

hyperparameter

Authors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland

*Corresponding author: alex.jung@aalto.fi

Abstract

A hyperparameter of a machine learning (ML) method is a quantity that selects one model from a family of models, or one variant of a training method, and that is fixed before training starts. Examples are the learning rate of a gradient-based method, the number of features of a linear model, the maximum depth of a decision tree, and the number of layers of a deep net. Whereas the model parameters are learned by minimizing the training error, a hyperparameter is chosen by minimizing the validation error, since the training error would favor the largest model. The choice is therefore a higher-level form of empirical risk minimization (ERM) on the validation set, carried out by a search over candidate values.

Definition

A hyperparameter associated with a machine learning (ML) method is a quantity that is used to select among a family of models. Typical examples include the learning rate used in a gradient-based method, the number of features used in a linear model, or the maximum depth of a decision tree. The usefulness of a specific hyperparameter choice can be assessed via validation. Similar to learning (or tuning) model parameters by empirical risk minimization (ERM) on a training set, hyperparameters can be learned (or tuned) via minimizing the validation error. Thus, in a sense, hyperparameters are higher-level model parameters that are learned via a higher-level form of ERM, i.e., minimizing the validation error obtained by the trained model with a given hyperparameter value. For example, when training a deep net, hyperparameters such as the learning rate, the number of layers, and the batch size must be chosen before training begins, and their values critically affect both training speed and generalization performance.

Cross References

model, validation, model parameter

References

- 1.