Reprensation Learning

Mazino

04 April 2024

What is representation learning? Representation learning is a type of machine learning that involves learning a representation of the input data. The goal of representation learning is to learn a representation of the input data that is useful for a particular task, such as classification, regression, or clustering. Representation learning is a powerful technique that has been used to achieve state-of-the-art performance on a wide range of machine learning tasks, including image recognition, speech recognition, and natural language processing.

In this lecture we will be talking about topological descriptors. What choices of topological descriptors are there? Their properties and how they can be used in representation learning.

First descriptor you know : Persistence Diagrams :

Points are tuples in $\mathbb{R} \times \mathbb{R} \cup \{\infty\}$

Persisstence corresponds to distance to the diagonal

Multiplicity of each point is not apparent

We can calculate distances between persistence diagrams using the bottleneck distance. The bottleneck distance between two persistence diagrams is the minimum cost of a matching between the points of the two diagrams, where the cost of matching a point to the diagonal is the distance of the point to the diagonal. The bottleneck distance is a metric on the space of persistence diagrams, and can be used to compare the similarity of two persistence diagrams.

$$d_B(D_1, D_2) = \inf_\gamma \max_{p \in D_1} d(p, \gamma(p))$$

where γ is a bijection between the points of D_1 and D_2, and $d(p, q)$ is the distance between two points p and q in the persistence diagram.

There is also the Wasserstein distance between persistence diagrams. The Wasserstein distance between two persistence diagrams is the minimum cost of a matching between the points of the two diagrams, where the cost of matching a point to another point is the distance between the two points raised to the power p. The Wasserstein distance is a metric on the space of persistence diagrams, and can be used to compare the similarity of two persistence diagrams.

$$W_p(D_1, D_2) = \left(\inf_\gamma \sum_{p \in D_1} d(p, \gamma(p))^p \right)^{1/p}$$

where γ is a bijection between the points of D_1 and D_2, and $d(p, q)$ is the distance between two points p and q in the persistence diagram.