André Artelt, Johannes Brinkrolf Bielefeld University

Introduction to Machine Learning (WS 2019/20)
6. Assignment

Released: Friday, 06.12.2019.
Due: Please solve the exercises in groups of three and submit your report by mail to your tutor or into post box 329 on CMD 5 in the main building by **Friday, 17.01.2020**, **11:59am**.

- Write all names of your group members and the tutor's name on the first page (have a look at given template).
- Mark each paragraph with the initials of the responsible group member. You find a new command `\initials{}`, therefore, in the template.
- The python toolbox needed for solving the practical exercises can be found at `http://scikit-learn.org/stable/`. It comes with an excellent online description and demos. You can also look at the given demo code.
- We recommend to install `anaconda` for this lecture because it install automatically all required package, namely `numpy`, `scipy`, `matplotlib`, `sklearn`, `jupyter`, and `Spyder` as a develoment environment.
- You have five weeks time for this sheet . Also there is a tutorial on 06.01.2020 where no other sheet will be discussed. You can use this session to work on this sheet and/or ask your tutor if you get into trouble.
- If you have any questions, please ask your tutor or write an email to intromachlearn@techfak.uni-bielefeld.de.

---

1 Beautiful pictures

(*10 Points*)

In this exercise you work with the *André's vacation pictures* (AVP) data set - available in the learning space as `avp_dataset.npz`. The goal is to build and train a ML model that performs as well as possible. The given input are RGB colored images. This is why you have to apply some pre-processing and extract features from the raw data. We provide some starter code `starter.py` for loading the data set and plotting the pictures.

We propose to use the following feature extraction: Place a grid on top of the image. Compute the mean pixel value for each grid cell. Concatenate the mean pixels of all grid cells into a single feature vector.

Your task is then to write down a short report which should include at least following sections:[1]

(a) *(2 Points)* Introduction and description of the given data - e.g. analyze the given data sets and answer the following questions: How many samples? How many classes? Which features are used? ...

(b) *(3 Points)* Describe shortly the classifiers you used (e.g. How does the classifier distinguish classes? How are hyperparameters chosen? Which hyper-parameters were used? ...)

(c) *(3 Points)* Show and describe your results (e.g. accuracy on train and test set) by using plots or tables with test errors. Note that accuracy is not the only measure you can use - think about balanced and unbalanced data sets.

(d) *(2 Points)* Discussion (Try to explain and discuss the results. Any further ideas?)

Prepare your results in a **full-text** report on two or three pages. We provide a LaTeX template `report_template.tex` which might help you.

You can come with further features which enable better results. Explain the features and compare them with the others in the report. This will give you up to *5 bonus points*.

ATTENTION: Only models known from the lecture are allowed!

[1] You might want to use this ordering for working on the task.

2 Competition (optional)

(*10 Points*)

To make things more interesting, we are hosting a competition where you can get ten bonus points if you perform well. Your goal is to use the data set from task 1 for building a classifier that generalizes as well as possible. We will use a secret test set for evaluating your model. You will get *10 bonus points* if your model performs at least as good as the model of the tutors[2]. Furthermore, your names will be added to the hall-of-fame[3].

Further details and guidelines can be found in `competition.pdf`.

Please note that participating in this competition is **not** mandatory!

ATTENTION: Only models known from the lecture are allowed!

[2] Of course the tutors do not have access to the secret test set!

[3] The hall-of-fame will be available in the learning space.