This is a collection of particles parametrized by \( \theta \):

\[
E_i, x_i = \{ w_i^\theta, y_i^\theta \}_{i=1}^m
\]

Examples: drawn from \( \mathcal{N}(\mu, \sigma) \), uniform circles, disks, triangles, uniform rectangles, etc.

The event to fit given by a collection of particles:

\[
Q = \{ E_i, x_i \}_{i=1}^n
\]

Each particle is passed individually to a Lipschitz 1 neural network \( f_\phi(x) \) with parameters \( \phi \). Linear layers are weight normed such that

\[
\|W^{(l)}\|_{p,\infty} \prod_{l=2}^{L} \|W^{(l)}\|_{\infty} \leq 1
\]

To approximate the EMD well, the NN architecture needs to be a universal approximator of Lip 1 functions. So we use GroupSort activation.

Using the KR duality, a lower bound on the EMD \( \phi(\mathcal{P}_\theta, \mathcal{Q}) \) is estimated by the NN which is then minimized by the shape parameters in the following minimax optimization problem:

\[
O(Q) = \min_{\theta} \max_{\phi} \left[ \sum_{i=1}^n E_i f_\phi(x_i) - \sum_{i=1}^m w_i^\theta f_\phi(y_i^\theta) \right]
\]

\( O(Q) \) is the observable that describes how well the event \( Q \) is described by the class of geometric objects \( \mathcal{P} \).