

Ensemble averaging (machine learning)

In machine learning, particularly in the creation of artificial neural networks, **ensemble averaging** is the process of creating multiple models and combining them to produce a desired output, as opposed to creating just one model. Frequently an ensemble of models performs better than any individual model, because the various errors of the models "average out."

Contents

Overview

Method

Benefits

See also

References

Further reading

Overview

Ensemble averaging is one of the simplest types of committee machines. Along with boosting, it is one of the two major types of static committee machines.^[1] In contrast to standard network design in which many networks are generated but only one is kept, ensemble averaging keeps the less satisfactory networks around, but with less weight.^[2] The theory of ensemble averaging relies on two properties of artificial neural networks:^[3]

1. In any network, the bias can be reduced at the cost of increased variance
2. In a group of networks, the variance can be reduced at no cost to bias

Ensemble averaging creates a group of networks, each with low bias and high variance, then combines them to a new network with (hopefully) low bias and low variance. It is thus a resolution of the bias-variance dilemma.^[4] The idea of combining experts has been traced back to Pierre-Simon Laplace.^[5]

Method

The theory mentioned above gives an obvious strategy: create a set of experts with low bias and high variance, and then average them. Generally, what this means is to create a set of experts with varying parameters; frequently, these are the initial synaptic weights, although other factors (such as the learning rate, momentum etc.) may be varied as well. Some authors recommend against varying weight decay and early stopping.^[3] The steps are therefore:

1. Generate N experts, each with their own initial values. (Initial values are usually chosen randomly from a distribution.)
2. Train each expert separately.
3. Combine the experts and average their values.

Alternatively, domain knowledge may be used to generate several *classes* of experts. An expert from each class is trained, and then combined.

A more complex version of ensemble average views the final result not as a mere average of all the experts, but rather as a weighted sum. If each expert is y_i , then the overall result \tilde{y} can be defined as:

$$\tilde{y}(\mathbf{x}; \alpha) = \sum_{j=1}^p \alpha_j y_j(\mathbf{x})$$

where α is a set of weights. The optimization problem of finding alpha is readily solved through neural networks, hence a "meta-network" where each "neuron" is in fact an entire neural network can be trained, and the synaptic weights of the final network is the weight applied to each expert. This is known as a *linear combination of experts*.^[2]

It can be seen that most forms of neural networks are some subset of a linear combination: the standard neural net (where only one expert is used) is simply a linear combination with all $\alpha_j = 0$ and one $\alpha_k = 1$. A raw average is where all α_j are equal to some constant value, namely one over the total number of experts.^[2]

A more recent ensemble averaging method is negative correlation learning,^[6] proposed by Y. Liu and X. Yao. Now this method has been widely used in evolutionary computing.

Benefits

- The resulting committee is almost always less complex than a single network which would achieve the same level of performance^[7]
- The resulting committee can be trained more easily on smaller input sets^[1]
- The resulting committee often has improved performance over any single network^[2]
- The risk of overfitting is lessened, as there are fewer parameters (weights) which need to be set^[1]

See also

- Ensemble learning

References

1. Haykin, Simon. Neural networks : a comprehensive foundation. 2nd ed. Upper Saddle River N.J.: Prentice Hall, 1999.
2. Hashem, S. "Optimal linear combinations of neural networks." Neural Networks 10, no. 4 (1997): 599–614.
3. Naftaly, U., N. Intrator, and D. Horn. "Optimal ensemble averaging of neural networks." Network: Computation in Neural Systems 8, no. 3 (1997): 283–296.
4. Geman, S., E. Bienenstock, and R. Doursat. "Neural networks and the bias/variance dilemma." Neural computation 4, no. 1 (1992): 1–58.
5. Clemen, R. T. "Combining forecasts: A review and annotated bibliography." International Journal of Forecasting 5, no. 4 (1989): 559–583.
6. Y. Liu and X. Yao, Ensemble Learning via Negative Correlation (<https://www.sciencedirect.com/science/article/pii/S0893608099000738>) Neural Networks, Volume 12, Issue 10, December

1999, pp. 1399-1404. doi:10.1016/S0893-6080(99)00073-8 (<https://doi.org/10.1016%2FS0893-6080%2899%2900073-8>)

7. Pearlmuter, B. A., and R. Rosenfeld. "Chaitin–Kolmogorov complexity and generalization in neural networks." In *Proceedings of the 1990 conference on Advances in neural information processing systems 3*, 931. Morgan Kaufmann Publishers Inc., 1990.

Further reading

- Perrone, M. P. (1993), *Improving regression estimation: Averaging methods for variance reduction with extensions to general convex measure optimization*
 - Wolpert, D. H. (1992), "Stacked generalization", *Neural Networks*, **5** (2): 241–259, CiteSeerX [10.1.1.133.8090](https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.133.8090) (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.133.8090>), doi:10.1016/S0893-6080(05)80023-1 (<https://doi.org/10.1016%2FS0893-6080%2805%2980023-1>)
 - Hashem, S. (1997), "Optimal linear combinations of neural networks", *Neural Networks*, **10** (4): 599–614, doi:10.1016/S0893-6080(96)00098-6 (<https://doi.org/10.1016%2FS0893-6080%2896%2900098-6>), PMID [12662858](https://pubmed.ncbi.nlm.nih.gov/12662858) (<https://pubmed.ncbi.nlm.nih.gov/12662858>)
 - Hashem, S. and B. Schmeiser (1993), "Approximating a function and its derivatives using MSE-optimal linear combinations of trained feedforward neural networks", *Proceedings of the Joint Conference on Neural Networks*, **87**: 617–620
-

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Ensemble_averaging_\(machine_learning\)&oldid=929430360](https://en.wikipedia.org/w/index.php?title=Ensemble_averaging_(machine_learning)&oldid=929430360)"

This page was last edited on 5 December 2019, at 19:52 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.