An Inductive Bias for Distances: Neural Nets that Respect the Triangle Inequality

The triangle inequality serves as a useful constraint when modelling distance metrics. However, most deep learning architectures rely on Euclidean distance which may not be sufficient to embed representations in latent space. This requires one to rethink the utilization of Euclidean metric with modern machine learning algorithms. To this end, the work introduces three novel architectures which satisfy the triangle inequality. Architectures are proved to be universally-approximate norm inducing metrics and outperform existing metric approaches in graph and reinforcement learning based scenarios.

Euclidean metrics restrict the representation of various embeddings which can otherwise be expressed using the triangle inequality. Motivated by this property, proposed architectures make use of norm inducing propositions which guarantee to satisfy the traingle inequality. Firstly, the DeepNorm architecture makes use of asymmetric semi-norms which are convex to yield expressive k layers. In particular, the pairwise maxReLU operates element-wise and preserves convexity and positive homogeneity. Secondly, WideNorms are combination of asymmetric semi-norms which are based on non-negative sums and max as valid combinations. This gives rise to maxmean which is a vector-wise combination. The Wide Norm architecture makes use of a k-component mixture of Wide Mahalanobis norm and is computationally beneficial due to pairwise distances in case of large minibatches. Lastly, Neural Metrics comprise of metrics defined on non-negaitve sum of metrics. This requires one to define metric-preserving functions which when applied to wide metrics and combined with MaxMean yield a valid metric. Neural Metrics allow improved modeling of certain metrics which may not be sufficiently expresseive in the case of Deep and Wide Norms.

Proposed metrics are validated on three different applications consisting of metric nearness, modelling graph distances and learning general value functions. Deep Norms depict accurate learning of metrics as a result of expressive implementations arising from the depth of architecture. Wide and Deep Norms depict improved performance when compared to Euclidian norms in modeling graph distances. Furthermore, proposed architectures highlight strong generalization properties when leanring value functions for the 4-room and Maze tasks. While induction of norms yields improved architectures, it presents a few caveats when modelling graph distances. Conventional Mahalanobis norm performs signficantly well in terms of generalization when compared to Deep and Wide Norms. This undermines and contradicts the strong generaization results observed when leanring value functions for reinforcement Learning navigation tasks. A possible explanation could be that the norms overfit to distance metrics in the conventional supervised setting which hinders their generalization across different training settings.

Norm-inducing architectures for distance-based learning provide two new directions for future work. Firstly, constructed architectures improve expressivity and scalability to large batch sizes which is an interesting direction for developing large-scale models in the unsupervied setting. Secondly, proposed architectures present room for generalization across different training scenarios which would extend their application to practical scenarios.

Euclidean metric restricts the representations of various expressions which highlight the need for a more suitable metric such as the triangle inequality. To this end, the work presents Deep Norms, Wide Norms and Neural Metrics which improve expressivity, scalabiity and the overall learning capacity of representation embeddings in latent space. Architectures are modelled using asymmetric norms and demonstrate improved performance in learning and generalization.