^{n} P_i \log P_i$$

Decision tree → info gain → $\phi_{parent} - \sum w_i \times \phi_{child}$
$\quad w_i = \frac{n_{child}}{n_{parent}}$

→ gini index → $1 - \sum_i p_i^2$

Hinge loss → SVM

grid search → best combination of hyperparameters for best fit
→ random search

Hyperparameter optimization

Skewness, kurtosis, coefficient of variation

Data imputation methods:
- → for cross-sectional datasets
- → for time series data

Cases:
- → MCAR
- → MAR
- → NMAR

① Using Mean/Median values :
→ poor results on encoded categorical features

② Using Mode
→ works with categorical features
→ can introduce bias

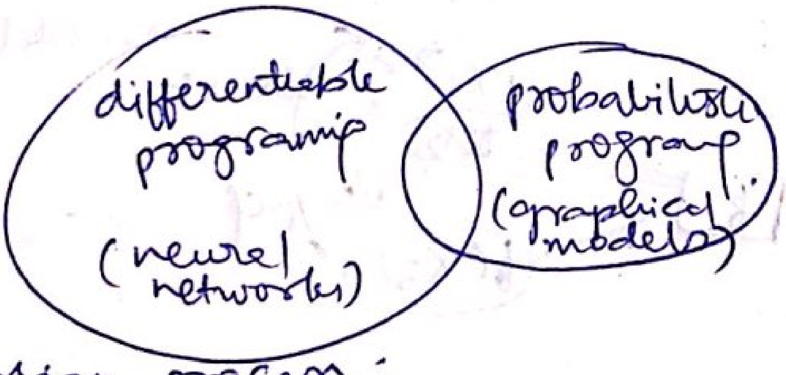
③ Using KNN algo.
→ impyute library
KDTree
KNN sensitive to outliers unlike SVM.

④ using multivariate imputation by chained equ.
→ Multiple imputations

⑤ Using DL.

⑥ using stochastic regression/extrapolation & intrapolation

Uncertainity quantification
→ gaussian process (GP) models.
(multivariate problems)



gaussian process:
→ use prior knowledg to make predictions.
→ assign probability to diff fns possible for fitt a dataset, getting mean of prob dist to get most probable ans.
→ incorporate confidence.

multivariate gaussian distribution :
→ each random variable → normal distributed
→ their joint dist → also normal
→ μ → mean
Σ → covariance matrix (symmetric & +ve semidefinite)
↳ gives σ_i^2 & σ_{ij}

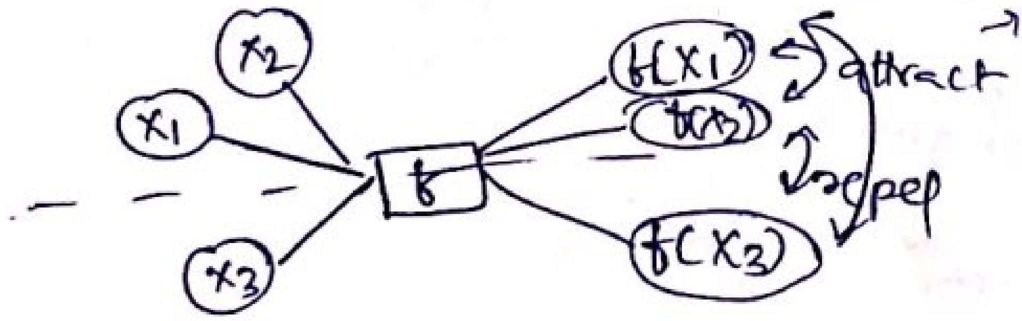
Semi supervised learning → model trained on dataset → small portion labelled data
↳ majority unlabelled data.

step 1 : cluster ~~similar~~ data into grps of similar data (unsup part)
step 2 : labelled remaing data in grp seeing the labelled data in the same grp.

manifold assumption →

contrastive learning:

adversial ML
→ data poisoning
→ evasion attack
→ model extraction



Few short learning :
↳ meta learning problem

data level approach
→ GAN's
→ data augmentation
→ using base datasets

parameter level approach
↳ limit parameter space