NDD October 7^{th}, 2021

1. Data experiment / OpenML
2. CNN – KDN → GNN?
3. New other data
4. High dimensional problem → sparse parity
5. Robust to contaminating distributions.
6. Run the benchmarks.

* The accuracy is better than the chance accuracy.

* FTE → no theoretically guarentees.
  can't say for sure since study's are empirical.
  Theory!!
  adversarial task! → hard to measure / task similarity

- recruiting gaussian distributions relevant to task.
- pruning the polytope
- KDF
- CNN??

- double counting → increase sample size.

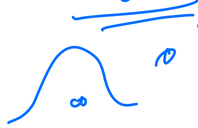
KDG / Jayanta Oct 11th 2021

2^5 possible polytopes. → 5

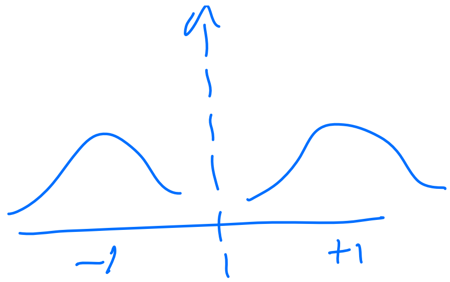
possibilities.

* penultimate layer →
* followed same path → same partition.
* CIFAR-10 - simple data
* covariance matrix ~ dimensionality re
  → high dimensions
* KDN → feature selection.
* Deep Boltzmann Machines.
* High Dimensionality → KDN
* MNIST → • test the KD-CNN on a lower dim dataset
  CIFAR-10 ↗
  • penultimate layer — ①
  • all the fully-connected layers ②
  • Deep Boltzmann machines
  • Dealing with high-dimensionality
  • Histogram of polytopes

Weighted Maximum Likelihood



→ weighting using the # paths.



weighting!

Dim-reduction!

sim_kdn

sim_kdf

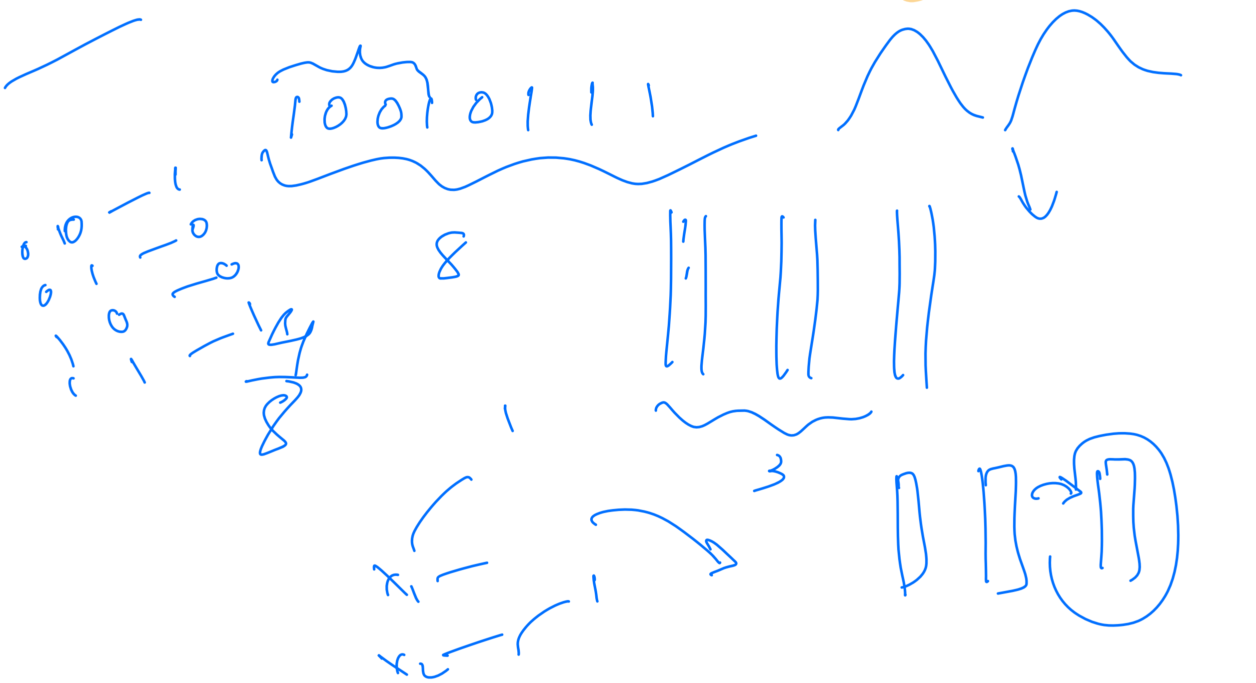
high-dim.

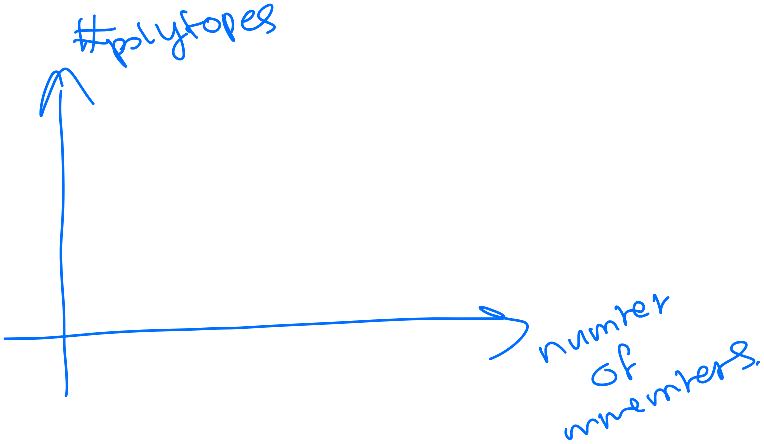
1 0 0 1 0 = 2

0 1 0 0 1 = 2

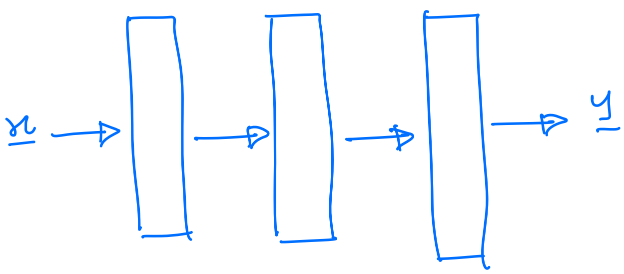
all the layers

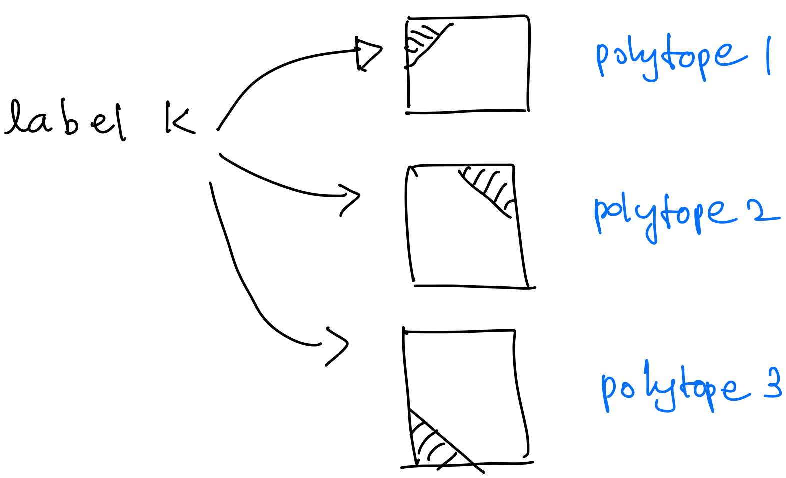
penultimate layers.



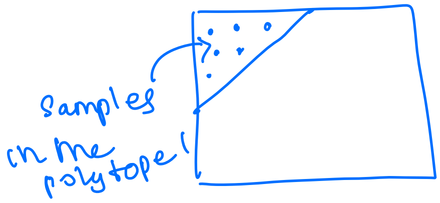


Kernel Density Network





consider polytope 1 $\Rightarrow$



samples $\Rightarrow (x_1, x_2, x_3, \ldots, x_m)$

where $x_i \in \mathbb{R}^d$ (d – feature dimension)

* We are interested in fitting a Gaussian over the samples (aka polytope members)

We need to estimate parameters μ, Σ $\leftarrow$ $\phi(x \mid \mu, \Sigma)$

d-variate Gaussian

The joint distribution of the observed m polytope members. $\Rightarrow$

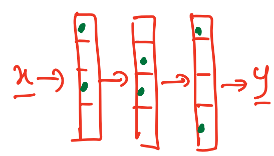
$$f(x_1, x_2, \ldots, x_m \mid \mu', \Sigma')$$

If x_i $(i = 1, \ldots, m)$ are iid,

$$\underbrace{f(x_1, x_2, \ldots, x_m \mid \mu', \Sigma')}_{} = \prod_{i=1}^{m} \phi(x_i \mid \mu, \Sigma)$$

$$\ell(\mu, \Sigma) = \prod_{i=1}^{m} \phi(x_i \mid \mu, \Sigma)$$

$$\hat{\mu}, \hat{\Sigma} = \underset{\mu, \Sigma}{\operatorname{argmax}} \prod_{i=1}^{m} \phi(x_i \mid \mu, \Sigma)$$



* In the KDN, all the sample points in a given polytope have the same activation pattern

eg:- $x_1 \rightarrow$ [1010 0110 1001]

$x_2 \rightarrow$ [1010 0110 1001]

$x_3 \rightarrow$ [1010 0110 1001]

$\vdots$

$x_m \rightarrow$ [1010 0110 1001]

* when $d \gg m$, it's not feasible to estimate Σ ($d \times d$ matrix) from just the m observations

* However, there might be other samples in the same class (but in different polytopes) that are approximately-matched with the activation patterns of the samples in the given polytope.

eg:-

$$\left.\begin{array}{l} x_1 \longrightarrow [1010\ 0110\ 1001] \\ x^* \longrightarrow [1010\ 1010\ 1001] \end{array}\right\} \underline{\underline{x_1 \approx x^*}}$$

* How does Jayanta's weighting help with the incorporation of approximately-matched samples?