

# CORREL metrics

---

## Definition

Using model ensembles is a popular technique that allows to improve prediction accuracy and obtain additional statistical measures for every prediction. Suppose we have an ensemble of  $M$  predictors  $\{P_1^i, P_2^i, \dots, P_M^i\}$  for a molecule  $i$ . The ensemble average:  $1/M_k \sum_k P_k^i$  is frequently used as the ensemble prediction for the molecule.

**CORREL** measure uses ensemble predictions for molecules  $i, j$  to determine similarity of their ensemble responses. It is calculated as Pearson linear or Spearman rank correlation coefficients.

The main assumption of CORREL is that if the two molecules are under-predicted or over-predicted by the same ensemble models, then the ensemble response for these two molecules is "similar".

## References

CORREL is used in the [Associate Neural Network \(ASNN\) method \[Tetko 2008\]](#) to provide bias correction of ensemble predictions.