CSE 547 - Assignment 1

Philip Pham

May 10, 2018

Problem 0

List of collaborators: I have not collaborated with anyone.

List of acknowledgements: None.

Certify that you have read the instructions: I have read and understood these policies.

Problem 1: Gaussian Random Projections and Inner Products

Let $\phi(x) = \frac{1}{\sqrt{m}}Ax$ represent our random projection of $x \in \mathbb{R}^d$, with A an $m \times d$ projection matrix with each entry sampled i.i.d from $N(0,1)$. Note that each row of A is a random projection vector, $v^{(i)}$.

The *norm preservation theorem* states that for all $x \in \mathbb{R}^d$, the norm of the random projection $\phi(x)$ approximately maintains the norm of the original x with high probability:

$$\mathbb{P}\left((1-\epsilon)\|x\|^2 \leq (1+\epsilon)\|x\|^2 \leq 1 - 2\exp\left(-\left(\epsilon^2 - \epsilon^3\right)m/4\right)\right), \tag{1}$$

where $\epsilon \in (1, 1/2)$.

Using the norm preservation theorem, prove that for any $u, v \in \mathbb{R}^d$ such that $\|u\| \leq 1$ and $\|v\| \leq 1$,

$$\mathbb{P}\left(|u \cdot v - \phi(u) \cdot \phi(v)| \geq \epsilon\right) \leq 4\exp\left(-\left(\epsilon^2 - \epsilon^3\right)m/4\right) \tag{2}$$

Proof. □

Problem 2: Locality-Sensitive Hashing (LSH) for Angle Similarity

Suppose our set of n points $D = \{p_1, \ldots, p_n\}$ are vectors in d dimensions. Our problem is: given a query point q find a point $p \in D$, which has a small angle with q. Recall that the angle between two vectors a and b is $\cos^{-1}\left(\frac{a \cdot b}{\|a\|\|b\|}\right)$.

As doing this exactly may be computationally expensive, let us try to do this approximately with a fast algorithm. The approximate objective is as follows: suppose there exists a point $p \in D$ which has cosine similarity larger than θ, then our goal is return a point with cosine similarity greater than $c\theta$.As doing this exactly may be computationally expensive, let us try to do this approximately with a fast algorithm. The approximate objective is as follows: suppose there exists a point $p \in D$ which has cosine similarity larger than θ, then our goal is return a point with cosine similarity greater than $c\theta$.

Let us try to do this with LSH. Let us consider the a family of hash functions, where $h(p) = \operatorname{sign}(u \cdot p)$ where we will sample u uniformly at random from a Gaussian (or from a unit sphere).

1. Provide an exact expression for $\mathbb{P}(h(p) = h(p'))$ based on some geometric relation between p and p'.

   ### Solution

   Define
   $$\operatorname{angle}(u, v) = \cos^{-1}\left(\frac{u \cdot v}{\|u\|\|v\|}\right), \tag{3}$$
   which is the angle between two vectors.

   Then,
   $$\boxed{\mathbb{P}(h(p) = h(p')) = 1 - \frac{\operatorname{angle}(p, p')}{\pi}.} \tag{4}$$

2. Provide an expression for P_1 and P_2 in terms of θ and $c\theta$. Note that since we want a large angle

3.