Ensembles

Peter Tiño
School of Computer Science
University of Birmingham
The power of Parliament

Many ‘experts’ arguing about the same problem

The ‘experts’ can locally disagree, but overall, all of them try to solve the same global problem

The collective answer is usually more stable to fluctuations in intelligence/emotions/biases of individual ‘experts’

There exist many types of formalisations of this idea.

Peter Tiño
Formalising the Parliament

Instead of a single learning model, we’ll have a collection (or ensemble) of learning models that produce a single collective answer.

Ensembles can be categorized according to

- how the individual expert opinions are formed during the training
- how the individual expert opinions are combined to form a single collective answer in the work (testing) mode
- how localized/specialised the individual experts are allowed to become
Convex combination of ensemble members

Regression problem
\[ X \subseteq \mathbb{R}^n \text{ - input space} \]
\[ Y \subseteq \mathbb{R} \text{ - output space} \]

Ensemble members
\[ f_i : X \rightarrow Y, \quad i = 1, 2, ..., M \]

Ensemble
\[ \bar{f} : X \rightarrow Y, \]
\[ \bar{f}(x) = \sum_{i=1}^{M} v_i f_i(x), \]
where \( v_i \geq 0 \), and \( \sum_{i=1}^{M} v_i = 1 \).
Squared error at test input $x$

The ‘desired’ (mean) response from the environment:

$$\mu(x) = E_{P(y|x)}[y]$$

Squared error: $(\bar{f}(x) - \mu(x))^2$

To simplify the notation, don’t explicitly write the test input $x$

Squared error:

$$(\bar{f} - \mu)^2 = \left( \sum_{i=1}^{M} v_i f_i - \mu \right)^2$$
Squared error at test input $x$ - Continued

Evaluate this:

$$\sum_{i=1}^{M} v_i (f_i - \mu)^2 = (\bar{f} - \mu)^2 + \sum_{i=1}^{M} v_i (f_i - \bar{f})^2$$

$$\geq (\bar{f} - \mu)^2$$

One reason for preferring ensembles:

Squared error of the ensemble is never worse than the average of squared errors of the ensemble components

This is not the only point scored by ensembles …
Variance of uniform ensembles

Uniform ensemble: \( v_i = \frac{1}{M}, \ i = 1, 2, \ldots, M \)

\[
\bar{f}(x) = \frac{1}{M} \sum_{i=1}^{M} f_i(x),
\]

Variance:

\[
Var(\bar{f}) = \frac{1}{M} Var_E + \left(1 - \frac{1}{M}\right) Cov_E,
\]

where

- \( Var_E \) is the average variance of ensemble members, and
- \( Cov_E \) is the average covariance between pairs of distinct ensemble members.
Variance of uniform ensembles - Continued

\[ \text{Var}_E = \frac{1}{M} \sum_{i=1}^{M} E[(f_i - E[f_i])^2] \]

\[ \text{Cov}_E = \frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j \neq i} E[(f_i - E[f_i])(f_j - E[f_j])] \]

Note that

\[ \text{Var}(\bar{f}) = \frac{1}{M} \text{Var}_E + \left(1 - \frac{1}{M}\right) \text{Cov}_E, \]

is a homotopy between \( \text{Cov}_E \) and \( \text{Var}_E \):

The more ensemble members we have, the closer is \( \text{Var}(\bar{f}) \) to \( \text{Cov}_E \).
Variance of ensembles - good news

\[ \text{Bias}_E = \frac{1}{M} \sum_{i=1}^{M} (f_i - \mu) \]
is the average bias of ensemble members.

Squared error:

\[ E[(\bar{f}(x) - \mu(x))^2] = \text{Bias}_E^2 + \frac{1}{M} \text{Var}_E + \left(1 - \frac{1}{M}\right) \text{Cov}_E \]

\( \text{Bias}_E^2 \) and \( \text{Var}_E \) are always non-negative, but \( \text{Cov}_E \) can be negative!

Deal with Bias/Variance dilemma by using ensembles and making the ensemble members negatively correlated (e.g. NCL).