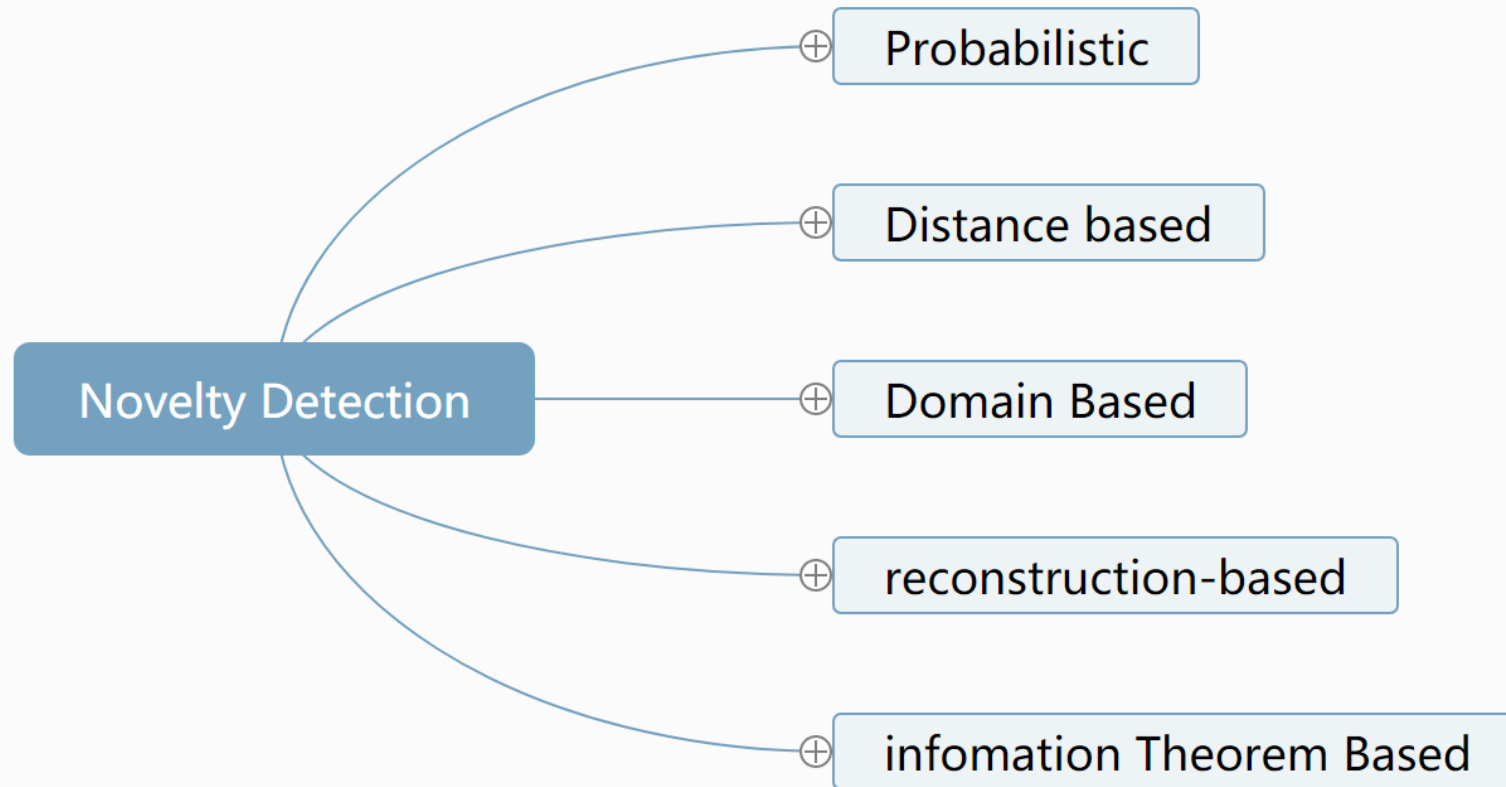


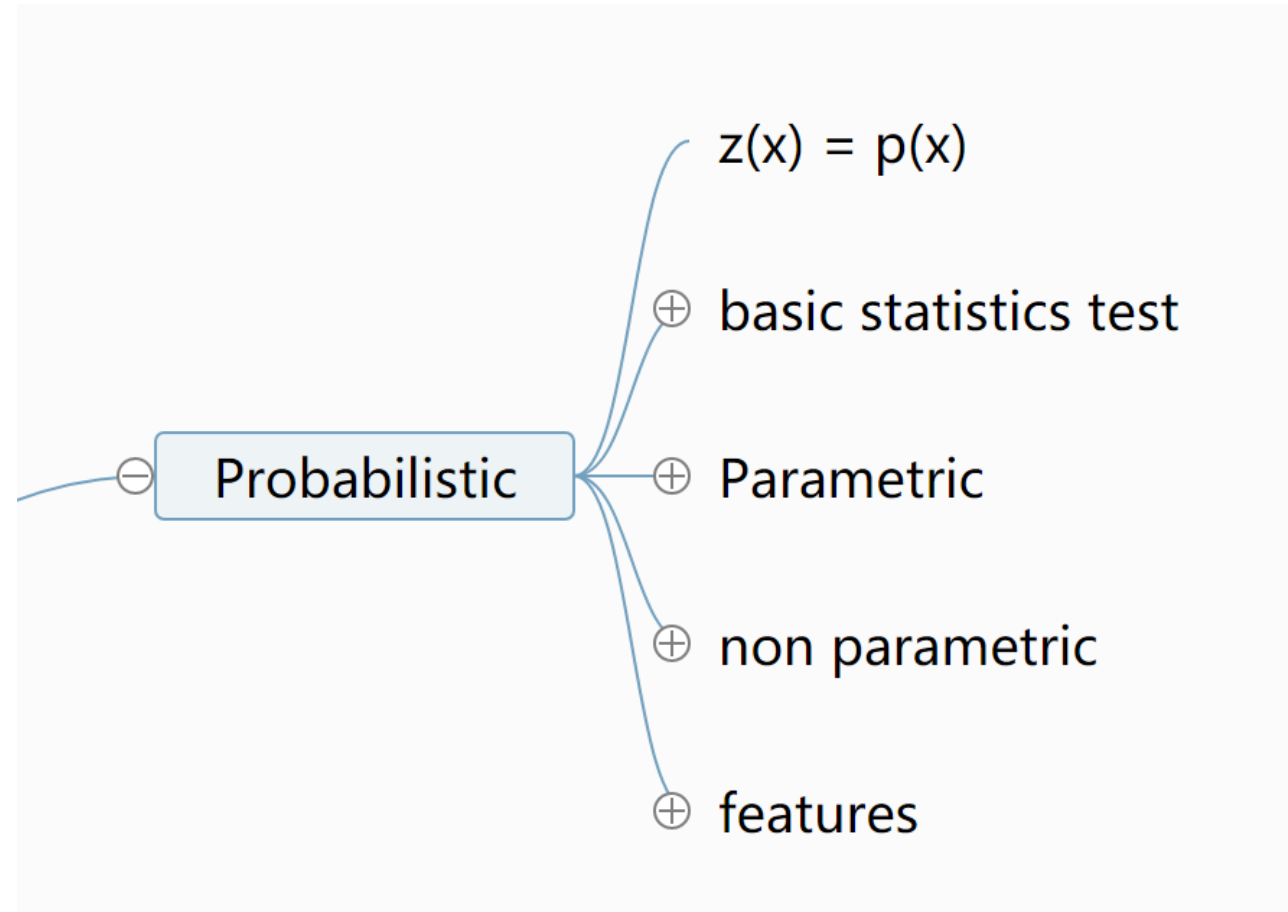
Novelty Detection Review

novelty = unknown

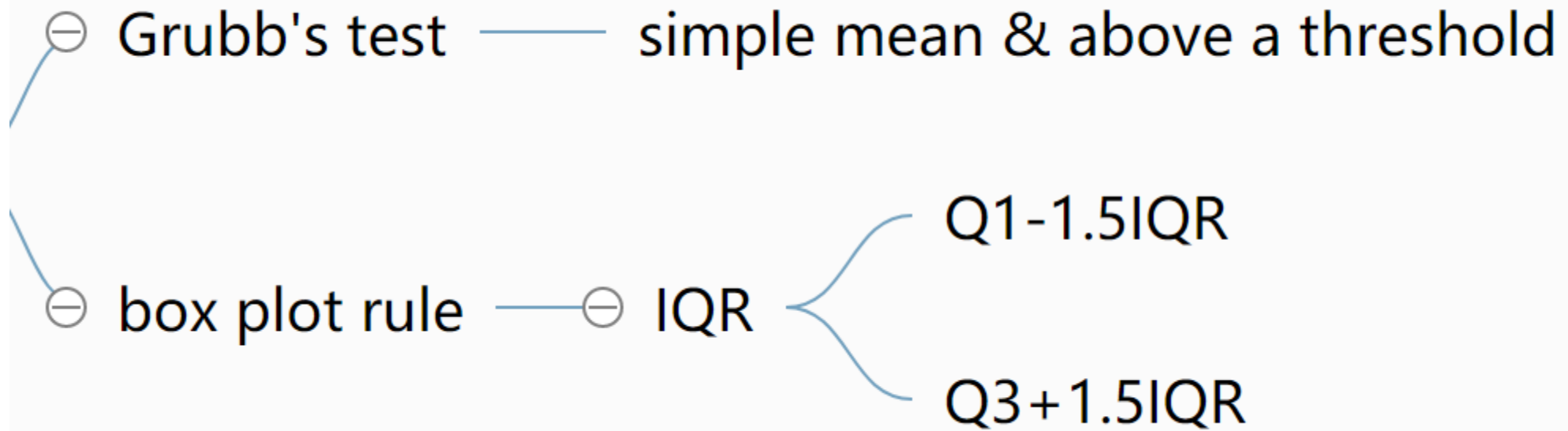
Outline



Probabilistic Approach



Basic Statistical Tests



Parametric

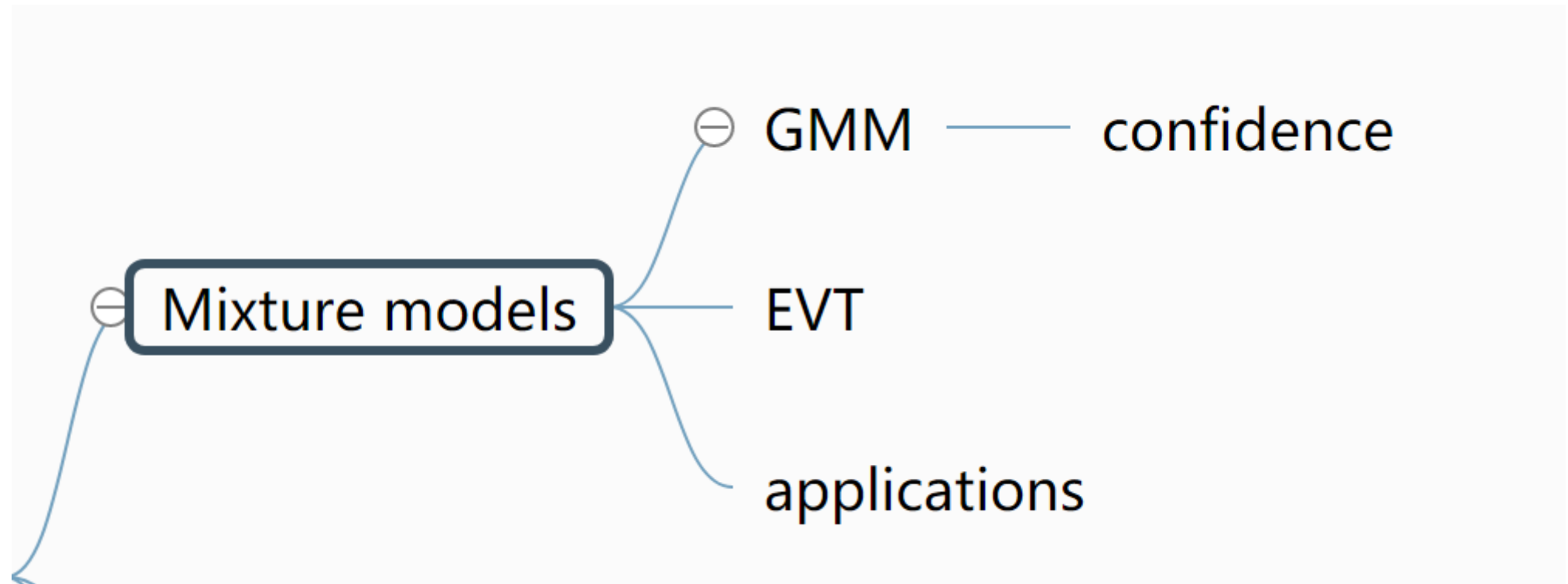
⊕ Mixture models

⊕ state-based models

⊕ features

- assume structure of model
- require priori knowledge

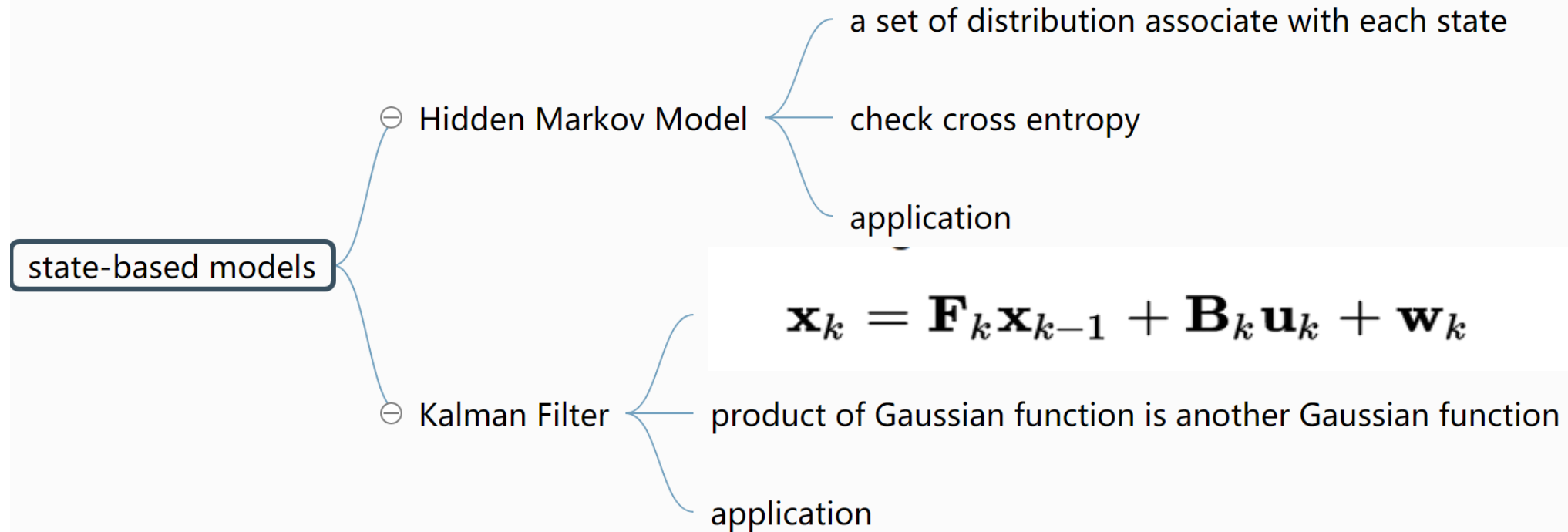
Mixture Model



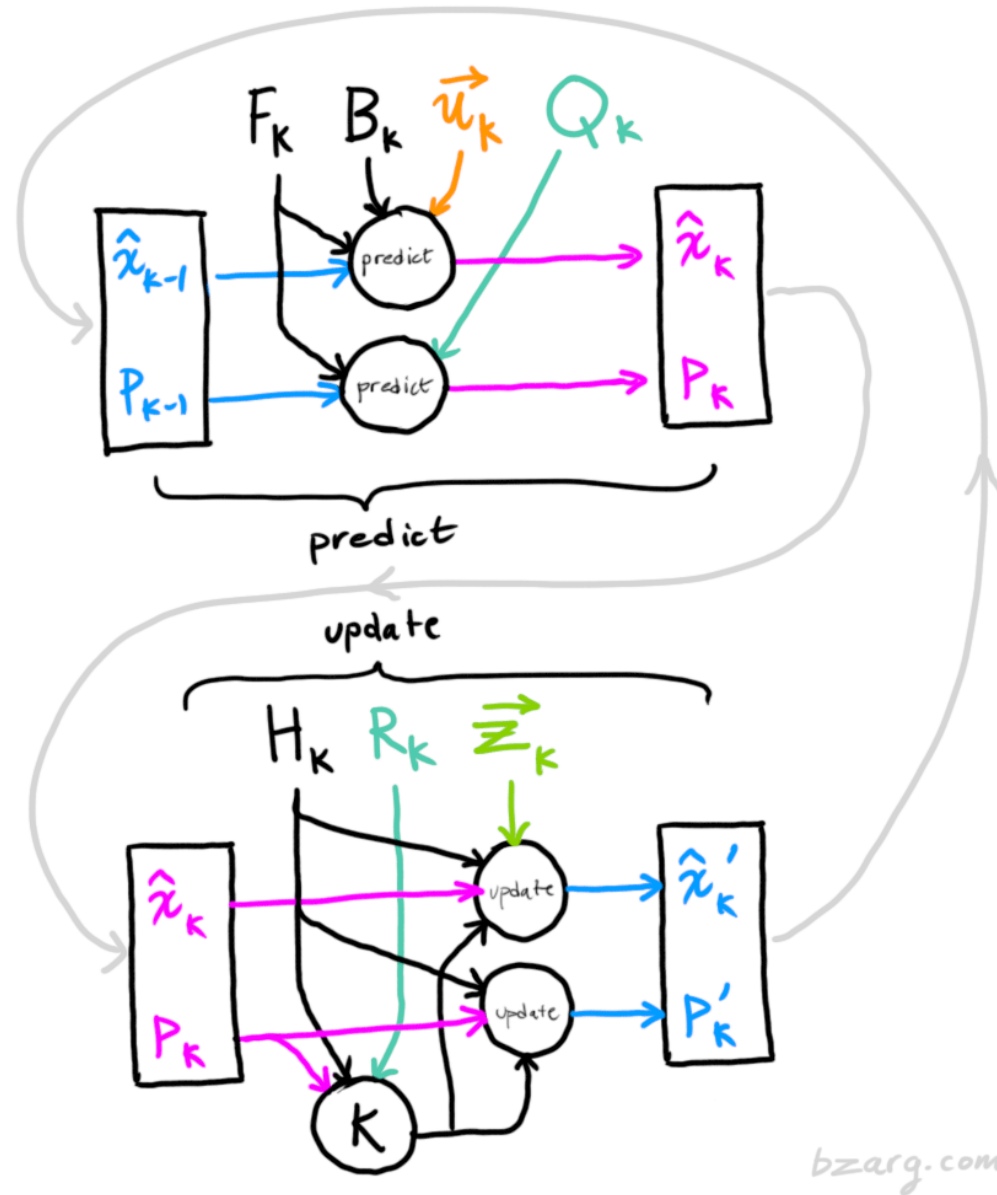
Applications

- Disease Outbreak Detection
 - hierarchical Bayesian & two GMMs
- label songs
 - ratio-reject: GMM
- real time machine status
 - k-means + GMM + greedy EM
- Pattern Recognition
 - use heuristic to adjust threshold
- online document clustering
- convert one-class classification to binary

State-based



Kalman Filter Information Flow



Application

- sound recognition
- unusual activity detection

Parametric Approach Feature

- minimal information to represent
- ⊖ poor scalability — when dimensionality increase, data points spread

Non-Parametric Approach

Kernel Density Estimator

negative selection

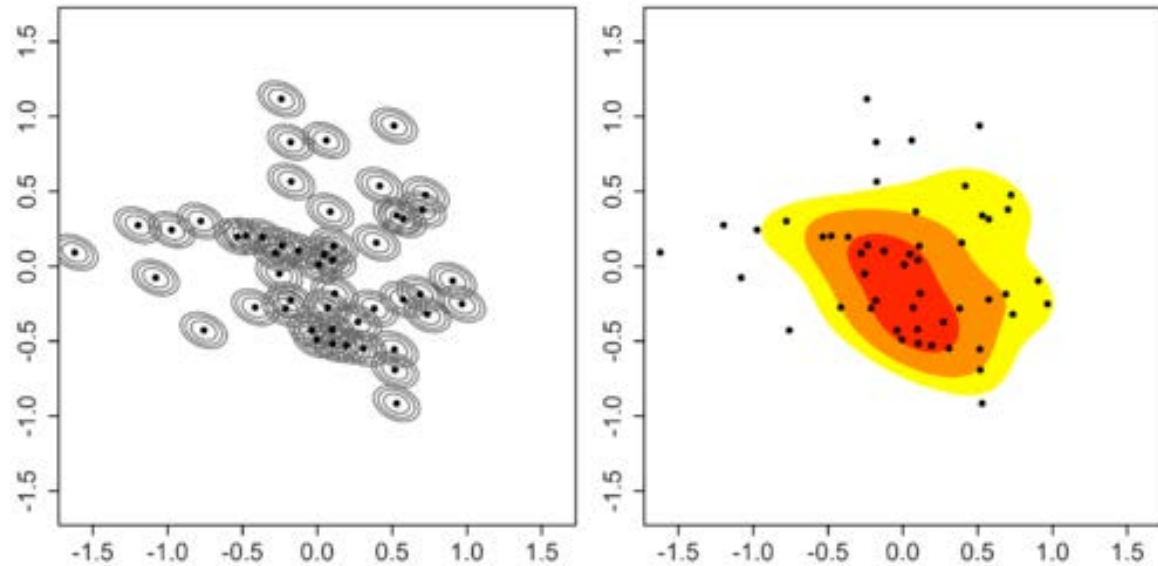
⊖ features

scale reasonably well

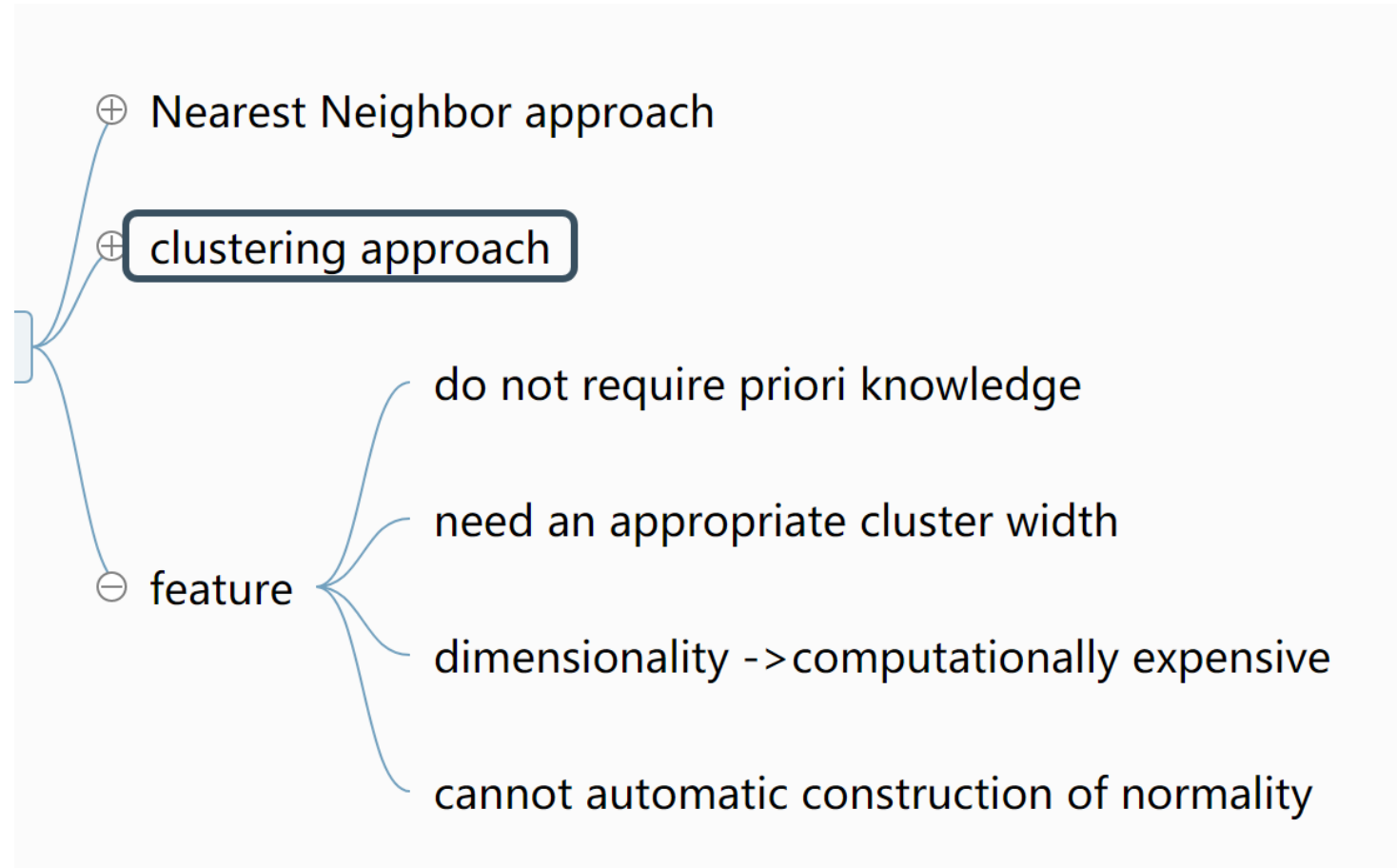
grow in size to fit data and accommodate the complexity of data

Kernel Density Estimator

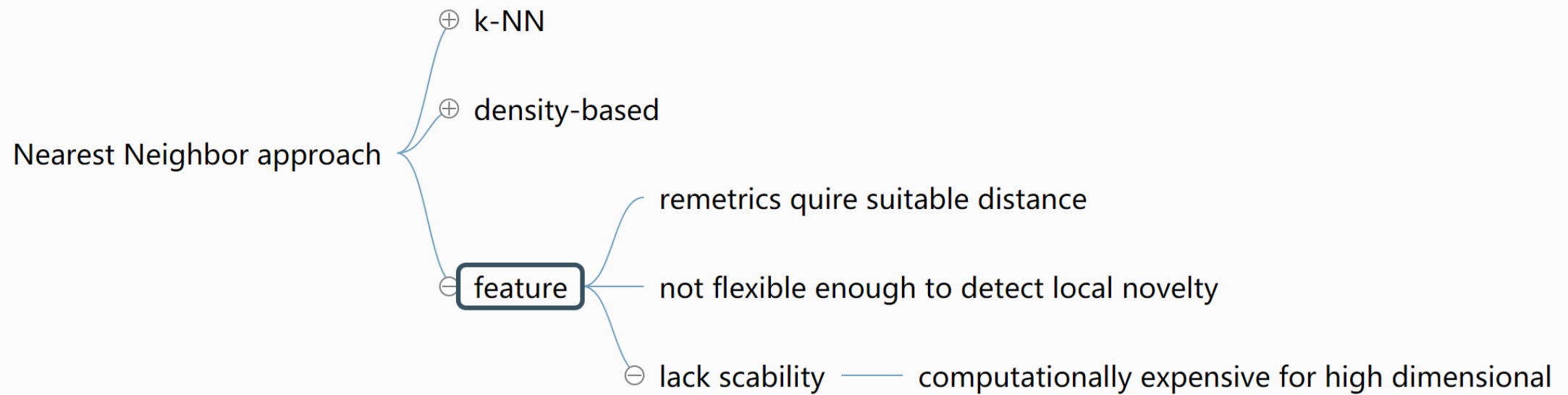
- place isotropic Gaussian kernel centered at each training point, with a single shared variance hyperparameter
- sum local contribution



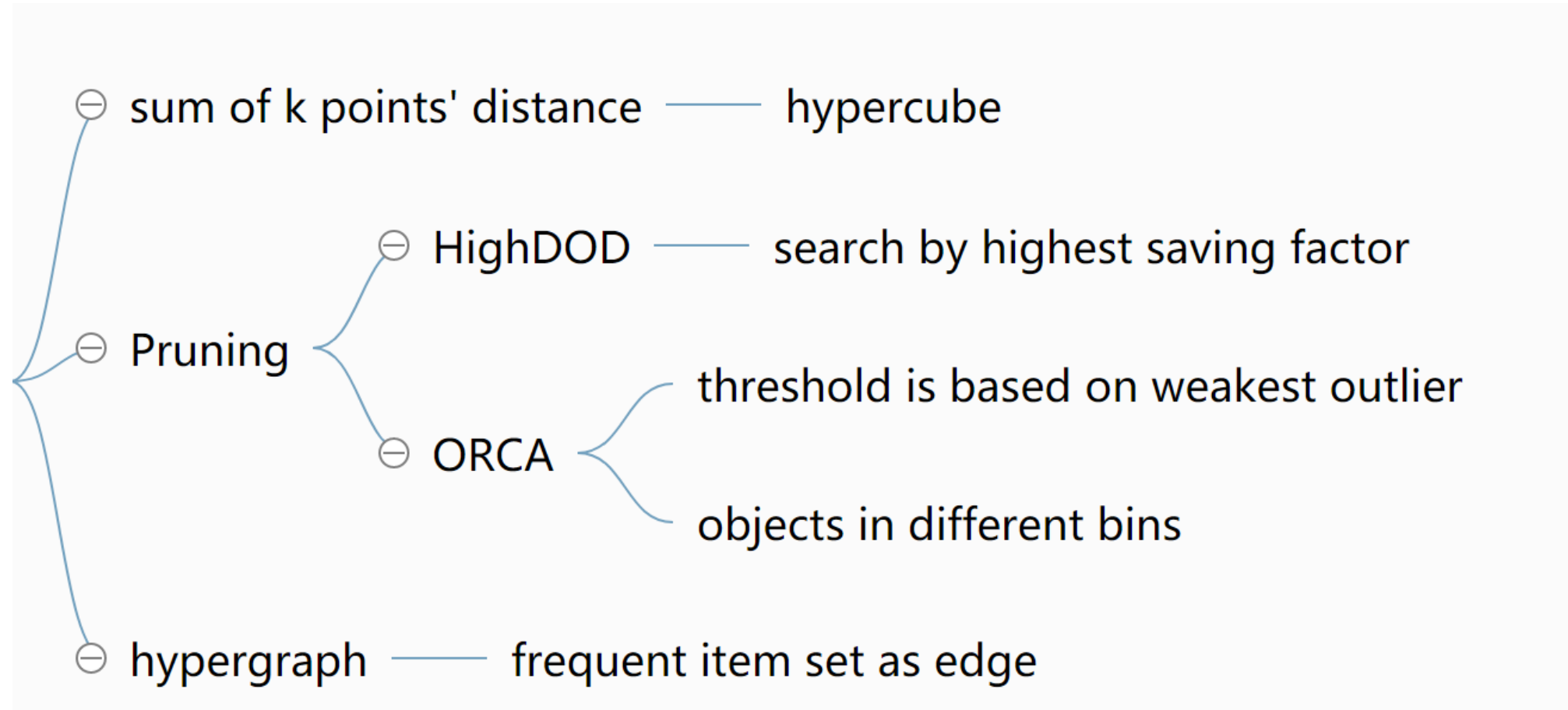
Distance based



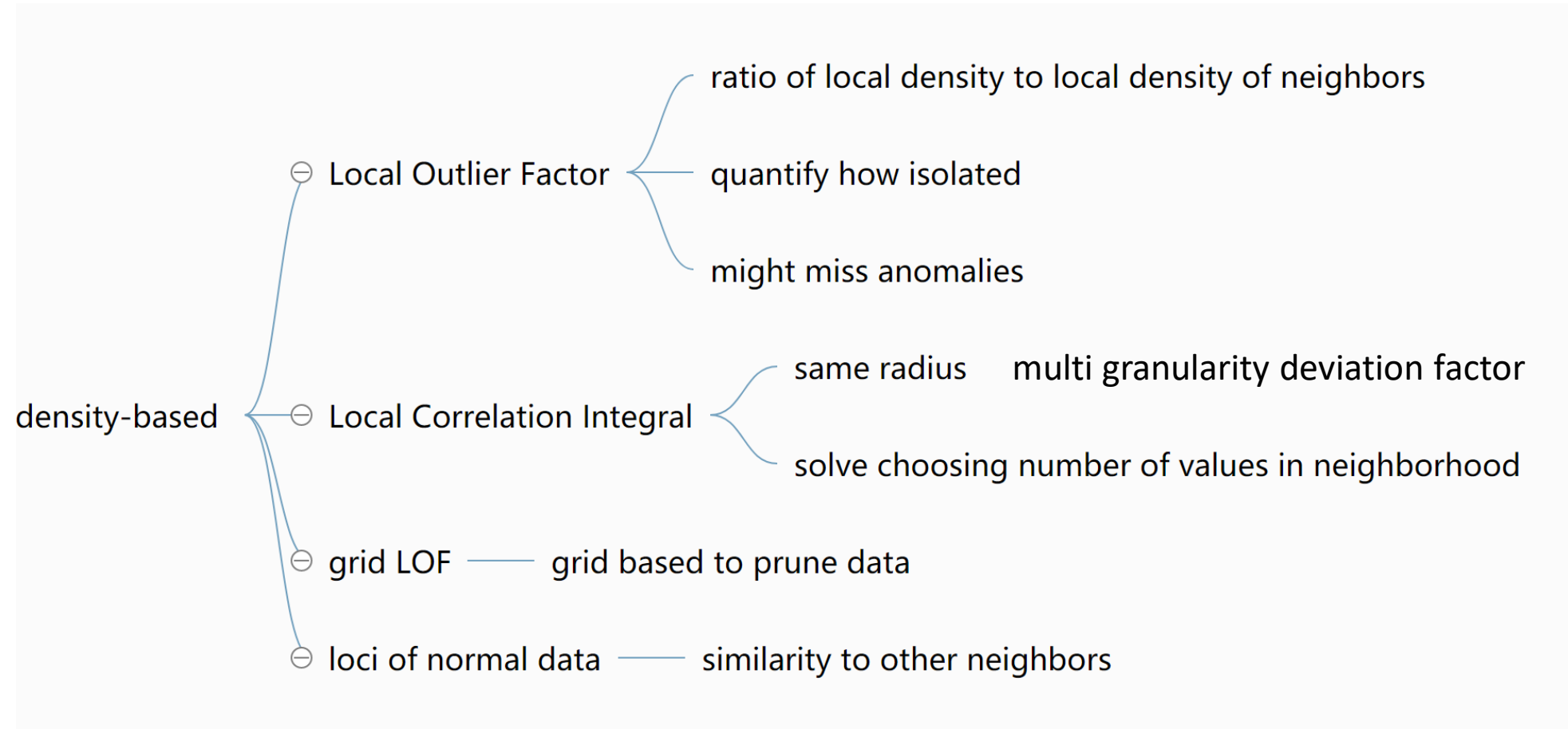
nearest neighbor



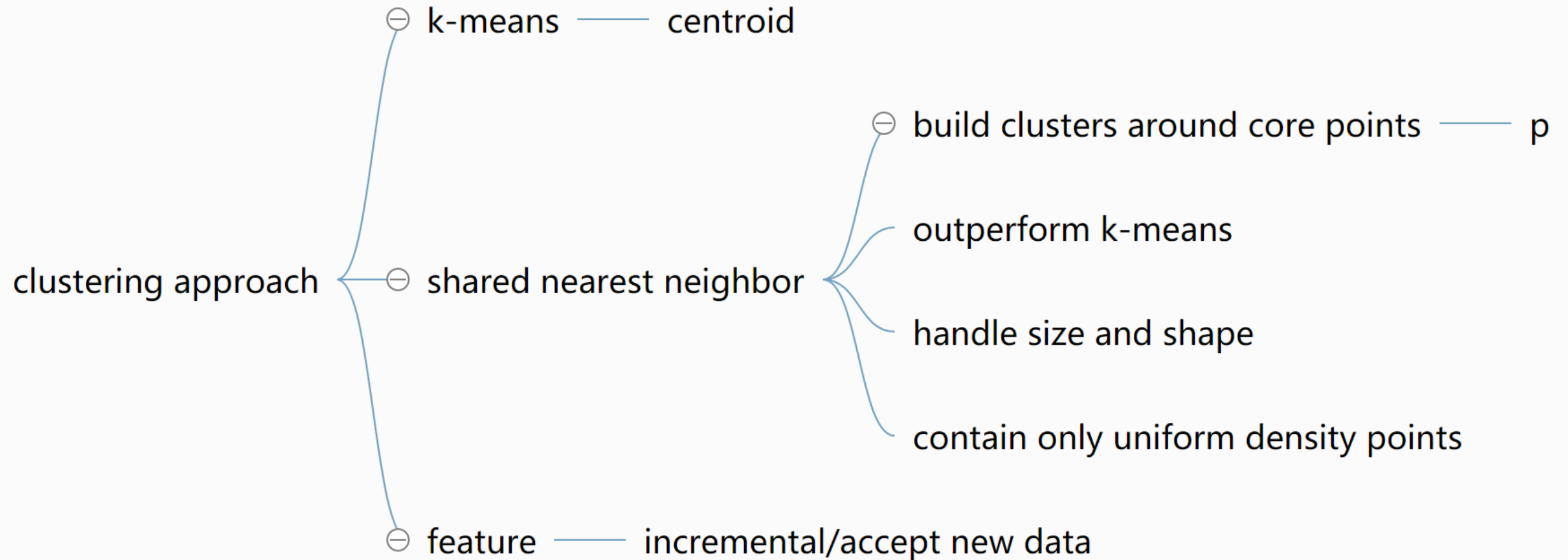
k-NN



density-based



clustering



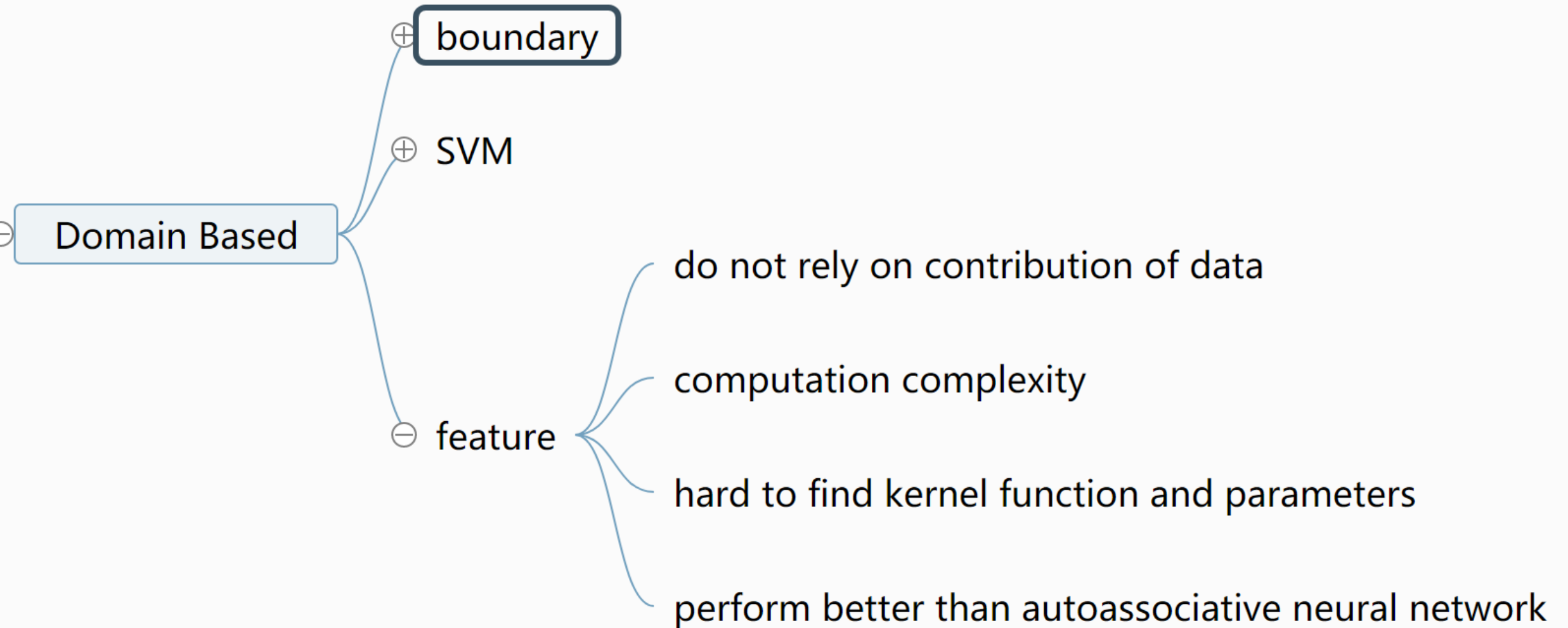
Applications

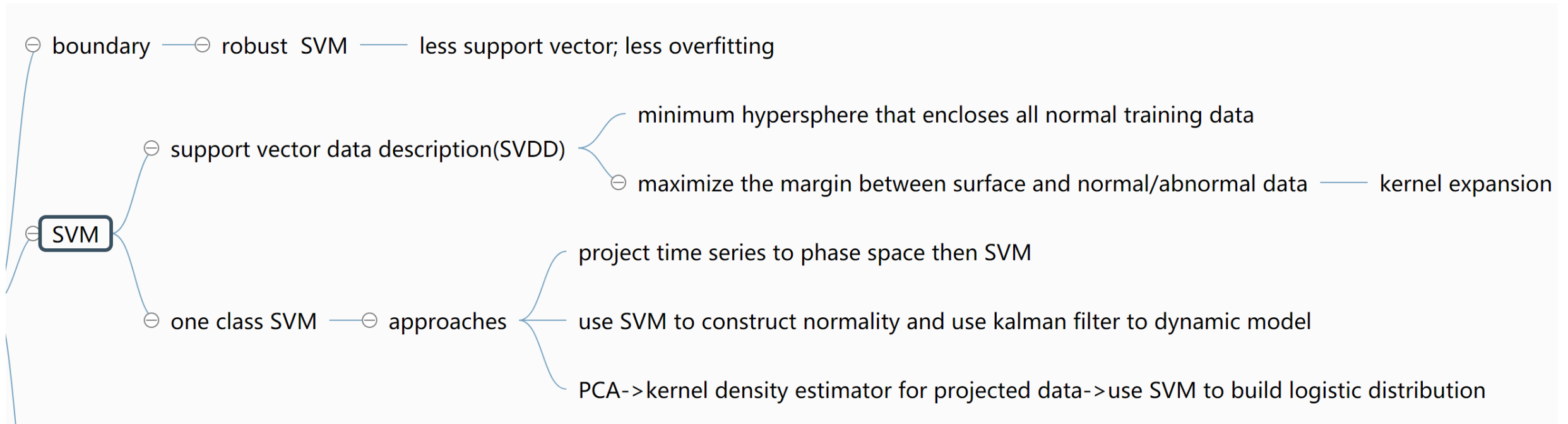
- sensor network
 - aggregation tree
 - not stable
- video scene classification & feature extraction
 - block co-occurrence matrix space
- simulation mesh data
 - distribute to every cluster, collect outlier and rebuild
 - good for local & global

Comparison Between Probabilistic and Distance-based Approach

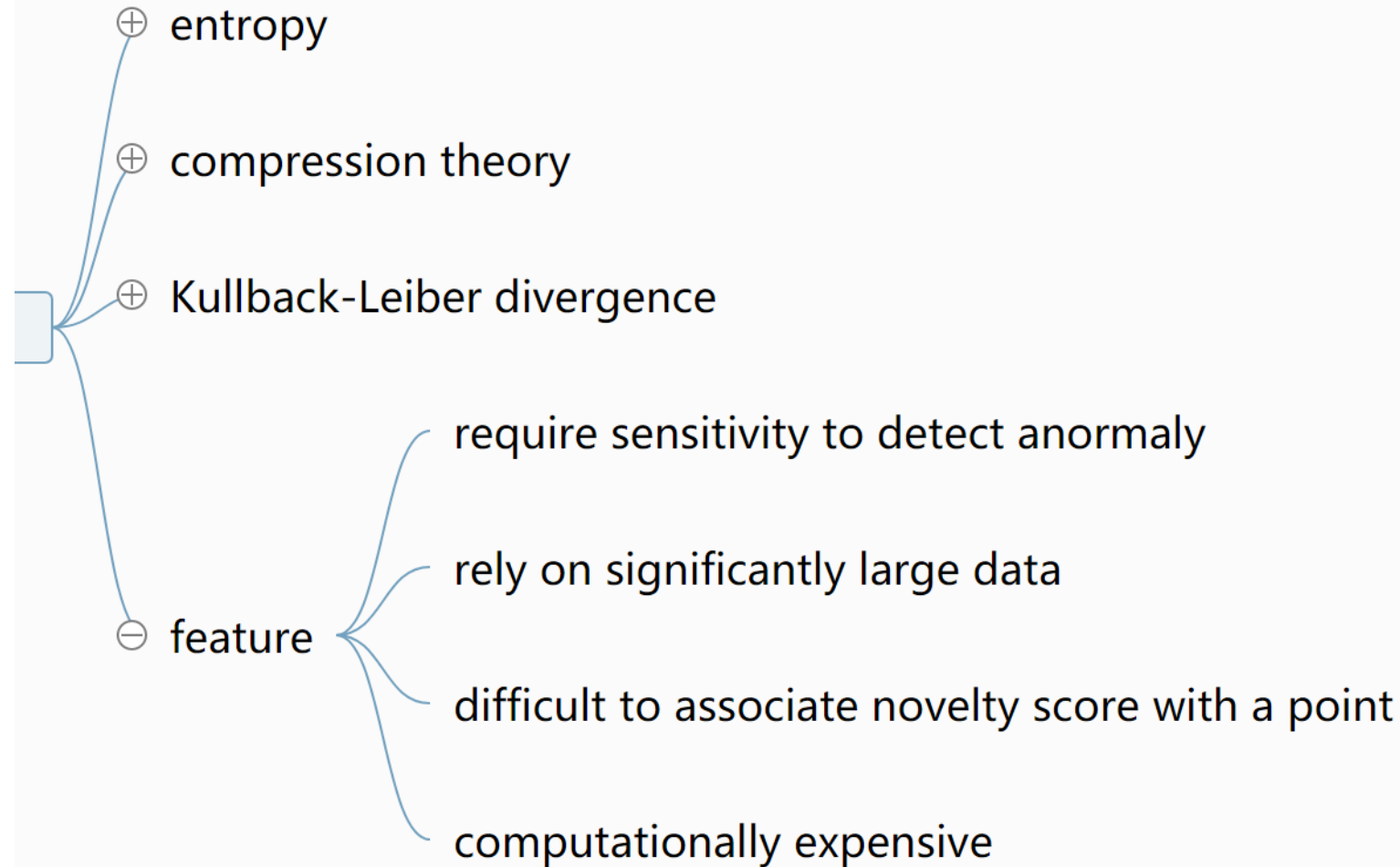
- Assumption
 - assume occupied by normal data and assign based on distance metric
- Difference
 - Distance: assume distance can discriminate points
 - Probabilistic: whether data comes from a same data model

Domain-based





Information theorem based

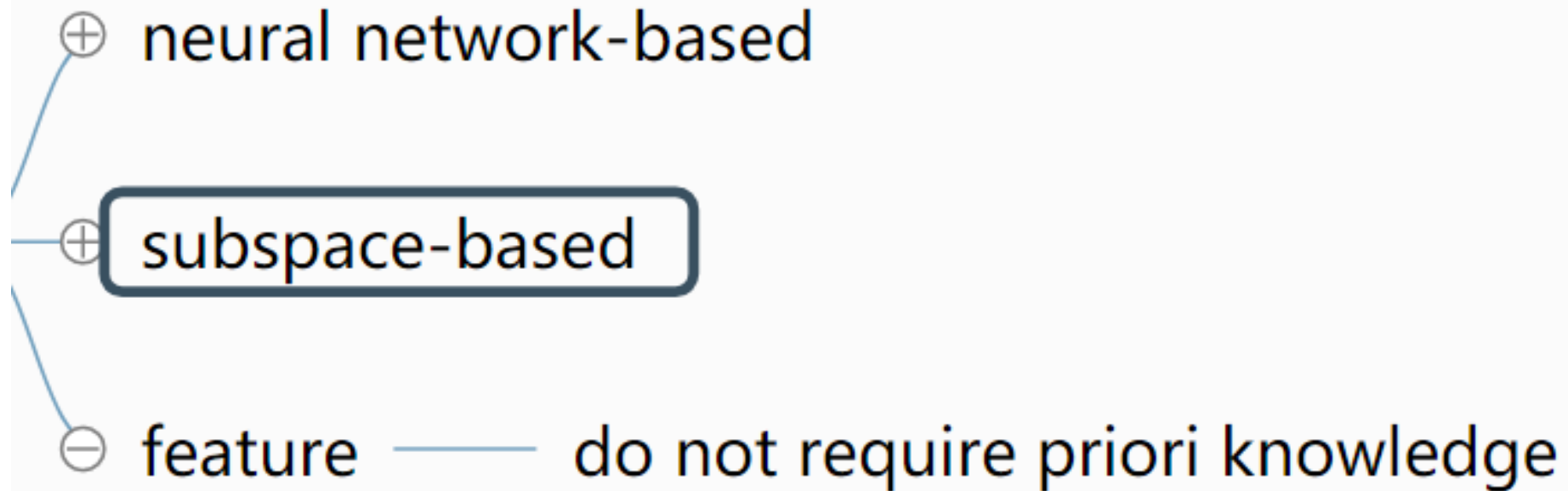


⊖ entropy — outlier: entropy decreases after removal

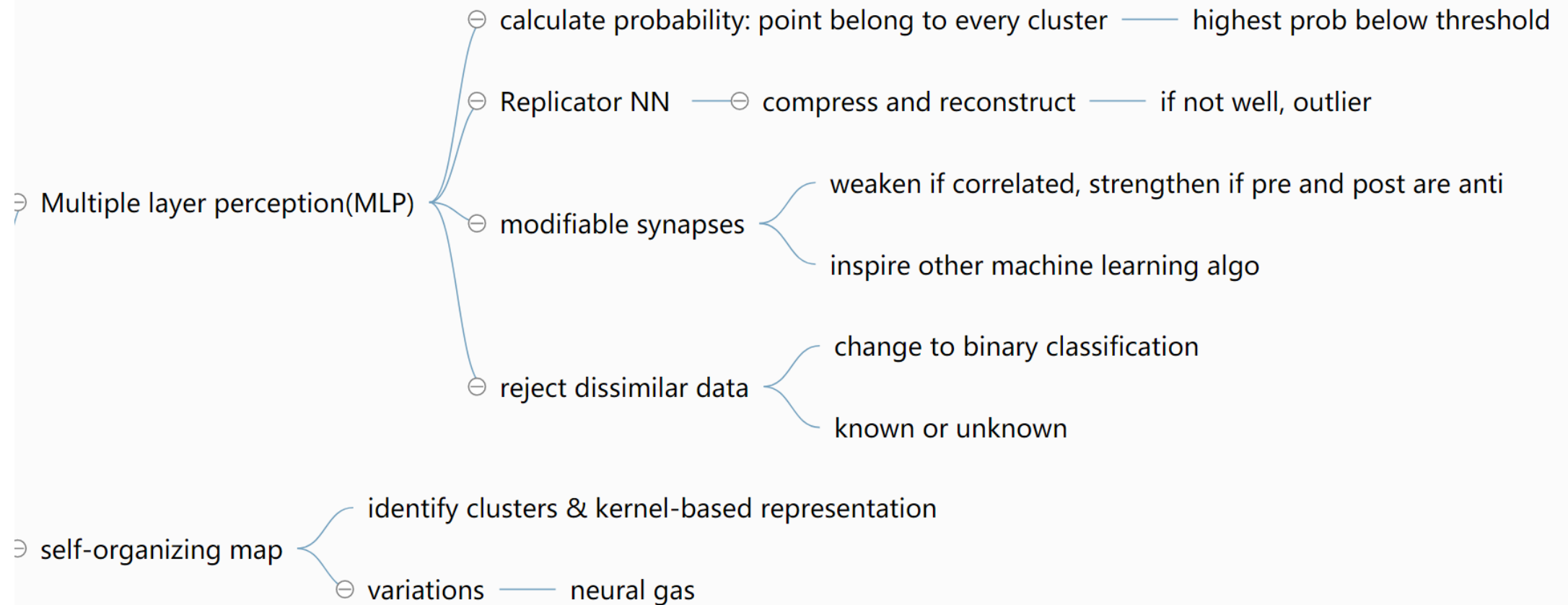
⊖ compression theory { dissimilarity method: how much computational resources to specify an object
calculate Euclidean distance of sub-sequence

⊖ **Kullback-Leiber divergence** { compare with and without by kl
do well in parametric
unknown for non-parametric
before and after "surprise"


Reconstruction-based



Neural Network



neural network based feature

- 
- require pre-defined parameters
 - ⊖ sensitive to parameters — problem with high dimensional data
 - ⊖ constructive — can grow in size
 - require good training method and stopping criterion

Applications

- outdoor scene classification: rejecter
- robot: growing neural gas
- real-time application: SOM

subspace based approach

