Machine Learning and Data Mining project: Leaf identification

Michele Scomina[1]

[1] problem statement, solution design, solution development, data gathering, writing

Course of AA 2023-2024 - DIA - Ingegneria Elettronica e Informatica (IN20)

1 Problem statement

1.1 Goal of the project

The goal of this project is to identify the type of leaf based on a group of features, which are derived from both the shape and texture of the leaf. The given solution must be able to classify the leaves in one of 30 categories of leaves.

1.2 Formal definitions

The problem can be formally defined as follows:

- Let X be the set of all possible leaves. Every element $x \in X$ is a leaf defined by a 14-dimensional attribute vector, representing various shape and texture features of the leaf.

- Let Y be the set of all possible categories of leaves. Every element $y \in Y = \{0, 1, 2, ..., 29\}$ represents a unique category of leaves.

- Let $f : X \to Y$ be the function that assigns each leaf to its category based on its attributes.

- Let $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be the training set, where $x_i \in X$ and $y_i \in Y$.

- The goal is to find a function $f_{\text{predict}} : X \times M \to P_Y$ that approximates f as closely as possible, effectively classifying the leaves into their correct categories, where M is a model learned from the training set D and P_Y is the set of all discrete probability distributions over Y.

2 Assessment and performance indexes

Since classes are balanced, but are composed of only a few samples, there's a risk of ill-defined metrics, as the model might not be able to predict some classes and not have a defined precision, recall or F1-score for those classes. Therefore, the weighted accuracy and the AUC (One-VS-Rest) will be used as performance indexes.

3 Proposed solution

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

4 Experimental evaluation

4.1 Data

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi.

Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

4.2   Procedure

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

4.3   Results and discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl,

n_{this}	n_{that}	LF	TF	#I	#S [$\times 10^6$]	t_l [s]
	10	0.37	0.45	552	0.59	52
50	25	0.43	0.44	3076	0.56	245
	50	0.45	0.43	637	0.64	715
	10	0.34	0.50	1138	2.76	110
100	25	0.40	0.48	1224	0.94	326
	50	0.38	0.49	443	0.44	1056

Table 1: Results (including learning time t_l) for different values of n_{this} and n_{that}.

blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec euismod nibh vitae elit dignissim, a maximus justo mattis. Cras elementum dolor at volutpat lacinia. Duis a magna tempus, lobortis arcu eget, blandit augue. Nulla facilisi. Sed nibh lorem, tempus eget pellentesque vel, sagittis id quam. Mauris interdum faucibus vulputate. Nunc a ex tincidunt, scelerisque turpis a, dictum metus. Nunc nec dui dui. Ut mollis massa ligula, at pretium nibh tristique at. Mauris suscipit fermentum congue. Vivamus in auctor nibh. Vestibulum magna nisl, blandit luctus eleifend vitae, pharetra et ipsum. Duis faucibus vulputate mauris, eu viverra elit mattis vel.