## 0368-3248-01-Algorithms in Data Mining

Fall 2012

Lecture 6: Assignment 2

Lecturer: Edo Liberty

Warning: This note may contain typos and other inaccuracies which are usually discussed during class. Please do not cite this note as a reliable source. If you find mistakes, please inform me.

## 1 Weak random projections

## setup

In this question we will construct a simple and weak version of random projections. That is, given two vectors  $x, y \in \mathbb{R}^d$  we will find two new vectors  $x', y' \in \mathbb{R}^k$  such that from x' and y' we could approximate the value of ||x-y||. The idea is to define k vectors  $r_i \in \mathbb{R}^d$  such that each  $r_i(j)$  takes a value in  $\{+1, -1\}$  uniformly at random. Setting  $x'(i) = r_i^T x$  and  $y'(i) = r_i^T y$  the questions will lead you through arguing that  $\frac{1}{k}||x'-y'||_2^2 \approx ||x-y||_2^2$ .

## questions

- 1. Let z = x y, and z' = x' y'. Show that  $z'(\ell) = r_{\ell}^T z$  for any index  $\ell \in [1, \dots, k]$ .
- 2. Show that  $E\left[\frac{1}{k}||z'||_2^2\right] = E\left[(z'(\ell))^2\right] = ||z||_2^2$ .
- 3. Show that

$$Var[(z'(\ell))^2] \le 4||z||_2^4.$$

Hint: for any vector w we have  $||w||_4 \leq ||w||_2$ .

4. From 3 (even if you did not manage to show it) claim that

$$\operatorname{Var}\left[\frac{1}{k}||z'||_2^2\right] \le 4||z||_2^4/k.$$

5. Use 3 and Chebyshev's inequality do obtain a value for k for which:

$$(1-\varepsilon)||x-y||_2^2 \le \frac{1}{k}||x'-y'||_2^2 \le (1+\varepsilon)||x-y||_2^2$$

with probability at least  $1 - \delta$ .