Machine Learning	Dr. Mario Fritz and Dr. Björn Andres
Exercise Sheet 6 - 08.12.2014	due: 15.12.2014

Exercise 16 - SVM with class balanced loss

We have seen already in an earlier exercise that for very unbalanced classes the minimization of the error can lead to the fact that one always predicts the dominant class. In this case it makes more sense to optimize other criteria as introduced in the lecture e.g. maximization of true positive rate (sensitivity) and true negative rate (specificity).

- **(2 Points)** Define costs

$$C(Y = 1, \operatorname{sign}(f(X)) = -1) = C_{1,-1}, \quad \text{and} \quad C(Y = -1, \operatorname{sign}(f(X)) = 1) = C_{-1,1},$$

  such that, the expected cost-sensitive risk

$$R^C(f) = \mathbb{E}[\, C(Y, \operatorname{sign}(f(X)))\, \mathbb{1}_{f(X)Y \le 0}],$$

  is equal to the sum of false positive rate and false negative rate.

- **(4 Points)** In practice, often the so called **class-balanced error rate** S is used,

$$S = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{1}_{f(X_i^k) \neq k},$$

  where K is the number of classes, N_k the number of training points in class k and $f : \mathcal{X} \to \{1, \ldots, K\}$.
  For the base of binary classification, $Y = \{-1, 1\}$, the class-balanced error rate reduces to

$$S = \frac{1}{2} \Big[\frac{1}{N_+} \sum_{i=1}^{N_+} \mathbb{1}_{f(X_i^+) \neq 1} + \frac{1}{N_-} \sum_{i=1}^{N_-} \mathbb{1}_{f(X_i^-) \neq -1} \Big],$$

  where $N_\pm$ is the number of positive/negative training points and $X_i^\pm$ are the training points from the positive/negative class. Show, that

$$\mathbb{E}[S] = \frac{1}{2} \Big[\mathrm{P}\big(f(X) = -1 \,|\, Y = 1\big) + \mathrm{P}\big(f(X) = 1 \,|\, Y = -1\big) \Big].$$

  Thus, the class balanced error rate is up to the factor $\frac{1}{2}$ just the empirical version of the risk functional of part a).

- **(3 Points)** Change the primal binary SVM problem

$$\begin{aligned} \min_{w \in \mathbb{R}^d,\, b \in \mathbb{R},\, \xi \in \mathbb{R}^n} \; & \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^{n} \xi_i \\ \text{subject to: } & Y_i(\langle w, X_i \rangle + b) \ge 1 - \xi_i, \quad \forall\, i = 1, \ldots, n, \\ & \xi_i \ge 0, \quad \forall\, i = 1, \ldots, n \end{aligned}$$

  so that the cost-sensitive loss of part a) is used.

Hints:

- for a) write down the false positive rate, $\mathrm{P}(f(X) = 1 \,|\, Y = -1)$, as an expectation and then as an integral. Do the same with the false negative rate and the cost-sensitive risk functional.
- for b) have a look at the proof where the expectation of the AUC has been derived.
- for c) check how the slack variables ξ_i are related to the hinge loss.

Exercise 17 - ROC and AUC

You are reviewer for an international conference on machine learning. Among the submissions are two independent papers, each proposing a solution to the important problem of automatic classification of hand-crafted contemporary art pottery. As you are an expert in this emerging new field, you are asked to write a review of the two papers. In both papers, the quality of the proposed classifier is evaluated by plotting the ROC curves. Author 1 reports an AUC of 0.73, Author 2 reports an AUC of 0.70.

Both authors made the results of their classifier available as supplementary material. However, the authors used different datasets to evaluate their methods. You can find the dataset Y_1 (the true labels) and the prediction f_1 of the first author as well as the dataset Y_2 and the prediction f_2 of the second author in the file `auc.mat`.

a. **(4 Points)** Inspect the datasets and the predictions made by the two methods. Reproduce the values of the AUC reported by the two authors by means of the function `[AUC,ERR]=PlotROC(Y,f,graphics)`. Perform 1000 runs of a random classifier on each dataset and compare the ROC curves and the AUC values obtained by the two methods with the results of the random classifiers. Report your results. Which classifier do you think works better? Justify your decision. (Written solution on paper. Submit plots/statistics which illustrate your results (histograms of AUC resp. error values).)