Write at least 300 words explaining why ensemble is a successful strategy in machine learning.

The ensemble method is a combination of several machine learning models that produce a better predictive performance than utilizing a single model. The efficiency of this type of machine learning method relies on its combination of several machine learning techniques that can help decreasing the variance (bagging methods), bias (boosting method), or improving the predictions (stacking).

The bagging technique is a form of bootstrap aggregation method that can reduces the variance of an estimate by constructing several trees that use voting for classification and averaging for regression in order to achieve a better prediction. A type of method that is found inside the bagging technique is random forest, which is a type of ensemble that built trees using bootstrap sample from the training set. The difference in this case is that instead of using all the available features, with random forest a subset of features is selected. By doing this, the trees are further randomized and as a result the bias increases slightly, but with less variance (James et al., 2013).

On the other hand, the boosting technique (which decreases bias) fits a sequence of weak learners to weighted versions of the data. “The general idea of most boosting methods is to train predictors sequentially, each trying to correct its predecessor. There are many boosting methods available, but by far the most popular are AdaBoost (short for Adaptive Boosting) and Gradient Boosting.” (Géron, 2017)

Finally, stacking is an ensemble technique that trains the model based on a complete training set, and the results are then used as features to train the subsequent models. By doing this, we end up with a “(…) set of k first-level models used to transform the feature space, and a combiner classifier at the second-level. For a test instance, the first-level models are used to create a new k-dimensional representation. The second-level classifier is then used to predict the test instance.” (Aggarwal, 2015)

Although ensembles are universally use with decisions trees, many other applications in machine learning are starting to ensemble or combine several methods that allows the machine to make a better prediction by reducing the generalization error. Example of this can be seen in deep learning, where stacking techniques are being use in other to have a better model.

Aggarwal, C. C. (2015). *Data Mining*. Springer International Publishing. https://doi.org/10.1007/978-3-319-14142-8

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7