

Unsupervised learning

Parametric
Ulearnip

non-
parametric

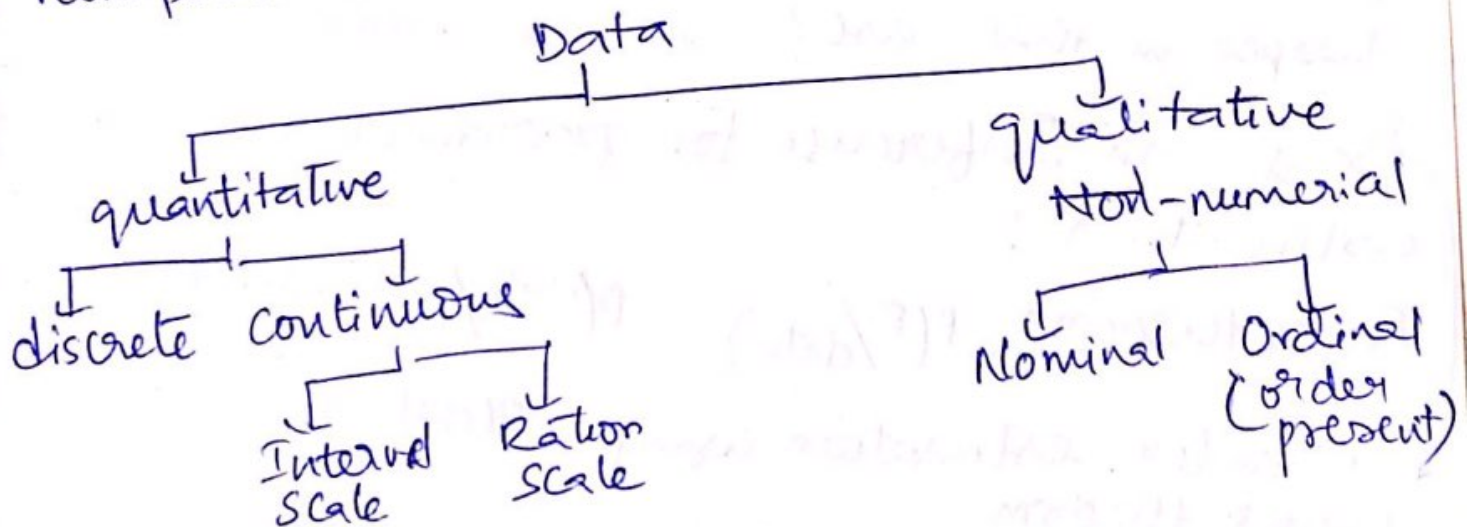
① construct
gaussian mixture
models

② Use expectation
maximization algo

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



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 NMs?))
→ estimating probability density function of random variable in a population from sample's help.

Q: Difference between probability density fn & probability distribution?
+n

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?
+n

Bayesian Inference for parameter estimation:

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

★ Parameter estimation using Bayes theorem

"prior distribution".

$\theta \rightarrow$ set of parameters ($\theta = \{\mu, \sigma\}$ for gaussian distribution)

$P(\theta/\text{data}) \rightarrow$ posterior distribution

$P(\theta) \rightarrow$ prior distribution

$P(\text{data}) \rightarrow$ evidence & data = $\{y_1, y_2, \dots, y_n\}$

\hookrightarrow normalizing const (helps making $\sum P(\theta/\text{data}) = 1$)

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 uncertainty
- mode \neq MAP estimate.

"gaussian distribution is conjugate to itself w.r.t gaussian likelihood function."

(Latent Dirichlet Allocation algo) *

Markov Chain Monte Carlo methods \rightarrow to calculate posterior distribution

*** Updating beliefs iteratively in real time. using bayesian inference \rightarrow kalman filter.

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

Q when does MAP estimate equals MLE?

Overfitting due to Bayesian priors \rightarrow pending

Marginalization \downarrow

$$P(x) = \int_y P(x, y) dy.$$

What is discriminant analysis?

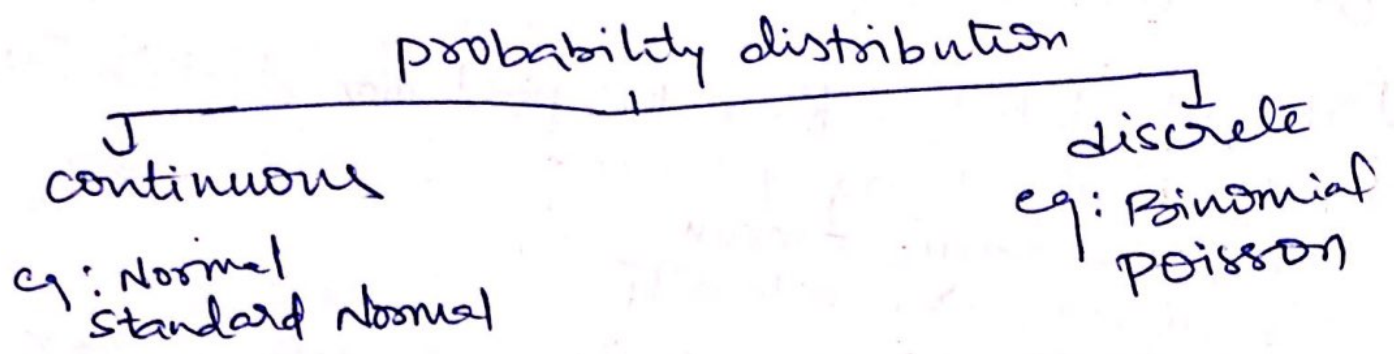
When is it used?

\downarrow
when dependant variable is categorical & predictor/independant variable is interval.

\downarrow
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. #



MCMC methods

→ monte carlo simulations

→ markov chains (are memoryless)

* bell curve, law of large nos.

Markov → Non independent events may also conform to patterns

(In long run, dist settle to pattern)

* if events are subj to fixed prob.,

interdependent events conform to average.

Q. How can bayesian inference be used to quantify uncertainty in predictions?

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

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

- quantifying uncertainty
- comparing models
- generative 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 occurrence of extreme values

Discrete distributions:

- Binomial
- Poisson
- Negative binomial
- geometric
- Bernoulli

Continuous

- Normal
- Exponential
- logistic Cauchy
- gamma
- chi squared

Q. lets say we have a distribution which is generated by combining multiple commonly known distributions. How do we find those & separate 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~~ ^{using} different ~~means~~ ^{distributions}.
- Q. How to get confidence interval for mean given we have a sample?