CIS 242 — Intro to Machine Learning

Winter 08 — Warmuth

handed out Tu 2-22-08 **Homework Assignment 2** due Tu 2-29-08 in class

1. Consider the Randomized Weighted Majority (WMR) algorithm, when the loss is the "dot loss".

   $\mathbf{w}^0$ = uniform over n experts
   for trial $t = 1, \ldots$
     Draw expert i with probability w_i^{t-1}
     Receive loss vector $\boldsymbol{\ell}^t \in \{0,1\}^n$
     Incur loss ℓ_i^t and expected loss $\mathbf{w}^{t-1} \cdot \boldsymbol{\ell}^t$
     Update weights to $w_i^t = w_i^{t-1} \beta^{\ell_i^t}$/normaliz.

   When $\beta = 0$, then this becomes the "Randomized Halving algorithm", which always picks an expert uniformly among the remaining consistent experts.

   Show that for any sequence of loss vectors for which there is at least one consistent expert, the Randomized Halving algorithm has expected loss at most $\ln n$.

2. We can modify the WMR algorithm as follows under the assumption that there is an expert which makes $\leq k$ mistakes:

   For loss vector $\mathbf{s} \in \{0, 1, \ldots\}^n$ specifying how much loss each expert has incurred so far, predict with distribution

   $$p_i(\mathbf{s}) = \begin{cases} \frac{\beta^{s_i}}{\text{normaliz.}} & \text{if } s_i \leq k \\ 0 & \text{otherwise.} \end{cases}$$

   Here $\beta \in [0, 1)$ as usual.

   Note that the original WM update does not set the probabilities of experts with more than k mistakes to zero. It always uses probabilities proportional to β^{s_i}.

   We showed in class in various ways that for the original WMR algorithm, for any sequence of loss vectors where the best expert has total loss at most k:

   $$\text{total loss of alg} \leq \frac{\ln n + k \ln(1/\beta)}{1 - \beta}.$$

Show that the same bound still holds for the modified algorithm .

Open ended question: Can you prove a better bound for the modified WMR algorithm.

3. Consider the implicit EG algorithm for linear regression:

$$\begin{aligned} \hat{y}_t &= \mathbf{w}^{t-1} \cdot \mathbf{x}^t \\ w_i^t &= w_i^{t-1} \exp^{-\eta(\mathbf{w}^t \cdot \mathbf{x}^t - y_t) x_i^t} / \text{normaliz.} \end{aligned}$$

Show how to compute this update with a line search.

Hint: Use the motivation of the update.

4. (Extra credit) Imagine an adversary producing the sequence of loss vectors with the restriction that there is one expert that has loss at most k. The goal of the adversary is to make the WMR algorithm incur a large (expected) loss. Prove that the optimum adversary always chooses unit vectors as loss vectors (i.e. one expert has loss one and the rest loss zero). It does this until all but one vector have loss k.

Open ended question: How could you use this fact to prove upper bounds for WMR?

Hint: Summarize the current state as a vector $s_i \in \{0, 1, \ldots, \}$, where s_i is the total loss of expert i incurred so far. Make up a graph with states as nodes. Sequences of unit loss vectors correspond to certain paths. ...