Machine Learning — Programming Assignment
Dice rolls

Claudio Cusano

Deadline 17–7–2020, 16.00

1 Problem definition and data

A Casino needs a software for monitoring dice games. We have to design a system that, given a image of a single six-sided die, automatically determines the outcome of the roll (a number in the 1–6 range). For this purpose a data set of images has been collected. The data set includes 400 grayscale images for each of the six possible outcomes. Images have been divided in training, vaidation and test sets with 80/20/20 proportions. The first part of the name of the image files encodes the outcome of the die. The image here below shows some examples of images in the data set:
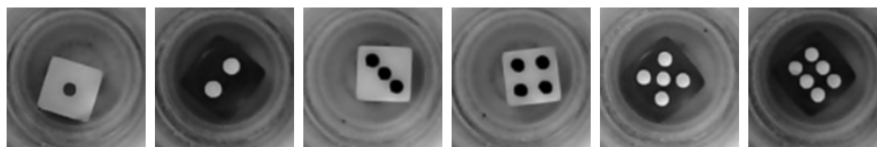

Image data can be read in a Python script as follows:

```python
import matplotlib.pyplot as plt
plt.imread("training/1-000.png")
```

Beware that, depending on the version of your libraries, you may get pixel values in the $[0, 1]$ or $[0, 255]$ range.

Warning: this problem can be "easily" solved by using suitable image processing operators. Here we want to tackle it with a machine learning approach, even if it may result less effective.

2 Assignment

For the programming assignment you are expected to:

1. design and implement a feature extraction procedure;

2. implement, train and evaluate one or more classification models;

3. use suitable data processing and visualization techniques to analyze the behavior of the trained model(s).

All the above should be implemented as scripts in the Python programming language. Any machine learning library (included `pvml`) can be used. In particular you can reuse the code developed during the lab activities. The image_features.py module is shipped with the data (but you may choose to not use it!).

3 Report

Prepare a report of three to five pages documenting all your work. Provide detailed instructions on how to reproduce the results. The report must be in the PDF format. Include your name in the report and conclude the document with the following statement: "I affirm that this report is the result of my own work and that I did not share any part of it with anyone else except the teacher."

Make a ZIP archive with the report and the Python scripts, and send it by e-mail before the deadline. To keep the size of the submission manageable, do not include files containing the original data, the features etc.

Deadline for the submission: Friday, July 17 at 16.00.