STATS 700, Fall 2024
LLMs and Transformers
HW 2

Ambuj Tewari

Due September 10

You are allowed to use LLMs provided you declare precisely how used them (but beware that they can hallucinate). You cannot search online or ask anyone. Discussing problems with other students in the course is allowed but all solutions should be your own work. If two students submit verbatim copies of any solution, it will be treated as a violation of academic integrity. Submit your solutions as a typeset PDF file (no handwritten scans please) on Canvas.

1 Relative Entropy is a Bregman Divergence (5 pts)

Let $f : S \to \mathbb{R}$ be a twice differentiable convex function defined on some convex set $S \subseteq \mathbb{R}^d$. Its Bregman divergence is defined as:

$$D_f(x\|y) = f(x) - f(y) - \nabla f(y)^\top (x - y) \ .$$

Show that (1 pt each except item 2 which is 2 pts):

1. Bregman divergence is always non-negative (using the fact that f is convex and differentiable).
2. If f is strongly convex, i.e., $\nabla^2 f \succeq \mu I$ (we say $A \succeq B$ for symmetric matrices if $A - B$ is positive semidefinite), then $D_f(x\|y) \geq \frac{\mu}{2}\|x - y\|_2^2$ where $\|x\|_2 = \sqrt{x^\top x}$.
3. When $f(x) = \frac{1}{2}\|x\|_2^2$, the Bregman divergence is $\frac{1}{2}\|x - y\|_2^2$.
4. When f is the entropy function, the Bregman divergence D_f is relative entropy.

2 Entropy of Powers (2 pts)

Let k be a positive integer and X be a real-valued random variable taking values in a finite set. What is the best general relationship that you can establish between the entropy of X and the entropy of X^k?

3 Entropy of Random Graphs (3 pts)

The classic Erdos-Renyi random (undirected) graph model on n vertices comes in two varieties. If we choose each possible edge independently with probability p, we get the $G(n, p)$ model. If we choose m edges uniformly at random from the total possible $\binom{n}{2}$ edges, we get the $G(n, m)$ model. Compute the entropy of an Erdos-Renyi random graph under both models. Show that the two entropies are close if $m = \binom{n}{2}p$ and n is large.