Ensemble learning

Nathanaël Carraz Rakotonirina

Mathématiques Informatique et Statistique Appliquées (MISA)
Université d'Antananarivo

Ensemble learning

- Ensemble methods combine multiple base models to get better performance than could be obtained from any of the base model alone.
- The ensemble will have similar bias to the base models, but lower variance, generally resulting in improved overall performance.

Basic approach

- **Regression:** The base models' predictions are averaged. The resulting model has the form
$$f(y|\boldsymbol{x}) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} f_m(y|\boldsymbol{x})$$
where f_m is the mth base model.
- **Classification:** A majority of the predictions is taken (committee method).

Stacking

- **Stacking** or **stacked generalization** learns how to combine the base models.
- It is composed of the base models and the **meta model**.
- The meta model take as input the predictions (real values for regression, probabilities or labels for classifications) of the base models on a separate dataset along with the ground-truth labels.
- The meta model is often a linear model (linear regression fore regression and logistic regression for classification). It can have the following form:

$$f(y|\boldsymbol{x}) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} w_m f_m(y|\boldsymbol{x})$$

Bagging

- **Bagging** or **bootstrap aggregation** is a form of ensemble learning which train different base models to different randomly sampled versions of the data.
- **Bootstrap** sampling is a statistical trick which samples with replacement until we have a total of N examples per model (an example may appear multiple times).
- On average, each model only sees 63% of the unique examples.
- The training instances that are not used by a given base model are called **out-of-bag** instances. They can be used to evaluate the model's performance instead of using cross validation.
- It prevents the ensemble from relying too much on any individual training example, which enhances robustness and generalization.
- It reduces the variance at the cost of increasing the bias.
- Bagging relies on the base estimators being unstable estimators (small changes in the dataset result in significant changes in model prediction) like decision trees.

Random forests

- It extends bagged decision trees by also randomizing the choice of features.
- **Random forests** train multiple decision trees based on a randomly chosen subset of features (at each node of the tree), as well as a randomly chosen subset of the dataset.
- This method tries to decorrelate the base learners even further.
- It can be trained in parallel.