Witness Autoencoder: Shaping the Latent Space with Witness Complexes

Abstract

We present a Witness Autoencoder (W-AE) – an autoencoder that captures geodesic distances of the data in the latent space. Our algorithm uses witness complexes to compute geodesic distance approximations on a mini-batch level and leverages topological information from the entire dataset while performing batch-wise approximations. This way, our method allows to capture the global structure of the data even with a small batch size, which is beneficial for large-scale real-world data. We show that our method captures the structure of the manifold more accurately than the recently introduced topological autoencoder (TopoAE) in the recent introduction.

Motivation

Currently, autoencoders (AEs) are widely used for non-linear dimensionality reduction in various applications, mainly due to the expressiveness of neural networks and the encoder-decoder architecture. However, one of the key issues of AEs is that their latent spaces do not necessarily reflect the geometric and topological structure of the true data manifold – i.e., they are not guaranteed to preserve relative distances between points and the topological structure of the data. Preserving this structure is beneficial not only for interpretability of the latent space, but also for generalization capabilities [1,2] and robustness to adversarial attacks [3].

Proposed Method

We present a Witness Autoencoder (W-AE), which introduces a novel loss term for autoencoders to enforce structure preservation in the latent space, i.e. alignment of the geodesic distances between data and latent space. Thus the total loss becomes,

\[ L(x) = L_e(x, d) + \lambda L_c. \]

The loss term is closely related to the ones presented in [10, 11] and relies on the construction of a neighborhood graph to approximate distances on the manifold. For the construction of such a neighborhood graph, we present a new method that is based on witness complexes. It improves geodesic distance approximations on a mini-batch level, which is desirable since false neighbors and missed neighbors are more likely to occur for small mini-batches since low density regions are more likely. In figure 1 our neighborhood graph construction is compared to the method used in TopoAE and k-NN neighborhood graphs for the Swiss roll dataset, where wrong neighbors occur as so called short-circuit errors.

Experimental Study on Swiss Roll Dataset

For evaluation W-AE was trained with different batch-sizes on the Swiss roll dataset and compared to TopoAE, UMAP, ISNE and a vanilla AEs. In figure 2 qualitative results are presented and compared to TopoAE, that show that W-AE approximates geodesic distances better even for small mini-batch sizes (n_{mb}). In the table quantitative and qualitative results for all methods are presented.

References