UNIVERSITÄT DES SAARLANDES
FACHRICHTUNG MATHEMATIK
Dr. Marwa Banna

High Dimensional Probability
with Applications to Big Data Sciences

Summer term 2020

Assignment 4: Part 1
Due: Monday, June 15, 2020

Problem 15 (Exercise from lecture: dimension reduction).
Let $\mathcal{X}$ be a set of N points in $\mathbb{R}^n$ and $\epsilon \in (0,1)$. Consider an $m \times n$ matrix A whose rows X_i are independent, mean zero, isotropic and sub-Gaussian random vectors. Consider the *Gaussian random projection:*

$$P = \frac{1}{\sqrt{m}} A$$

and assume that $m \geq C\epsilon^{-2} \log N$ where C is an appropriately large constant depending only on the sub-Gaussian norms of the vectors X_i. The Johnson-Linderstrauss Lemma then states: with high probability, say 0.99, the map P preserves the distances between all points of $\mathcal{X}$ with error ϵ; i.e.

$$(1-\epsilon)\|x-y\|_2 \leq \|Px - Py\|_2 \leq (1+\epsilon)\|x-y\|_2 \qquad \text{for all } x, y \in \mathcal{X}.$$

1. Show that it is sufficient to prove that, with high probability,

   $$1-\epsilon \leq \|Pz\|_2^2 \leq 1+\epsilon \qquad \text{for all } z \in T,$$

   where

   $$T := \left\{ \frac{x-y}{\|x-y\|_2} : \quad x, y \in \mathcal{X} \text{ distinct points} \right\}.$$

2. Prove that

   $$\mathbb{P}\left\{ \max_{z \in T} \left| \frac{1}{m} \sum_{i=1}^{m} \langle X_i, z \rangle^2 - 1 \right| > \epsilon \right\} \leq |T| \cdot 2\exp(-c\epsilon^2 m),$$

   for some constant $c > 0$.

3. Deduce.