Technische Universität München
Fakultät für Informatik
Prof. N. Navab, Ph.D.

WS 2013-2014
Exercise Sheet 6
November 27, 2013

Practical Course: Machine Learning in Medical Imaging

Boosting

For your submission, please send the completed script 'adaboost.m' producing the different plots of exercise 1, and a small report containing your observations and the proof of exercise 2, by **Wednesday, December 4th 10:30am** to pierre.chatelain@in.tum.de.

1 Implementation of AdaBoost

The script 'adaboost.m' contains an incomplete code for running AdaBoost on a synthetic example. It generates a training set $(\mathbf{X}_{\text{train}}, \mathbf{Y}_{\text{train}})$ and a testing set $(\mathbf{X}_{\text{test}}, \mathbf{Y}_{\text{test}})$ sampled from the same distribution, where $X \in \mathbb{R}^2$ are the feature values and $Y \in \{-1, 1\}$ are the labels. The goal of the exercise is to train a strong classifier from a family of weak classifiers using the AdaBoost algorithm on the dataset $(\mathbf{X}_{\text{train}}, \mathbf{Y}_{\text{train}})$, in order to predict the labels of the testing set $\mathbf{X}_{\text{test}}$.

The proposed weak classifiers are decision stumps, defined as $f(x) = s(2[X_d > \theta] - 1)$, where d is the feature dimension along which the decision is taken, $[\cdot]$ is 1 if $\cdot$ is true and 0 otherwise, θ is the threshold applied along the dimension d, and $s \in \{-1, 1\}$ is the *polarity* of the decision stump (*i.e.* which side of the threshold corresponds to which label).

The AdaBoost algorithm iterates the following steps:

- Find $\hat{f} = \arg\min_f \epsilon(f)$, where $\epsilon(f) = \frac{\sum_i w_i [y_i \neq f(x_i)]}{\sum_i w_i^t}$.
- Update the weights: $w_i \longleftarrow \frac{w_i}{Z} \exp(-\alpha y_i f(x_i))$, where $\alpha = \frac{1}{2} \log \frac{1-\epsilon}{\epsilon}$.
- Update the strong classifier: $F \longleftarrow F + \alpha f$

The optimization of the weak classifiers is already implemented, as well as the computation of the training and testing errors. Your tasks for this exercise are the following (see the 'TODO' comments in the script 'adaboost.m'):

1. Compute the parameter α and the updated weights w within the boosting loop (second step of the algorithm above).
2. Plot the testing set $\mathbf{X}_{\text{test}}$, displaying the points with a different color for each label (*e.g.* red for $Y = -1$, blue for $Y = +1$). Use two different plots: one with the true labels $\mathbf{Y}_{\text{test}}$, and one with the labels predicted by your boosted classifier ($\mathbf{F}$), and compare the two plots.
3. Plot the evolution of the training error and testing error during 100 iterations and report what you observe.

2 (Bonus question) Optimization with exponential loss

We claimed during the lecture each iteration of AdaBoost optimizes an additive model with exponential loss:

$$(\hat{\alpha}, \hat{f}) = \arg\min_{\alpha, f} \sum_{i=1}^{N} \exp\left(-y_i(F(x_i) + \alpha f(x_i))\right) \tag{1}$$

where $(x_i) \in \mathcal{X}$ are the samples with labels $(y_i) \in \{-1, 1\}$, $f : \mathcal{X} \longrightarrow \{-1, 1\}$ are the weak classifiers, and $F : \mathcal{X} \longrightarrow \mathbb{R}$ is the current strong classifier.

We propose here to demonstrate this result:

1. Show that for all α,

$$\sum_{i=1}^{N} \exp\left(-y_i(F(x_i) + \alpha f(x_i))\right) = \sum_{i=1}^{N} w_i(e^{-\alpha} + (e^{\alpha} - e^{-\alpha})\epsilon(f)) \tag{2}$$

   where $w_i = \exp(-y_i F(x_i))$ and $\epsilon(f) = \frac{\sum_{i:y_i \neq f(x_i)} w_i}{\sum_i w_i}$.

2. Deduce from this equality that $\hat{f} = \arg\min_f \epsilon(f)$.

3. Show that $\hat{\alpha} = \arg\min_\alpha \left[\sum_{i=1}^{N} w_i(e^{-\alpha} + (e^{\alpha} - e^{-\alpha})\epsilon(\hat{f}))\right] = \frac{1}{2} \log \frac{1-\epsilon(\hat{f})}{\epsilon(\hat{f})}$

4. Conclude.