Learned Embeddings for Efficient Clustering

Key message: Learned embeddings offer a large improvement in efficiency and purity of doublet-building, seeding, and clustering strategies for particle tracking.

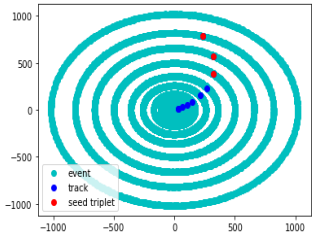
- Learn an embedding model as follows:
  - Dataset $T = \left\{\left(t_1^{(1)}, t_2^{(2)}\right), \ldots, \left(t_n^{(1)}, t_n^{(2)}\right)\right\}$, t_i a triplet of hits
  - Embedding model $\phi(x) \in \mathbb{R}^m$, where ϕ is an MLP and m is the embedding dimension
  - Supervise MLP with hinge loss l:

  Let $d = \|\phi(t^{(1)}) - \phi(t^{(2)})\|_2$ be the distance between $t^{(1)}$ and $t^{(2)}$. Then

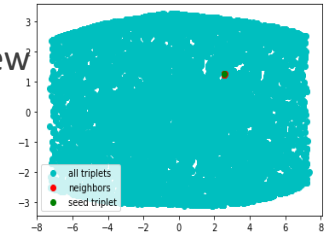
$$l(t^{(1)}, t^{(2)}) = \begin{cases} d & \text{if } t^{(1)}, t^{(2)} \text{ belong to same track} \\ \max(0, 1-d) & \text{else} \end{cases}$$

- Building triplets using embeddings achieves a **TrackML score of** $\approx$ **0.78** of 0.82 maximum recoverable.

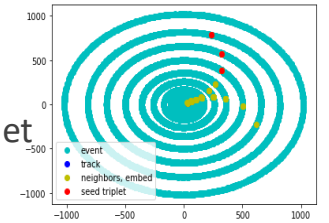
1) Choose seed from candidate triplets



2) Represent points in new space with Euclidean metric



3) Filter neighboring triplets against seed triplet



4) Output remaining hits as final track candidate