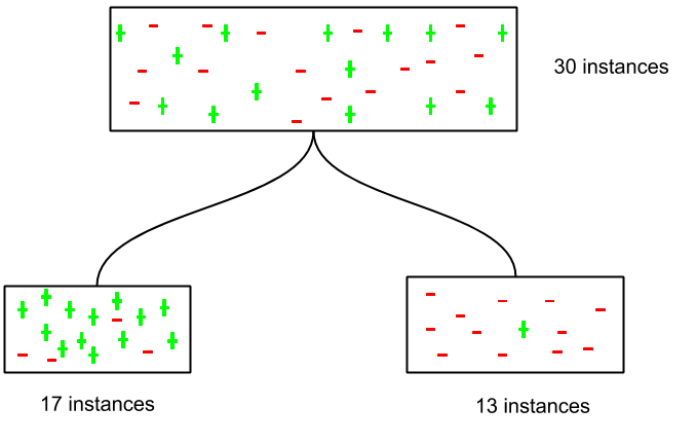
Fiche d'auto-évaluation 05

May 6, 2021

1. Consider training a neural network. Draw an example plot of training and validation loss vs. epochs in which training ended too soon (it would have been better to continue with more epochs).
2. Draw a plot like the previous, assuming now that your resulting model overfits the training dataset.
3. Draw a plot like the previous, assuming now that your resulting model underfits the training dataset
4. Are there cases, in Machine Learning, when gradient descent is guaranteed to converge?
5. Tell if the following way of applying scaling (Min-Max scaling, in this case) is correct or not. If not, correct it.
   (a) Partition the dataset $\mathcal{D}$ in training dataset $\mathcal{D}^{\text{train}}$ and test dataset $\mathcal{D}^{\text{test}}$.
   (b) Put the test set aside.
   (c) Compute the minimum value $\min_j^{\text{train}}$, the maximum value $\max_j^{\text{train}}$ and average μ_j^{train} of each column j on the training set.
   (d) Transform all original values $x_j^{(i)}$ in $\frac{x_j^{(i)} - \dots}{\dots - \dots}$ (Complete the formula).
   (e) Train a Machine Learning Model on the transformed training set.
   (f) Take the test set $\mathcal{D}^{\text{test}}$, compute the minimum value $\min_j^{\text{test}}$, the maximum value $\max_j^{\text{test}}$ and the average μ_j^{test} of each column j on the test set $\mathcal{D}^{\text{test}}$.
   (g) Transform $\mathcal{D}^{\text{test}}$ with a similar formula as before, now using $\min_j^{\text{test}}, \max_j^{\text{test}}, \mu_j^{\text{test}}$
   (h) Use the trained model to make predictions on the transformed test set.
6. If we apply 7-fold cross validation, how many models do we need to train?
7. You want to do Model selection: you want to use a type model, say Polynomial Regressions, and you want to select the best hyper-parameters

(give some example of hyper-parameter). How can you use grid search and cross-validation together? Explain the correct procedure, step by step. (this procedure is implemented in sklearn-model_selection.GridSearchCV).

8. When using Random Forests, are there cases where, increasing the number of trees, the accuracy of the forest decreases? If yes, which case?
9. What is the difference between random forest, bagging trees and extra-trees?
10. If you use a set of trees to do your prediction, is it better they are all similar or is it better they are different?
11. What is the CART algorithm used for? Describe it step by step. Then, explain how CART algorithm is integrated in a random forests, writing a pseudo-code.
12. If we increase the number of trees in a random forest, does the variance of the model increase or decrease?
13. Is a decision tree a linear or non-linear classifier? And a random forest?
14. In which sense a random forest is interpretable? (Give at least two arguments)
15. How can you compute the feature importance based on an ensemble of trees?
16. Is it necessary or at least useful to scale the dataset before using a decision tree or an ensemble of trees?
17. What does "linear classifier" mean?
18. Consider the following split[1]



[1] Picturefromhttps://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/ml-decision-tree/tutorial/

Compute the information gain of the split, considering entropy and Gini impurity index.

19. Can information gain be negative?

20. Why is pruning important? For training or testing? What are the criteria that can be used to prune a tree?

21. How is randomness used in Bagging, Random Forests and Extra-trees?

22. Is the concept of bagging only limited to tree-based learning? Or can it also be applied with other types of models?