

5b. ML 2: Optimal classifier

The machine learning task has two parts - **symbol classification** and **determination of the optimal classifier parameter**. Check the [Upload system \[https://cw.felk.cvut.cz/upload/\]](https://cw.felk.cvut.cz/upload/) for the due date and notice that both tasks are submitted **separately**. It is expected that the classifiers will be implemented by students, usage of advanced pattern recognition libraries is forbidden. If unsure, ask the TA.

Selection of optimal classifier



Very often, there are many classifiers that can be used for a ML task, and we have to make a decision about which is the best classifier for the task. In the zip archive [classif_result_tables.zip](https://wiki/_media/courses/b3b33kui/cviceni/strojove_uceni/classif_result_tables.zip) [\[wiki/_media/courses/b3b33kui/cviceni/strojove_uceni/classif_result_tables.zip\]](https://wiki/_media/courses/b3b33kui/cviceni/strojove_uceni/classif_result_tables.zip), there are all files that will be used for the task. The task shall be solved in Python. What is asked to be uploaded are a **pdf report** and one function related to the part **Safety first**.

We have 5 different learned binary classifiers. The result of the classification of each classifier depends on the value of the α parameter. Thus, the result of the classification of a given classifier can be expressed as a function $C(\mathbf{x}, \alpha) \in \{0, 1\}$, where \mathbf{x} is a vector which belongs to the sample we want to classify.

We tested all classifiers on a test set $X = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{99}\}$. At the same time, we tried all possible values of $\alpha \in \{\alpha_0, \alpha_1, \dots, \alpha_{49}\}$. For a classifier $i \in \{1, 2, \dots, 5\}$ we obtain a table with values $C_i(\mathbf{x}_j, \alpha_k) \in \{0, 1\}$, where $j \in \{0, 1, \dots, 99\}$, $k \in \{0, 1, \dots, 49\}$ (see $C1, C2, C3, C4, C5$ in [classif_result_tables.zip](#)). The real labels of the samples $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{99}$ from the test set are available (see GT in [classif_result_tables.zip](#)).

1 Selection of appropriate parameter

In this section, suppose that the classifiers are used for binary classification of images (e.g. whether a dog is on the picture or not). For the classifier 1 (table $C1$), determine the best value for parameter $\{\alpha_0, \alpha_1, \dots, \alpha_{49}\}$. Be aware that you don't know the concrete task for which the classifier will be used. Therefore, it is necessary to use a sufficiently general approach. In other words, the classifier should not be one that is optimal for a particular task but globally inefficient for most other tasks.

In a short (one A4 page) **pdf report** explain the choice of the parameter (use terms such as sensitivity, false positive, ROC curve etc.). Inside the report, put the figure of a ROC curve with a marked point on the curve which correspond to the optimal value of the parameter.

2 Top secret!

Imagine, that you are Agent 00111 and you want to use your fingerprint to secure some top secret documents. The data is very sensitive, so it is better to delete it than secure it poorly. You also know

that you will always have enough time to unlock the data. Five trained classifiers (with different α values) are available. The input of the classifier is a fingerprint scan. For your fingerprint, desired output of the classifier is 1 (data will be unlocked), 0 otherwise (if it is not your fingerprint). All classifiers were tested using the test set X for all possible values of the parameter α . Results of the test for the classifiers are saved in tables $C1$, $C2$, $C3$, $C4$, $C5$ (see above). Ground truth values (real fingerprint affiliation) of the different scans are also available (see GT)

Select the most suitable classifier and its α parameter.

In the **pdf report** write your choice and explain the criterias you used for the choice.

3 Safety first

This part is a continuation of the previous part **Top secret!**. A colleague, also an agent, will send you his classifier which also depends on the parameter α . However, you are not sure about his loyalty, he may be a double agent. Thus, it will be necessary to find if his classifier is better than the classifier you selected in the previous section.

For security reasons, you will have to make the decision about his classifier using a function that will be created in advance. Input of the function will be table $C6$ with the results of the classification on the set for different α parameters (same format as $C1$, $C2$, etc.) and eventually other input parameters of your choice. The output of the function should be the decision if the new classifier is better than the one that you selected yourself (**true** if the obtained classifier is better than the previous one, **false** otherwise). In the **pdf report** explain the criterias that the function use. Submit also the function.

Selection of optimal classifier -- Evaluation

- Send a PDF report where you determine the optimal parameter. Based on the chosen parameter and the method of choosing the parameter you receive the points. [0–5 points]

[courses/be5b33kui/semtasks/05_ml2/start.txt](#) · Last modified: 2024/02/20 10:28 by xposik

Copyright © 2024 CTU in Prague | Operated by [IT Center of Faculty of Electrical Engineering](#) |
Bug reports and suggestions [Helpdesk CTU](#)