

Project-II for PCML

PCML-Home Dates&Deadlines

Project-II in 5 easy steps

STEP 1: Understand real-world tasks.

STEP 2: Understand data and do fun ML stuff.

STEP 3: Predict test data.

STEP 4: Write your finding in a report.

STEP 5: Submit your report, your predictions, and your code.

News

Nov. 24, 2015: Information about the project is online.

Important dates and deadlines

These dates will be updated from time to time.

Nov. 24, 2015 (Tuesday, 22h00): Project II details are online.

Dec. 17, 2015 (Thursday, 13h00): Test data made available.

Dec. 21, 2015 (Monday, 13h00): Project due.

Summary and Goals

This project is about applying machine learning to a real-world problem. During the course of this project, you will learn:

- To process, analyze, and visualize real data.
- To formulate real-world problems as machine learning problems.

- To analyze and compare various ML methods on a real-world application.

Total marks are 100 and they constitute 30% of the overall grade. Out of this, you will get 80 marks for your analysis and predictions. You will get 10 Marks for your code and 10 marks for your report content.

Project 2 Wiki

We set up a wiki, which we periodically update with answers to common questions.

To access it [click here](#).

STEP 1: Understand the problem

Your goal is to build a system that recognizes the object present in an image. Click *description* to see details, *data* to download the data, and *code* to get started with the project.

- [Description](#)
- [Download the data](#)
- [Example code](#)

STEP 2: Understand the data and apply ML methods

You should visualize the data and do basic exploratory data analysis, similar to what you did for project I.

You should understand the performance measures, e.g. why do you need to compute a balanced error measure instead of the 0-1 loss or the log-loss (similar to project I)? Performance measures often play a big role in being able to come up with a good ML method.

You should choose your methods carefully. Think about the reasons for choosing a particular method. How do you expect it to behave compared to the others you tried before? If it does not give you the improvement you expected, then you should think about the reasons behind it. You should write your analysis in the report. Please only include the most important experiments, not all of them.

Software: You are allowed to use existing software (Matlab toolboxes / code, etc). There is no need to implement all the methods yourself. **However, you have to describe and discuss the methods you use in the report and show that you understand them, as well as what there parameters / hyper-parameters mean.**

If you have implemented methods on your own, you should indicate this in the report (put a section named 'implementation details'). This may help you on the 10 marks reserved for your code.

STEP 3: Predict test data

The test data will be available at the date **indicated above**. It will contain images along with their fetures, similar to the training data, but without any class labels. You need to provide the following predictions:

1. **Binary Prediction:** Write your binary predictions in a mat file named `pred_binary.mat`. This mat file should contain a vector 'Ytest' which contains the prediction for each sample. The size of Ytest must be 11453x1, with 0 for class Other and 1 for the rest. One way to create Ytest is shown below:

```
% assign your predicted scores to Ytest first, then:  
save('pred_binary', 'Ytest');
```

2. **Multi-class Prediction:** Write your multi-class predictions in a mat file named `pred_multiclass.mat`. This mat file should contain a vector 'Ytest' which contains the prediction for each sample. The size of Ytest must be 11453x1, where each element contains the predicted class, either 1, 2, 3 or 4.

Check your prediction files format by running the following script:

testMyPredictions.m.

STEP 4: Write your findings in a report

The maximum number of pages is limited to 8. 5 marks will be deducted per extra page. Please write your report similarly to project I.

If you have implemented methods on your own, you should indicate this in the report (put a section named 'implementation details').

STEP 5: Submit

The submission page will be available soon on Moodle. You have to submit the following files in a zip file (size limitation 20MB).

- report.pdf (your report).
- pred_binary.mat (binary predictions). See **details** of the file format
- pred_multiclass.mat (multi-class predictions). See **details** of the file format
- A subfolder called code which contains all your code (size limitation 20MB).

Late submissions are not allowed.

Marking: Total marks are 100 and constitute 30% of the overall grade. Out of this, you will get 80 marks for your analysis and predictions. You will get 10 Marks for your code and 10 marks for your report content.