Using encoders as anomaly detectors

Nadia van Beurden
Masterthesis
Radboud University – Sascha Caron

The Idea

If we train a model very well on 4-top events, it gives false predictions on non 4-top events.

Training on: particles generated by the 4-top event

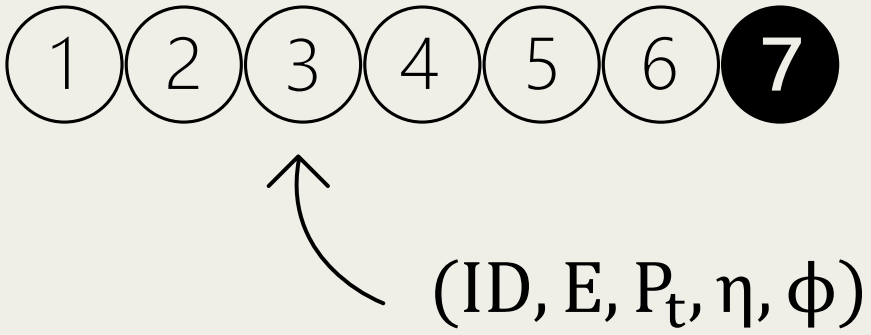
Predicting: next particle generated

The Idea

If we train a model very well on 4-top events, it gives false predictions on non 4-top events.

Training on: particles generated by the 4-top event

Predicting: next particle generated

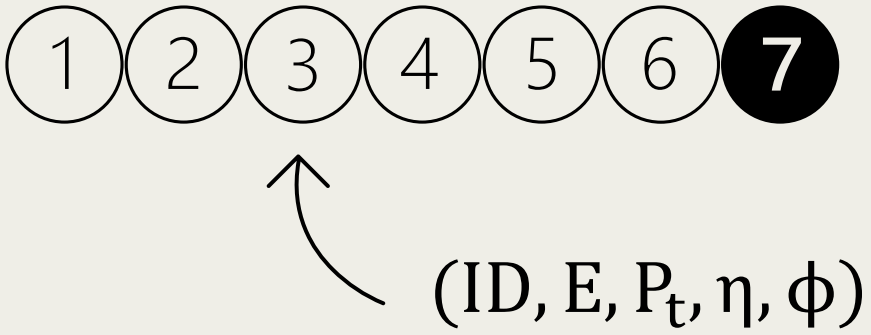


The Idea

If we train a model very well on 4-top events, it gives false predictions on non 4-top events.

Training on: particles generated by the 4-top event

Predicting: next particle generated

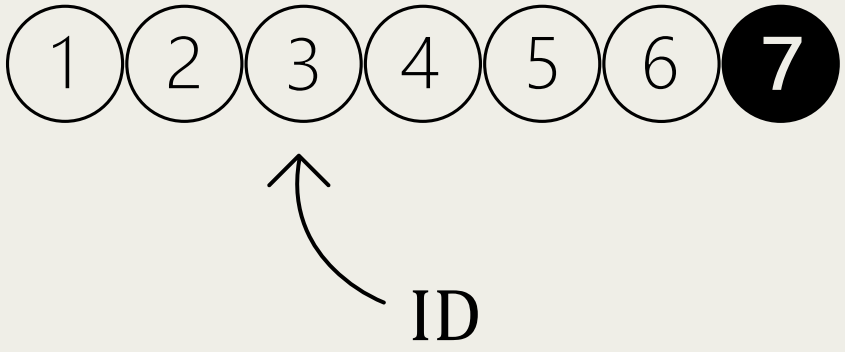


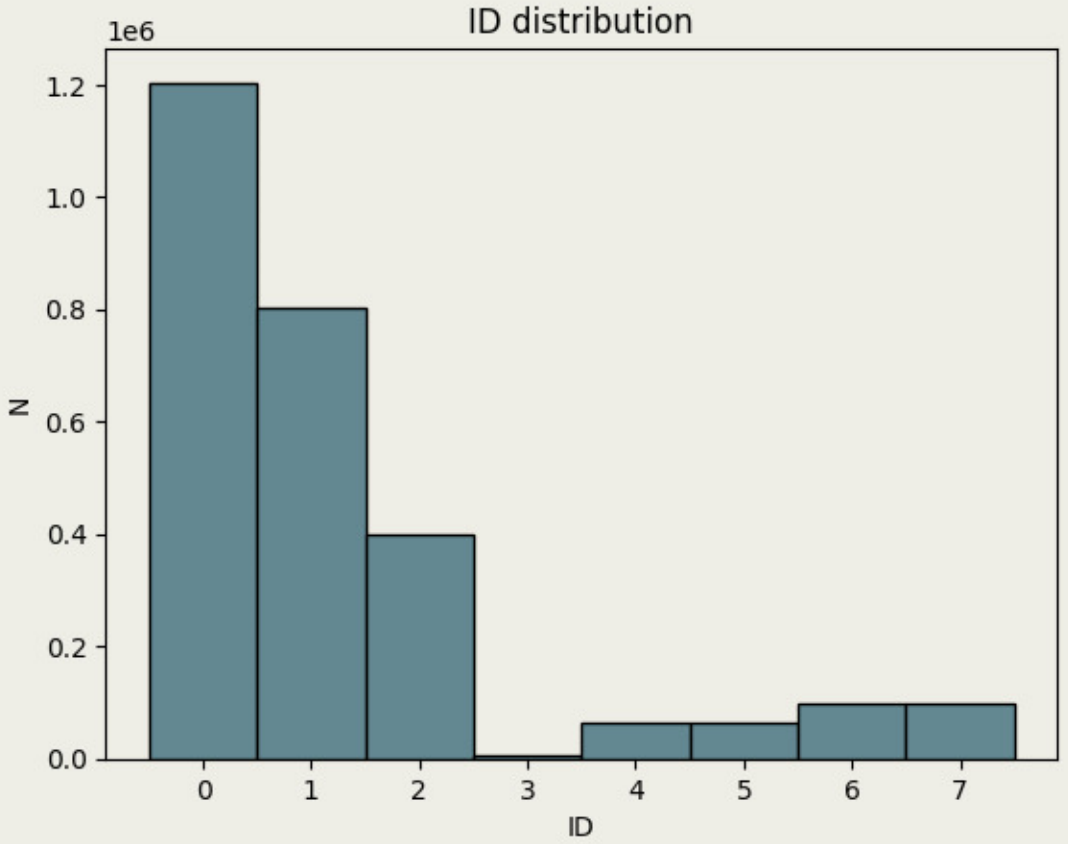
Attention and Embedding should capture relations and intrinsic information

Focus on particle ID

Can we predict the next particle type in the sequence of generated particles?

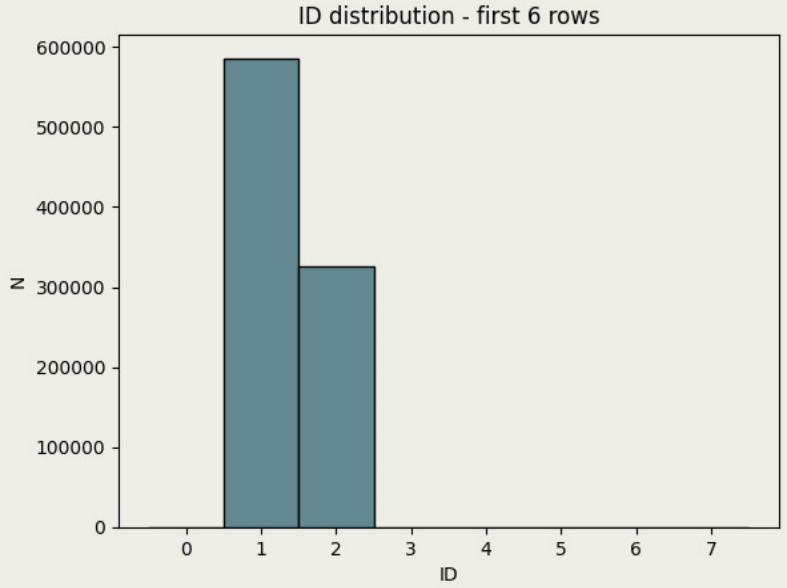
Already a discrete variable

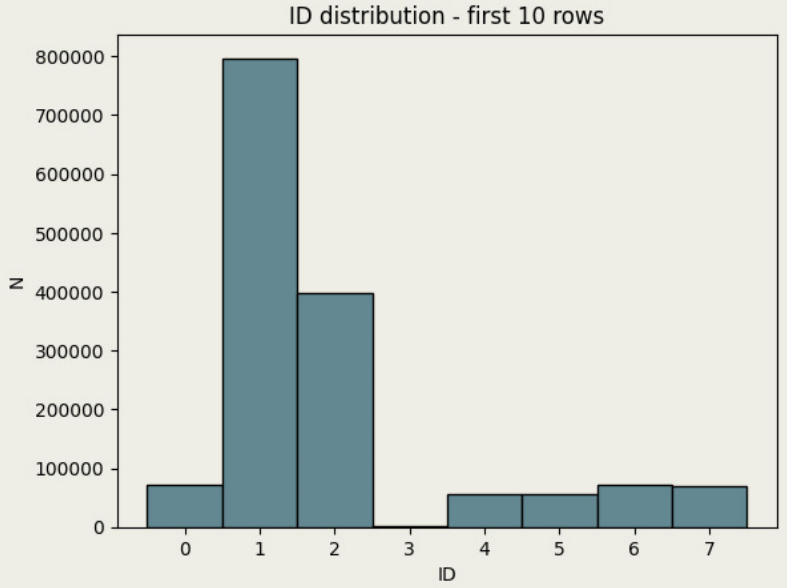


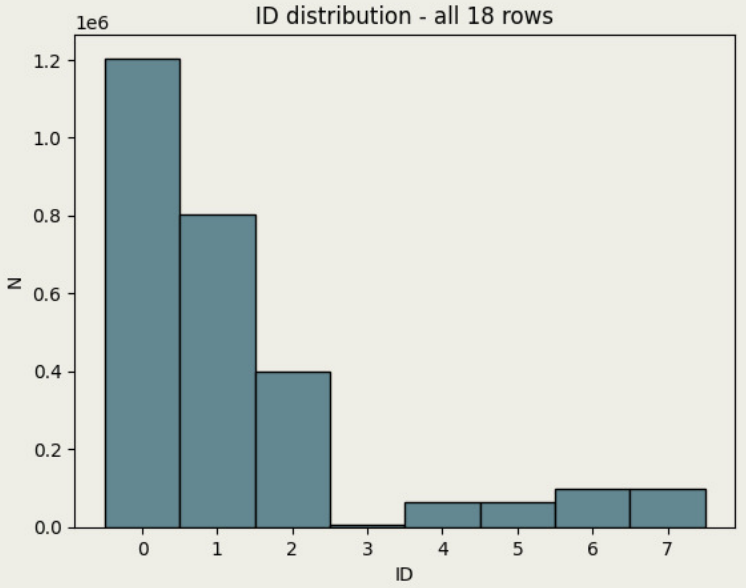


Focus on particle ID

Can we predict the next particle type in the sequence of generated particles?



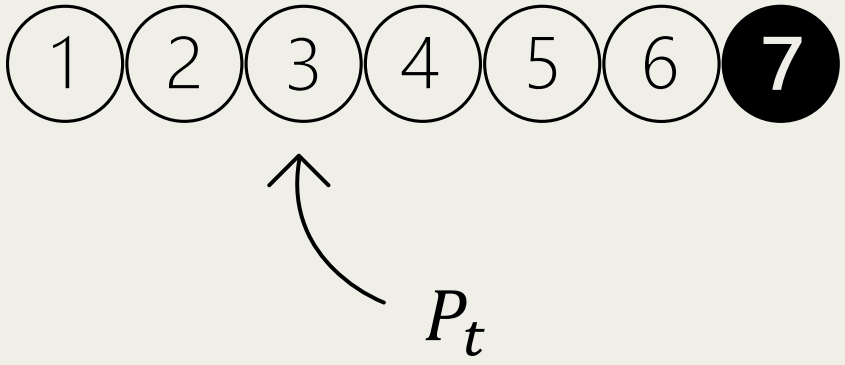


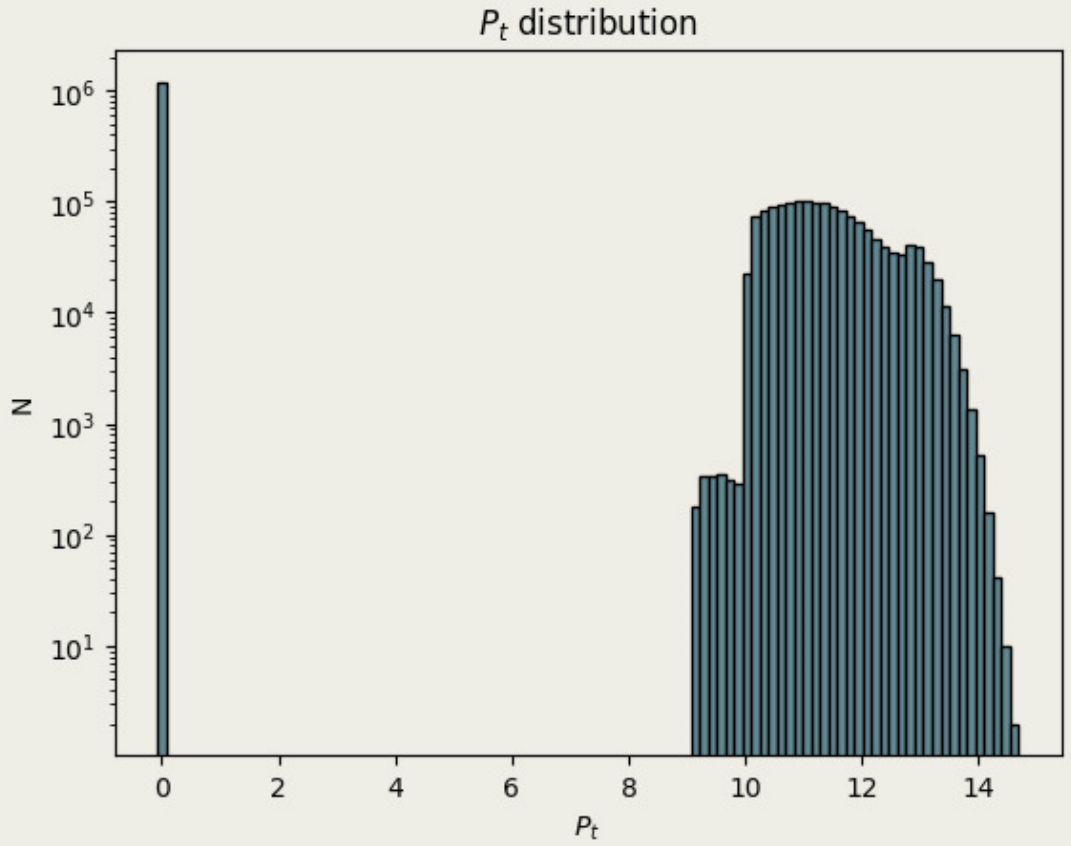


Focus on momentum

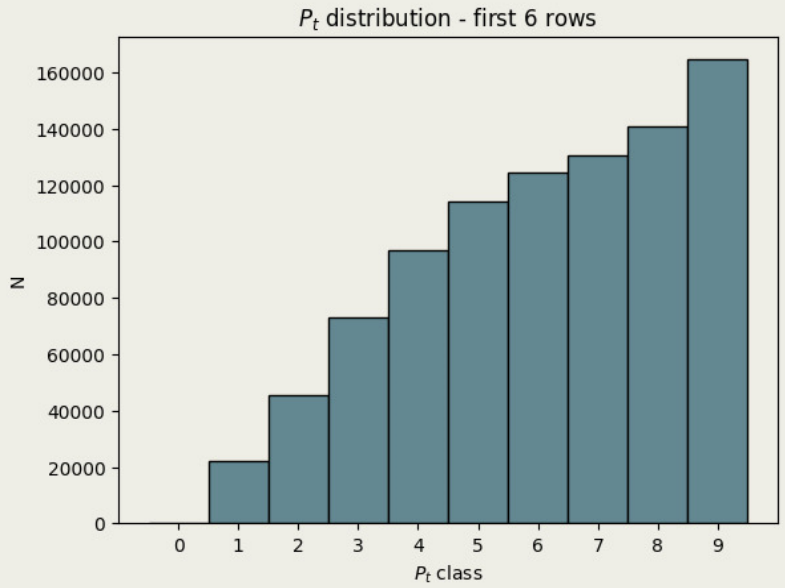
Continuous variable

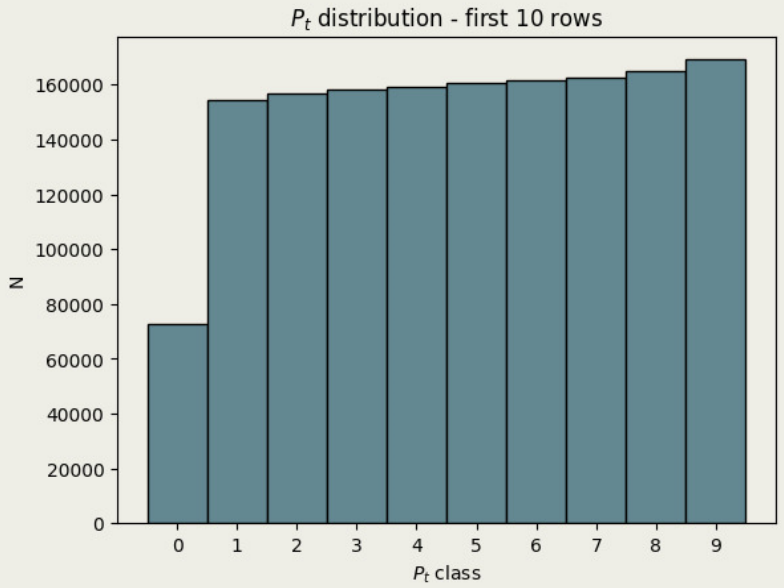
Discretize by binning

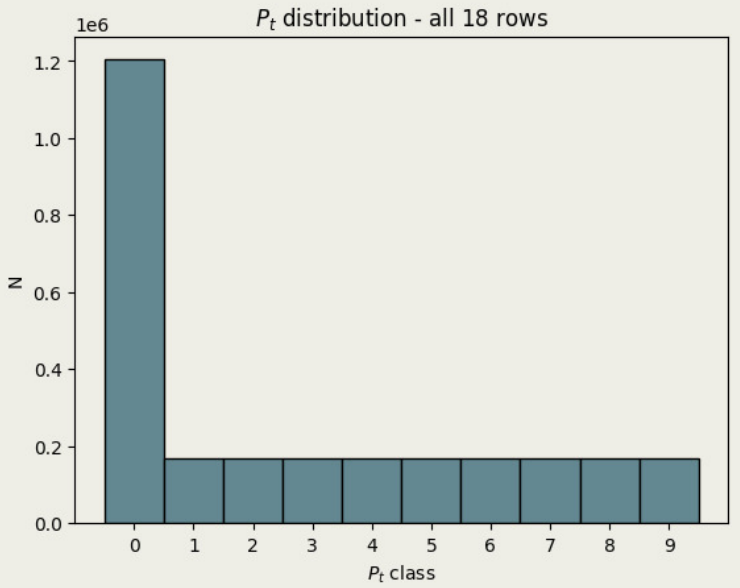




Focus on momentum







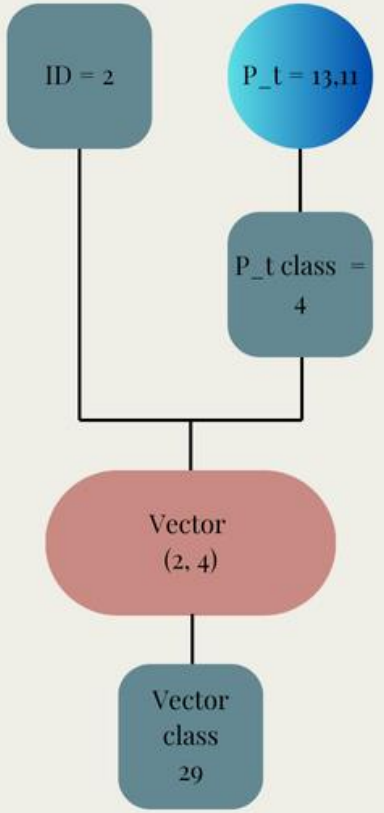
Combining into vectors

Only combining ID and P_t

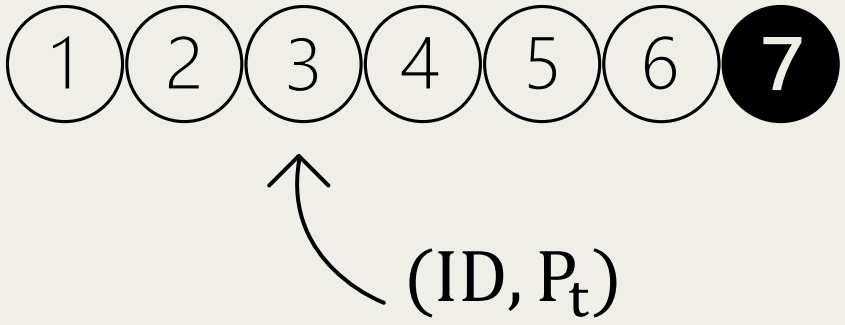
Use dictionary to covert each unique vector to a class.

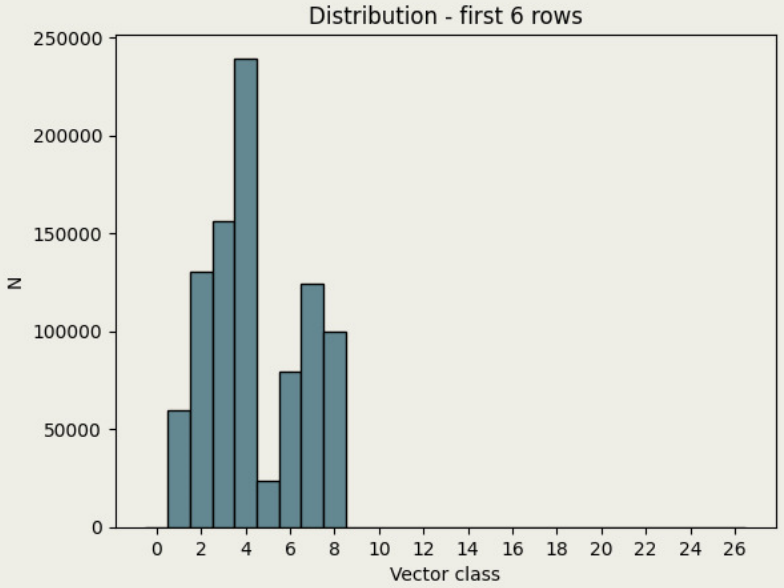
Class ordering should not matter

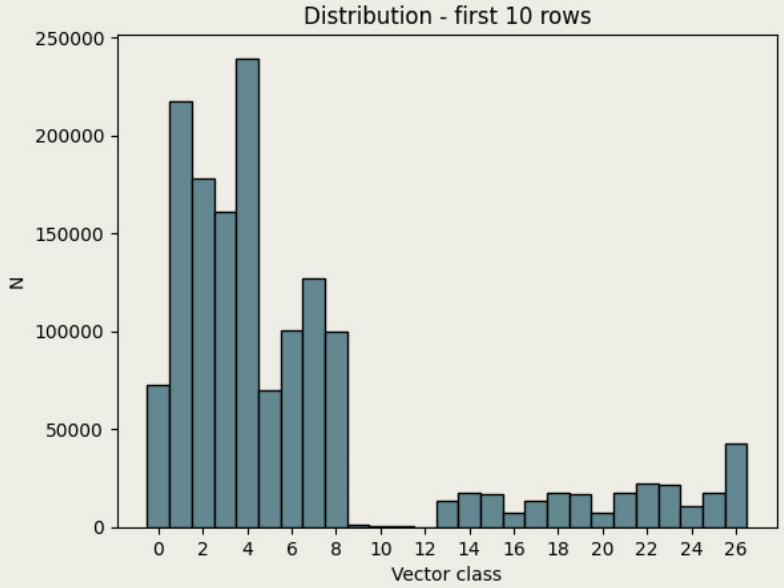
Train on predicting the correct vector class

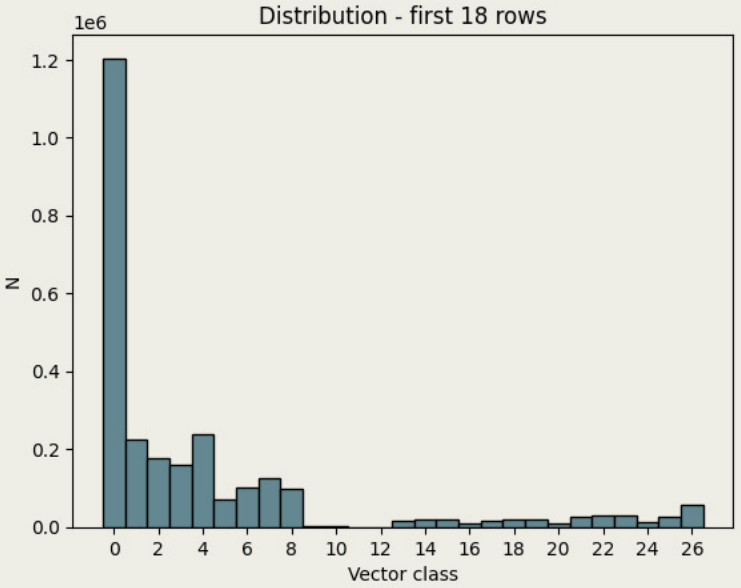


Combining into vectors







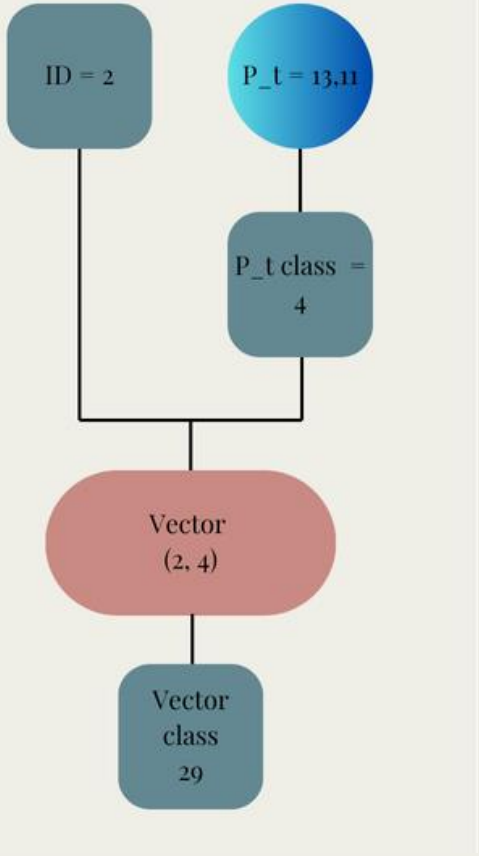


Inlcude more data into vectors

Can also include Energy, pseudorapidity and angle

Vector of length 3, 4 or 5

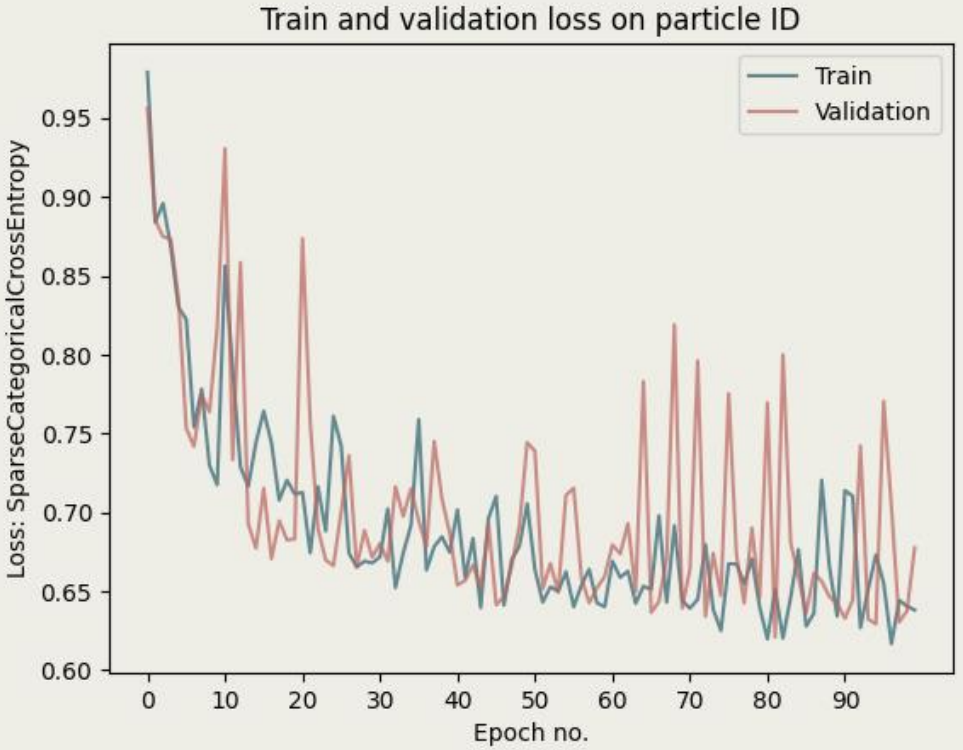
Increases ‘Vocabulary size’ and Embedding dimensions

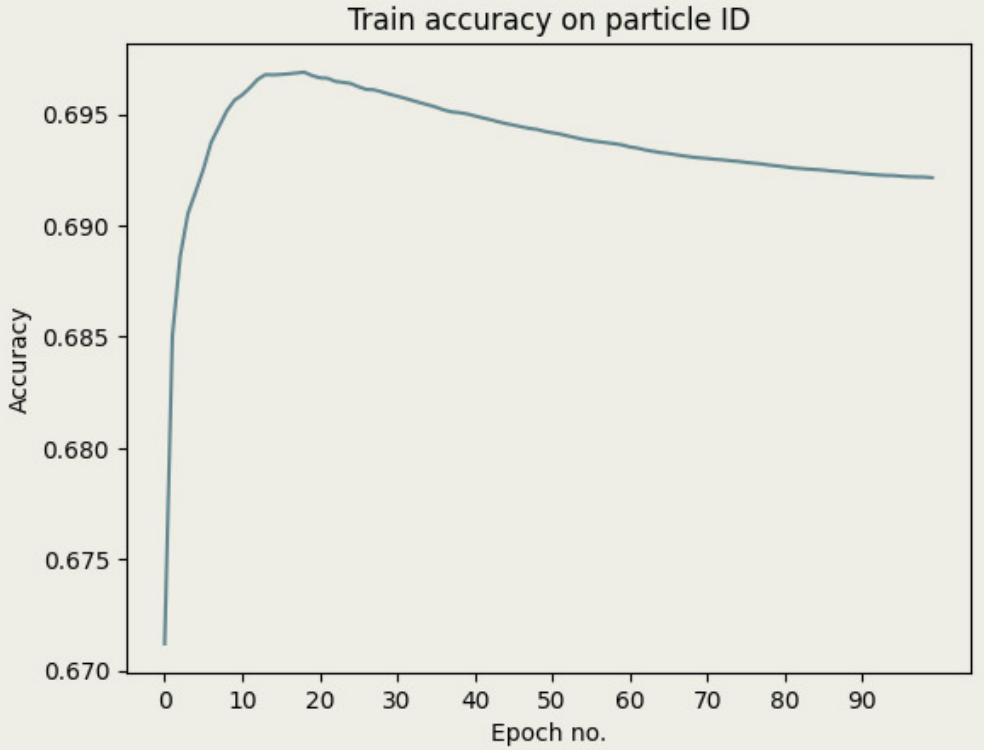


Sneak peak at the results

Sneak peak at the results – particle ID

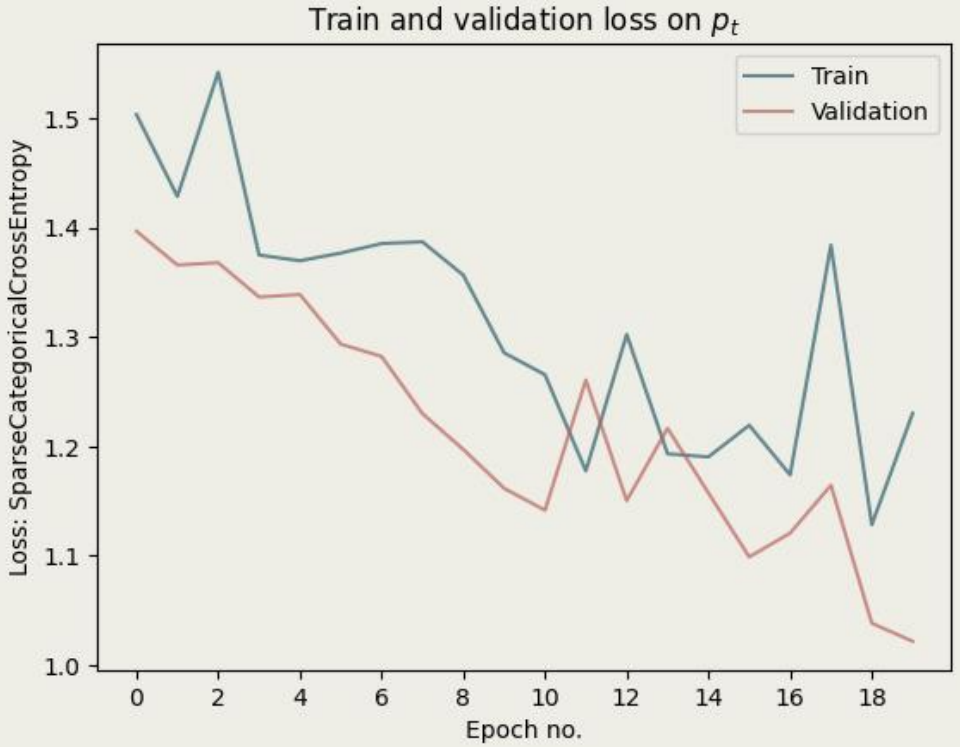
Can we predict the next particles type in the sequence of generated particles?

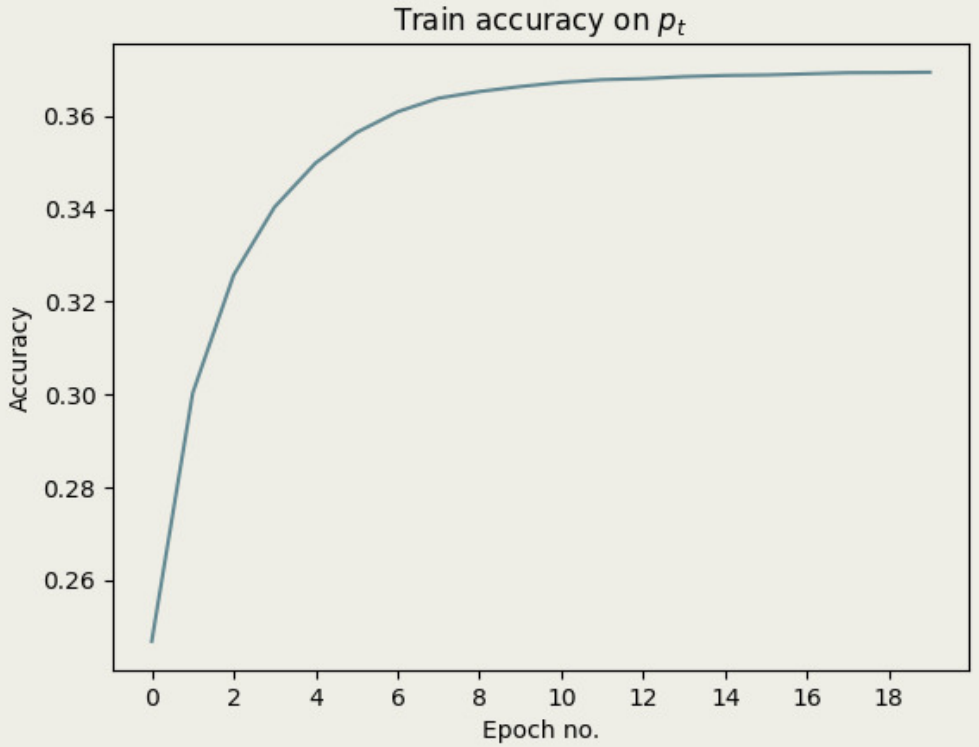




Sneak peak at the results – Momentum

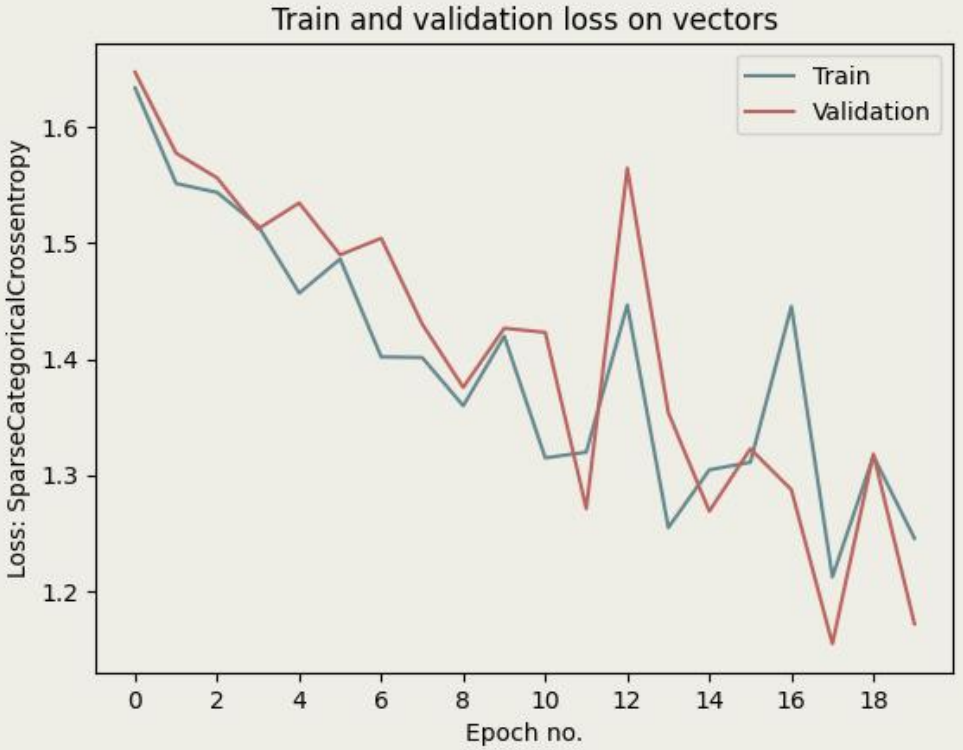
Can we predict the next particles momentum in the sequence of generated particles?

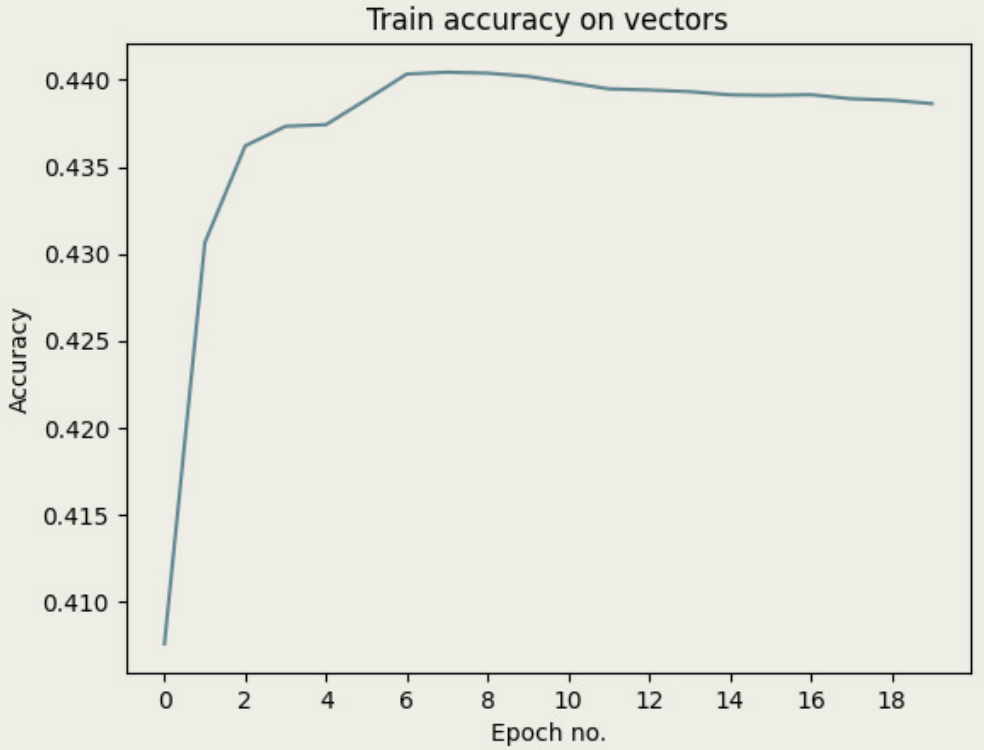




Sneak peak at the results – 2D Vector

Can we predict the next vector in the sequence of generated particles?





Sneak peak at the results – background

Does the loss increase on non 4top data?