Instituto Superior de Agronomia, ULisboa, Green Data Science

Practical Machine Learning/Aprendizagem Automática Aplicada

Questionnaire #9, May 26, 2023

Name:_____

Topic: Ensemble methods

Input: IDT(a decision tree inducer),T(the number of iterations), S(training set), μ(the subsample size), N(number of attributes used in each node)

Output: M_t:$\forall t = 1, ..., T$

for *each* t *in* $1, ..., T$ **do**

$S_t \leftarrow$ Sample μ instances from S with replacement.

Build classifier M_t using $IDT(N)$ on S_t

$t++$

end

Figure 1: pseudo-code to create an ensemble of trees with Random Forest

Given: $(x_1; y_1), \ldots, (x_m; y_m)$, $x_i \in X$, $y_i \in Y = \{-1, 1\}$.
Initialize $D_1(i) = 1/m$.
For $t = 1 \ldots T$:

1. Train weak classifier using distribution D_t.
2. Get weak hypothesis $h_t : X \rightarrow \{-1, 1\}$ with error $\varepsilon_t = \sum_{i:h_t(x_i) \neq y_i} D_t(x_i)$
3. Choose $\alpha_t = \frac{1}{2} \log(\frac{1-\varepsilon_t}{\varepsilon_t})$
4. Update:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \cdot = \begin{cases} e^{-\alpha_t} & \text{if instance } i \text{ is correctly classified} \\ e^{\alpha_t} & \text{if instance } i \text{ is not correctly classified} \end{cases}$$

where Z_t is a normalization factor (chosen so that $\sum_{i=1}^{m} D_{t+1} = 1$).

Figure 2: pseudo-code to create an ensemble of classifiers with AdaBoost

1. What is called a “weak learner” or a “weak classifier”? Give an example based on a decision tree.

Response: A weal learner is a classifier with a very simple structure that is expected to perform better than random. For trees, and example of a weak learner is a tree stomp, that defines a single split.

2. Consider the AdaBoost pseudo-code where each iteration creates a weak classifier at step 1 and distribution D_t indicates the weight given to each example.
    a. Are the sequence of classifiers dependent or independent, i.e. does classifier T_{i+1} depend on classifier T_i ? Justify your answer.

Response: The sequence of classifiers is dependent. Classifier T_{i+1} depends on classifier T_i since the distribution D_t is updated as result of the errors of classifier T_i. Then T_{i+1} uses the updated distribution, so it depends on T_i.

    b. The AdaBoost algorithm is based on the assumption that each (weak) classifier is better than random, which implies that $\alpha_t > 0$. What is the rationale for weight update in AdaBoost? Refer to the pseudo-code for your response.

Response: If the example is correctly classified its weight (before normalization) is $\exp(-\alpha_t)$ and therefore it is less than 1. If it is not correctly classified, the weight is $\exp(\alpha_t)$, which is larger than 1. Therefore, the misclassified examples will end up having a larger weight than the correctly classified examples in the following iteration.

3. AdaBoost computes a “weighted loss”. Indicate what is the expression of the weighted loss in the pseudo-code.

Response: $\varepsilon_t = \sum_{i:h_t(x_i)\neq y_i} D_t(x_i)$ is a loss function since it compares true labels with predicted labels. It is weighted by the distribution of weights given to examples on the t-th iteration.

4. Both Random Forest and AdaBoost are considered to iterate over different datasets. Explain how those different datasets are formed:

    a. For Random Forest

Response: Random Forest creates different data sets by resampling the set of examples with replacement (bootstrap)

    b. For AdaBoost

Response: For AdaBoost, the new datasets are created by changing the weights given to each example in each iteration