1 Introduction

This week’s practical focuses on comparing performance of the classifiers seen in class, namely mixture of weak classifiers trained with boosting, SVM and Gaussian Mixture Models with Bayes. While most classification techniques try to update a complex model to fit the data in order to optimize an objective function (log-likelihood in GMM, minimum margin and slack for SVM), Boosting chooses weak classifiers (WC) iteratively among a large set of randomly created WC and combines them to create a strong classifier by increasing the weights of datapoints not well classified by the previous combination of WC.

We will start with a qualitative comparison on 2D datasets drawn by hand using MLDemos. We will then perform quantitative comparison of the classifiers using datasets from the UCI database. A sensitivity analysis will then be performed as a tie-breaker to determine, when two algorithms give similar good performance, which one is more or less sensitive to the choice of hyperparameters.

2 Algorithm options

We will use only C-SVM (leaving out ν-SVM). For this algorithm, two parameters need to be set. These are the penalty factor C and the kernel width. With GMM, a single parameter needs to be chosen. This is the number of Gaussian. MLDemos assume same number of Gaussians in each class. Note that this is a limitation of MLDemos, as GMM with Bayes can in principle use arbitrary number of Gaussians to model each class separately. The two parameters of a Boosting algorithm are the choice of weak classifiers (WC) available and the boosting update rule. The WC ranges from simple stumps to Support Vector Machines. Here are the ones available

- **Decision stumps**: Threshold on one dimension.
- **Random projections**, i.e. linear classifier: threshold on a projection that can be a linear combination of several dimensions. (Extended case of decision stump not aligned on the dimensions).
• **Random rectangles/circles/Gaussians**: The limits of one class are defined by a rectangle/circle/Gaussian.

• **Random Support Vector machines**: The classifier is a Support Vector Machine with randomly chosen support vectors and parameters (gammas range from 1e-5 to 1 and alphas from -10 to 10).

### 2.1 Boost type

During the weight update phase of the algorithm, the error of each weak classifier is taken into account. There are several ways to compute this error:

- **Discrete**: If a point is misclassified, its error is 1. This does not take into account how much the point is misclassified (ie. close to the border of the class)

- **Real**: Takes into account the confidence of the WC so that points close to the border but misclassified do not weight as much as misclassified points away from the border.

- **Logit**: The exponential loss function used as an upper bound on the error is replaced by a binomial log likelihood

- **Gentle**: A modified version of real Adaboost, less sensitive to outliers.

### 3 Datasets

The comparison must be performed both on provided and hand-drawn datasets.

#### 3.1 The provided datasets

Two high-dimensional datasets will be provided. These are taken from the UCI database and are:  

- **House votes**: Records of votes in the Senate. The goal is to determine the political affiliation of senators given their votes.

- **Ionosphere**: Classification of radar returns from the ionosphere.

You can project the datasets with PCA before running the classification.

#### 3.2 Standard Dataset Generator

MLdemos includes a tool to draw predefined types of points distributions. To access it, click on the ”Add Data” button on top of the main window (See Figure 1). Choose the type and parameters of the distribution of points and click on ”Add Data”. The first three choices are for classification, the other three are for regression and are not used in this practical.

1[^datasets]

4 What to do

4.1 Get an intuition of the algorithm, perform a qualitative evaluation

- Try to draw some datasets by hand and use the different weak classifiers, SVM and GMM with Bayes to get an intuition in 2D of the way the algorithms work. Starting from 1 Weak Classifier, incrementally increase the number of WC, change the kernel width and the penalty factor in SVM and increase the number of Gaussian in GMM, to see how the WC, Support Vectors and Gaussians are changed, to improve classification.

- Find the best suited set of parameters in each case to classify each of the three distribution types in the Standard Dataset Generator: Checkerboard, concentric circles and swiss roll (keep the number of swirls / circles / grid count to 2).

4.2 Compare the algorithms quantitatively

Second, use the provided datasets and the Standard dataset generator to answer the following questions in a quantitative way (use the compare tool described in the previous practical):

- Test the different weak classifiers available for boosting and compare their performance on the provided datasets (House votes and ionosphere).
• Compare the performance of boosting of SVMs and boosting of Gaussians with classic SVM and GMM; which perform better and what are the advantages/drawbacks of each?

• Use the MLDemos tool "Gridsearch" to choose a range of hyperparameters and conduct a sensitivity analysis for each of the methods (See Figure 2).

Figure 2: Set the two parameters you want to study, their range and the number of steps. You can also modify the number of cross-validations folds. Finally, hit the Run button. You can choose to display the raw error of the algorithm or the F-measure of the classification.