Literature review of the paper(recommended by Dr. Wang)

**LEARNING EMBEDDINGS INTO ENTROPIC WASSERSTEIN SPACES**

Background:

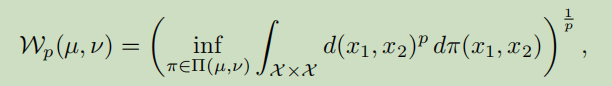
* Euclidean embedding unable to capture structure data. Euclidean space makes assumption about the neighbourhood size and connectivity, hence may not be able to represent complex relationships.
* Wasserstein space:

Input: probability distributions

Function: endow probability distributions with an optimal transport metric, which measures the distance traveled in transporting the mass in one distribution to match another. Flexible that it allows other metric spaces to be embedded to it while preserving the original distance metrics. - flexibility may allow representing complex relationships.

* Wasserstein distance:

The p-Wasserstein distance between two probability distributions μ and v over a metric space X is: Giving the optimal transport plan.



Infimum(inf): of a subset S of a partially ordered set T is the greatest element in T that is less than or equal to all values of S, if such element exists. If a set has a smallest number, it is the infimum of the set, also called the minimum.

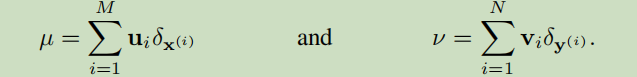
∏: transport plans that distribute the mass in μ to match that in v.

d(x1,x2): ground metric on X giving the cost of moving a unit of mass from support point x1 ∈ X underlying distribution µ to point x2 ∈ X underlying ν.

This paper:

* Embed input data as discrete distributions, instead of Gaussian in previous papers.

Discrete distribution:



**u,v** are non-negative weight vectors summing to 1. x(i),y(i) are the supporting points in Rn. There are M x points and N y points.