COMS 4773 Spring 2024 HW 4 (due Apr. 19 at noon)

Problem 1. Suppose P is a distribution over $\mathbb{R}^d \times \{-1, 1\}$, and $(X, Y) \sim P$. Here, we regard $\{-1, 1\}$ as a subset of $\mathbb{R}$.

(a) The *mean squared error* of a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\mathbb{E}[(f(X) - Y)^2].$$

Characterize the function $f^\star\colon \mathbb{R}^d \to \mathbb{R}$ that minimizes the mean squared error in terms of the conditional distribution of Y given X. Justify your answer.

(b) Suppose $f^\star\colon \mathbb{R}^d \to \mathbb{R}$ is the minimizer of the mean squared error (as defined above), and suppose $h^\star\colon \mathbb{R}^d \to \{-1, 1\}$ is the Bayes classifier for P. Prove that for any $f\colon \mathbb{R}^d \to \mathbb{R}$,

$$\Pr(h(X) \neq Y) - \Pr(h^\star(X) \neq Y) \leq \sqrt{\mathbb{E}[(f(X) - Y)^2] - \mathbb{E}[(f^\star(X) - Y)^2]}$$

where h is defined by $h(x) = \operatorname{sign}(f(x))$.

Problem 2. Suppose $J\colon \mathbb{R}^d \to \mathbb{R}$ is an L-Lipschitz convex objective function that you would like to approximately minimize (for some $L > 0$). However, instead of a first-order (gradient) oracle, you only have access to the following "noisy oracle" (for some fixed $V > 0$):

> Given any $w \in \mathbb{R}^d$ as input, the noisy oracle returns a random vector $G = (G_1, \ldots, G_d)$ such that, for each $i \in [d]$,
>
> $$\mathbb{E}[G_i \mid \mathcal{F}] = \frac{\partial J}{\partial w_i}(w) \quad \text{and} \quad \mathbb{E}[(G_i - \mathbb{E}[G_i \mid \mathcal{F}])^2 \mid \mathcal{F}] \le V,$$
>
> where $\mathcal{F}$ denotes the results of all previous queries to the oracle.

(a) Analyze the following algorithm for approximately minimizing J (with hyperparameter $\eta > 0$):

- Set $w_0 = (0, \ldots, 0) \in \mathbb{R}^d$.
- For $t = 1, 2, \ldots, T$:
  * Query the oracle at w_{t-1} to obtain $G_t = (G_{t,1}, \ldots, G_{t,d})$.
  * Set $w_t = w_{t-1} - \eta G_t$ for some fixed step size $\eta > 0$.
- Return $\bar{w} = \frac{1}{T} \sum_{t=0}^{T-1} w_t$.

In particular, prove a bound on $\mathbb{E}[J(\bar{w}) - \min_{w \in B^d} J(w)]$ in terms of L, V, η, and T, where $B^d = \{w \in \mathbb{R}^d : \|w\|_2 \le 1\}$ is the unit ball; and give the value of the hyperparameter η that minimizes the bound (up to constant factors).

(b) Suppose you are tasked with implementing the noisy oracle. You have access to a first-order oracle for J, but you are told that your implementation of the noisy oracle must only output vectors with at most one non-zero component. Design and write the pseudocode for such an implementation, and prove its correctness. You should try your best to make it so V is as small as possible.

Problem 3. For $\gamma \in (0, 1/2)$, let $f_\gamma \colon [-1, 1] \to \mathbb{R}$ denote the piecewise linear function defined by

$$f_\gamma(t) = \begin{cases} -1 & \text{if } -1 \le t \le -\gamma, \\ t/\gamma & \text{if } -\gamma < t < \gamma, \\ 1 & \text{if } \gamma \le t \le 1. \end{cases}$$

Prove that, for any $\epsilon \in (0, 1/2)$ and any $\gamma \in (0, 1/2)$, there is a polynomial p of degree at most $\frac{1000}{\gamma} \ln \frac{1}{\epsilon}$ such that

$$|f_\gamma(t) - p(t)| \le \epsilon \quad \text{for all } t \in [-1, -\gamma] \cup [\gamma, 1].$$

(The constant 1000 is probably more than necessary.)

Hint: Combine a polynomial obtained from Jackson's theorem (presented in lecture) together with a particular linear combination of Bernstein polynomials (also presented in lecture).