COST SENSITIVE ERROR CORRECTING OUTPUT CODES

1. The Algorithm

For all subsets s, for all thresholds t, $\{(x, I(\frac{\sum_{y\in s} c_y}{\sum_y c_y} \geq t), \sum_y c_y : (x, c_1, ..., c_k) \in S\}$

Apply A to get hypothesis h

To predict, choose:

$$\arg\min_y E_t E_s c(x) I(y \in s) + (1 - c(x)) I(y \notin s)$$

2. The Analysis

Theorem 2.1. *Error rate ϵ for binary classification implies loss rate at most $2\epsilon E_{(x,c_1,\ldots,c_k)} \sum_y c_y$*

Proof. We only need to show the reduction to importance weighted classification. Since we use as an importance $\sum_y c_y$, we need only show that the statement holds when $\sum_y c_y = 1$. To do the prove, we first analyze the expected sum for some prediction y'. Without loss of generality(since we can always consider the complement of any set), we can assume for all s and $y' \in s$.

$E_t E_s c_{st}(x)$
$= \frac{1}{|\{s\}|} \sum_s E_t c_{st}(x)$
$= \frac{1}{|\{s\}|} \sum_s \sum_{y\in s} c_y$
$= \frac{1}{|\{s\}|} \sum_s c_{y'} + \frac{1}{2} \sum_{y\neq y'} c_y$
$= c_{y'} + \frac{1}{2} \sum_{y\neq y'} c_y$
$= c_{y'} + \frac{1}{2}(1 - c_{y'})$
$= \frac{1}{2}(c_{y'} + 1)$

So the expected value is $\frac{1}{2}(c_{y'} + 1)$ when the classifiers are perfectly correct. When an ϵ fraction err, the expected value changes by at most ϵ. Consequently, an ϵ fraction erring implies at most a 2ϵ increase in cost since we might reduce one expectation by ϵ and increase another by ϵ simultaneously. □

3. What's wrong here?

If we consider the embedding of multiclass classification into cost sensitive classification, we get $\sum_y c_y = k - 1$ where k is the number of labels. Applying the above reduction implies ϵbinary errors $\Rightarrow 2\epsilon(k-1)$ multiclass errors or less. This is not very impressive in comparison to standard ECOC ($\epsilon \to 4\epsilon$). Can we find an as-efficient cost sensitive reduction?

If we know the problem is an embedded multiclass problem, then we can choose $t = 0.5$. In this case we have:

$E_t E_s c_{st}(x)$ for y'
$= \frac{1}{|\{s\}|} \sum_s E_t c_{st}(x)$
$= 0$ or $\frac{1}{2}$ depending on whether $y' =$ the target.

So ϵ error rate $\Rightarrow 4\epsilon$ multiclass error rate, as expected.

The essential difference here is the range of t. When the decision is between balanced sets, having t near the boundary of balance is important. However, this does not cover the range of interest in cost sensitive classification.

Is there a modification of the cost sensitive error correcting code algorithm which achieves optimality on the multiclass problem? The best I can think of is letting t range over the set of "interesting values" observed in the training set, but I am not sure this is reasonable.