E1 245 - Online Prediction and Learning, Aug-Dec 2015
Homework #1

(10 marks per problem)

1. *Exponential inequalities*
   Prove the following inequalities (useful in showing mistake bounds for Weighted-Majority, for instance):

   (a) $\forall x \in \mathbb{R} \quad e^x \geq 1+x$

   (b) $\forall x \geq 0 \quad e^{-x} \geq 1-x+x^2$

   (c) $\forall |x| \leq \frac{1}{2} \quad e^{-x-x^2} \leq 1-x$

2. *1-Bit prediction problem, one-shot version, iid outcomes*
   Let $n \geq 1$ be a fixed integer. Consider the problem of predicting the next bit of an independent and identically distributed (iid) Bernoulli sequence given the previous $n-1$ bits. i.e., let $Y_1, Y_2, \ldots, Y_n$ be iid random variables distributed as Ber(p), $0 < p < 1$. We seek an algorithm that outputs $\hat{Y}_n \in \{0,1\}$, possibly depending on $Y_1, \ldots, Y_{n-1}$ but not on Y_n, to minimize the prediction error (relative to the best constant predictor 0 or 1), defined as $\mathbb{P}\left[\hat{Y}_n \neq Y_n\right] - \min\{p, 1-p\}$. Show that the prediction error for the MAJORITY prediction rule[1] admits the upper bound[2]

$$\mathbb{P}\left[\hat{Y}_n \neq Y_n\right] - \min\{p, 1-p\} = O\left(1/\sqrt{n}\right).$$

   (Hint: Suppose $p < 1/2$. Begin by conditioning on the value of $\hat{Y}_n$. Use Hoeffding's inequality to bound the probability of error in terms of p. Conclude by bounding the worst that this can be as a function of p.)

3. *1-Bit prediction problem, sequential version, iid outcomes*
   Consider now the game of *sequentially* predicting bits of an iid Ber(p) sequence (using only the previously observed bits) for a total of n rounds. Using the result from the previous problem, argue that the same $O(1/\sqrt{n})$ bound on the expected mean no. of mistakes (relative to the best constant predictor 0 or 1), i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{P}\left[\hat{Y}_i \neq Y_i\right] - \min\{p, 1-p\} = O\left(1/\sqrt{n}\right)$$

   holds when the MAJORITY rule is used to predict at each round.

4. *1-Bit prediction problem, sequential version, arbitrary outcomes*
   Consider the game of sequentially predicting the next bit of an *arbitrary, fixed* bit sequence for a total of n rounds. What happens to the mean no. of mistakes (relative to the best constant predictor 0 or 1) for the MAJORITY rule? (in other words, what is the worst that it can be?) Can you show that this behaviour must hold more generally for any *deterministic* (and causal) prediction rule?[3]

---

[1] The MAJORITY prediction rule is $\hat{Y}_n := \mathbb{1}\left\{\sum_{i=1}^{n-1} Y_i \geq \frac{n-1}{2}\right\}$ or $\hat{Y}_n := \mathbb{1}\left\{\sum_{i=1}^{n-1} Y_i > \frac{n-1}{2}\right\}$, depending on how you may want to break ties.

[2] Big-Oh notation: We say that $f(n) = O(g(n))$ if there exist constants α and n_0 such that $f(n) \leq \alpha g(n)$ $\forall n \geq n_0$.

[3] It is enough to give order-wise (in n) bounds or estimates.

5. *Tweaking the MAJORITY algorithm in the absence of perfect experts*
We showed that the MAJORITY or HALVING algorithm for binary prediction makes at most $\log_2 N$ mistakes using the advice of N experts whenever some expert is always predicting correctly. Show that a straightforward modification of MAJORITY makes at most $O((m+1)\log_2 N)$ mistakes whenever the best expert makes $m \geq 0$ mistakes. (Hint: What if all the experts get thrown out?)