

Final Project

TA email: ml2014ta@csie.ntu.edu.tw

RELEASE DATE: 11/19/2014

COMPETITION END DATE: **01/11/2015 NOON ONLINE**

REPORT DUE DATE: **01/21/2015 NOON ONLINE**

Unless granted by the instructor in advance, no late submissions will be allowed.

Any form of cheating, lying, or plagiarism will not be tolerated. Students can get zero scores and/or fail the class and/or be kicked out of school and/or receive other punishments for those kinds of misconducts.

You should write your solutions in English or Traditional Chinese with the common math notations introduced in class or in the problems. We do not accept solutions written in any other languages.

Introduction

In this final project, you are going to be part of an exciting machine learning competition. Consider a startup company that features a coming product on the mobile phone. The core of the product is a robust character recognition system. The company has collected some written characters in order to build a prototype of the product, and wants to demo the robustness of the prototype. The board of directors of the company decided to hold a competition and make the problem of building the prototype open to experts like you. To win the prize, you need to fight for the leading positions on the score board. Then, you need to submit a comprehensive report that describes not only the recommended approaches, but also the reasoning behind your recommendations. Well, let's get started!

Data Sets

The problem is formalized as a multiclass classification problem, where the goal is to classify all the characters accurately. We will have two tracks of competition. You will be provided with examples of the form (y, \mathbf{x}) , where \mathbf{x} is the pixels of some character, and y is the label. The prototype uses the Chinese zodiacs and upper/lower cases of Chinese numbers as the possible characters, and hence there are 32 possible labels. The data will be divided to the training set, the first test set, and the second test set. For the test sets, only \mathbf{x} will be shown. Details will be announced later.

To maximize the level of fairness, you are not allowed to manually label the test examples or write (and add) any additional characters at any time.

Evaluation Criterion

- (1) For the first track of the competition, your classifier g for an example (\mathbf{x}, y) will be evaluated by the following pointwise error function:
 - if $y = g(\mathbf{x})$, no error
 - otherwise ($y \neq g(\mathbf{x})$), if they still correspond to the same number, no error
 - else, an error of 1

Effectively, it means we use 0/1 error on the 22 effective classes, one for each zodiac and one for each number, regardless of cases.

- (2) For the second track of the competition, your classifier $g(\mathbf{x})$ will be evaluated as follows:
 - if $y = g(\mathbf{x})$, no error
 - **number/zodiac error.** otherwise ($y \neq g(\mathbf{x})$), if one of them is a zodiac and the other is a number, an error of 4
—such as “mouse” predicted as “one” or “two” predicted as “cow”

- **within zodiac error.** otherwise ($y \neq g(\mathbf{x})$), if one of them is a zodiac and the other is a zodiac, an error of 2
—such as “mouse” predicted as “cow”
- **case error.** otherwise ($y \neq g(\mathbf{x})$), if one of them is a number and the other is the same number with a different case, an error of 1
—such as upper-case “one” predicted as lower-case “one”
- **within number error.** otherwise ($y \neq g(\mathbf{x})$), an error of 2
—such as upper-case “one” predicted as “five”

The error function above can be put in a simpler way. The usual error is 2; case error is considered light and penalized less (1); number/zodiac error is considered serious and penalized more (4).

This kind of multiclass classification problem is often called *cost-sensitive classification*, in case you want to look up references online.

Survey Report

You are asked by the board to study at least THREE machine learning approaches using the training set above. Then, you should make a comparison of those approaches according to some different perspectives, such as efficiency, scalability, popularity, and interpretability. In addition, you need to recommend THE BEST ONE of those approaches as your final recommendation **for each track** and provide the “cons and pros” of the choice.

The survey report should be no more than SIX A4 pages with readable font sizes. The most important criterion for evaluating your report is replicability. Thus, in addition to the outlines above, you should also describe how you pre-process your data; introduce the approaches you tried and provide specific references, especially for those approaches that we didn’t cover in class; list your experimental settings and the parameters you used (or chose) clearly. Other criteria for evaluating your survey report would include, but are not limited to, clarity, strength of your reasoning, “correctness” in using machine learning techniques, the work loads of team members, and properness of citations.

For grading purposes, a minor but required part in your survey report for a two- or three-people team (see the rules below) is how you balance your work loads.

Competition

The submission site would be announced later. We will limit each team with 6 submissions per day **for each track** to check your performance on the first test set. But use your submissions wisely—you do not want to leave the board with a bad impression that you just want to “query” or “overfit” the test examples. After submitting, there will be a score board showing the test error. A few days before the competition deadline, the second test set will be released, and the board will actually evaluate you on those.

The competition ends at noon on 01/11/2015. We’ll have a mini-ceremony to honor the best team(s) on 01/12/2015. The competition site will continue to be open until the due day of the report.

Misc Rules

Report: Please upload one report per team electronically on CEIBA. You do not need to submit a hard-copy. The report is due at noon on 01/21/2015.

Teams: By default, you are asked to work as a team of size THREE. A one-person or two-people team is allowed only if you are willing to be as good as a three-people team. It is expected that all team members share balanced work loads. Any form of unfairness, such as the intention to cover other members’ work, is considered a violation of the honesty policy and will cause some or all members to receive zero or negative score.

Algorithms: You can use any algorithms, regardless of whether they were taught in class.

Packages: You can use any software packages for the purpose of experiments, but please provide proper references in your report for replicability.

Source Code: You do not need to upload your source code for the final project. Nevertheless, please keep your source code until 02/28/2015 for the graders' possible inspections.

Grade: The final project is worth 600 points. That is, it is equivalent to three usual homework sets. At least 540 of them would be reserved for the report. The other 60 may depend on some minor criteria such as your competition results, your discussions on the boards, your work loads, etc..

Collaboration: The general collaboration policy applies. In addition to the competitions, we still encourage collaborations and discussions between different teams.

Data Usage: You can use only the data sets provided in class for your experiments, and you should use the data sets properly. Using any tricks to query the labels of the test set is strictly prohibited.