

Loss functions for PES modelling

When training a model of the PES on some dataset, we typically have access to ground truth labels for both total energy, E and forces, \mathbf{F} , for a given system.

To train the model, performance on the test set, as measured by some loss function, is minimised using gradient descent (or some variant thereof).

It is thus vital that minimising the loss really does correspond to improving the model’s ability. Several possible pitfalls exist here. In the absence of any other agenda, we want to ensure that the quality of the final model’s predictions (and hence loss function used to train it) are independent of:

- **system size** : consider the case where a training set has some small (low number of atoms per unit cell) crystalline structures and some larger amorphous structures. If the loss function inadvertently evaluates to lower values (leading to smaller gradients and a smaller learning signal) for larger systems, the model will be preferentially optimised for performance on the smaller, crystalline structures at the expense of performance on the larger, amorphous structures.
- **composition** : consider an analogous case where the loss function inadvertently evaluates to lower values for some compositions than others (independently of how difficult the structures are to model). The model will be preferentially optimised for performance on some compositions at the expense of others, even in the case where all compositions are available in equal proportions in the training set.

Hence, we need to ensure that the loss function is invariant to these properties.

Total energy

If we model the total system energy, E , as a sum of local contributions, ε_i , and further assume that these local contributions are distributed about some mean, μ_z , with some variance, σ_z^2 , for a given species, z , then the following transform converts a raw system energy, E , into a unitless, standardized value, E' :

$$E' = \frac{E - \sum_i \mu_{z_i}}{\sqrt{\sum_i \sigma_{z_i}^2}} \sim \mathcal{N}(0, 1)$$

The expected value of an error metric applied to these standardized values, $\langle \text{metric}(\hat{E}', E'_{\text{true}}) \rangle$, will be independent of system size and composition.

Considering the single component case, this reduces to:

$$E' = \frac{E - N \cdot \mu}{\sqrt{N} \cdot \sigma} \sim \mathcal{N}(0, 1)$$

i.e. the error as applied to raw, unscaled energies is expected to scale with \sqrt{N} , not the commonly assumed N !

Forces

We can use a similar transform to convert raw per-atom force predictions, F_i , into standardized, unitless values, F'_i :

$$F'_i = \frac{F_i}{\eta_{z_i}} \sim \mathcal{N}(0, 1) \quad \forall z_i \in \mathcal{Z}$$

where η_{z_i} is a per-species scaling factor. Note that since atoms are a per-atom property, we do not need to worry about system size here.

Ratios

Using the above pre-metric transforms, weighting of the energy and force loss components can be done in a more principled manner, which depends a lot less on the specifics (e.g. units, compositions, system sizes) of the training set.