Submitted by <**name**>

This problem set will cover concepts from the units on computational complexity and statistical physics.

1 (60 PTS.) SQ Dimension Bounds for Generalized Linear Models (10/23)

Motivation: In this exercise you will try your hand at proving a CSQ lower bound for a supervised learning problem. This will involve integrating what you learned about statistical query dimension with tools from Hermite analysis that were touched upon in the unit on supervised learning.

Setup: For every $\ell \in \mathbb{Z}_{\geqslant 0}$, let ϕ_ℓ denote the normalized degree-ℓ Hermite polynomial given by

$$\phi_\ell(x) = \frac{1}{\sqrt{\ell!}}\mathrm{He}_\ell(x)\,,$$

where He_ℓ is the ℓ-th probabilist's Hermite polynomial. As discussed in class, these form an orthonormal basis for the space of functions which are L_2 integrable with respect to the standard Gaussian measure.

Let $\sigma : \mathbb{R} \to \mathbb{R}$ denote an activation function with Hermite expansion $\sigma(x) = \sum_{\ell=0}^{\infty} c_\ell \phi_\ell(x)$. Suppose σ satisfies the normalization condition

$$\mathbb{E}_{g\sim\mathcal{N}(0,I)}[\sigma(g)^2] = 1\,.$$

Suppose we are given samples $\{(x_i, y_i)\}_{i=1}^N$ where $x_i \sim \mathcal{N}(0, I)$ and $y_i = \sigma(\langle w, x_i\rangle)$ for some unknown $w \in \mathbb{S}^{d-1}$. In this exercise, we will establish a CSQ lower bounds for this problem.

1.A. (25 PTS.) Given vectors $u, v \in \mathbb{S}^{d-1}$, prove the identity

$$\mathbb{E}_{x\sim\mathcal{N}(0,I)}[\phi_\ell(\langle u, x\rangle) \cdot \phi_{\ell'}(\langle v, x\rangle)] = \mathbb{I}[\ell = \ell'] \cdot \langle u, v\rangle^\ell\,.$$

You may use the fact that for $g \sim \mathcal{N}(0, 1)$, $\mathbb{E}[e^{tg}] = e^{t^2/2}$ for all $t \in \mathbb{R}$, and that the probabilist's Hermite polynomials He_ℓ are given by the generating function

$$e^{tg-t^2/2} = \sum_{\ell=0}^{\infty} \frac{1}{\ell!}\mathrm{He}_\ell(g)t^\ell\,.$$

1.B. (15 PTS.) Let ℓ^* denote the *information exponent* of σ, that is, the smallest $\ell \in \mathbb{Z}_{\geqslant 0}$ for which $c_\ell \neq 0$. Use the above to prove that $\mathbb{E}[\sigma(\langle u, x\rangle)\sigma(\langle v, x\rangle)] \leqslant \langle u, v\rangle^{\ell^*}$.

1.C. (10 PTS.) Recall the proof from class that a lower bound on statistical query dimension implies a lower bound on the correlational statistical query complexity of the corresponding supervised learning problem. Adapt the proof to show the following slightly refined variant.

Suppose a finite set of functions $\mathcal{F} = \{f_1, \ldots, f_m\}$ satisfies $|\mathbb{E}_{x\sim\mathcal{N}(0,I)}[f_i(x)f_j(x)]| \leqslant \epsilon$ and $\mathbb{E}_{x\sim\mathcal{N}(0,I)}[f_i(x)^2] = 1$ for all $i \neq j$. Then any CSQ algorithm for supervised learning functions in $\mathcal{F}$ over Gaussian examples requires at least $m(\tau^2 - \epsilon)/2$ queries or tolerance less than τ to produce a function that achieves test loss at most $2 - 2\epsilon$.

1.D. (10 PTS.) In the previous pset you used the existence of an exponentially large family of pairwise separated unit vectors. We will use it again here. Formally, you may use that for any ϵ, there exists a set S of $\Omega(e^{C\epsilon^2 d})$ vectors in $\mathbb{S}^{d-1}$, for some absolute constant $C > 0$, such that any two distinct vectors $u, v \in S$ satisfy $|\langle u, v\rangle| \leqslant \epsilon$.

Prove that for any $q \geqslant 1$, any CSQ algorithm for supervised learning functions of the form $\sigma(\langle w, \cdot\rangle)$ over Gaussian examples requires at least q queries or tolerance at most $O(\log^{\ell^*/4}(qd)/d^{\ell^*/4})$ to produce a function that achieves test loss at most 1.

Solution:

1.A.

1.B.

1.C.

1.D.

2 (70 PTS.) A SIMPLE PHASE TRANSITION (11/4)

Motivation: In this problem we will explore a simple example of a phase transition for an Ising model.

Setup: Given inverse temperature $\beta \geqslant 0$, consider the Gibbs measure

$$\mu_{n,\beta}(x) = \frac{1}{Z_n(\beta)} \exp\Big(\frac{\beta}{n} \sum_{1 \leqslant i < j \leqslant n} x_i x_j\Big), \quad x \in \{\pm 1\}^n,$$

where $Z_n(\beta) \triangleq \sum_x \exp(\frac{\beta}{n} \sum_{1 \leqslant i < j \leqslant n} x_i x_j)$ is the partition function.

Given x, we define its *magnetization* by $\overline{x} \triangleq \frac{1}{n} \sum_i x_i \in [-1, 1]$.

We will show that above a critical temperature, the magnetization of a typical sample from the Gibbs measure is concentrated around 0, but below the critical temperature, the magnetization concentrates around two distinct values.

2.A. (20 PTS.) For $m \in [-1, 1]$, define $\psi_\beta(m) \triangleq \beta m^2/2 + H(\frac{1+m}{2})$, where $H(z) \triangleq -z \ln z - (1-z)\ln(1-z)$ denotes the entropy function. Prove that for any $m \in \{-n/n, (-n+2)/n, \ldots, (n-2)/n, n/n\}$,

$$\frac{1}{n+1} \leqslant \Pr_{x \sim \mu_{n,\beta}} [\overline{x} = m] \cdot Z_n(\beta) e^{-n\psi_\beta(m) + \beta/2} \leqslant 1.$$

(**Hint:** You may find the bound $\frac{1}{n+1} e^{nH(m)} \leqslant \binom{n}{mn} \leqslant e^{nH(m)}$ useful.)

Next, we will show that the free energy $\log Z_n(\beta)$ can be approximated variationally in terms of the ψ_β functional.

2.B. (10 PTS.) Let $\phi_*(\beta) \triangleq \sup_{m \in [-1,1]} \psi_\beta(m)$. Use the previous part to prove that

$$|\log Z_n(\beta) - n\phi_*(\beta) + \beta/2| \lesssim \log(n).$$

2.C. (30 PTS.) Compute the maximizers of ψ_β over $m \in [-1, 1]$. Your answer will depend on whether β lies above or below some threshold β_c, which you should identify. For $\beta > \beta_c$ larger than a threshold, you will not be able to express the solutions analytically, so provide a simple equation in m whose solutions correspond to the maximizers of ψ_β.

2.D. (10 PTS.) Use the preceding parts to prove that for any constant $\epsilon > 0$, there is a $\delta_{\beta,\epsilon} > 0$ which is *independent of* n such that

- When $\beta \leqslant \beta_c$,

$$\Pr_{x \sim \mu_{n,\beta}} [|\overline{x}| \leqslant \epsilon] \geqslant 1 - e^{-n\delta_{\beta,\epsilon}}$$

- When $\beta > \beta_c$, there exists a constant $0 < m_*(\beta) \leqslant 1$ such that

$$\Pr_{x \sim \mu_{n,\beta}} [|\overline{x} - m_*(\beta)| \leqslant \epsilon] = \Pr_{x \sim \mu_{n,\beta}} [|\overline{x} + m_*(\beta)| \leqslant \epsilon] = 1/2 - e^{-n\delta_{\beta,\epsilon}}$$

2.E. (0 PTS.) [Optional] Carry out an analogous computation to the preceding parts for the Gibbs measure given by

$$\mu_{n,\beta,h}(x) \propto \exp\Big(\frac{\beta}{n} \sum_{1 \leqslant i < j \leqslant n} x_i x_j + \langle h, x \rangle\Big)$$

for parameter $h > 0$.

Solution:

2.A.

2.B.

2.C.

2.D.

2.E.