Prof. Dr. Johannes Blömer 22.10.2015

submission due: 5.11.2015, F1.110, 11:15

# Clustering Algorithms

WS 2015/2016

Handout 1

The *square of the euclidian distance* between $x, y \in \mathbb{R}^d$ is given by

$$D_{l_2^2}(x,y) := \|x-y\|_2^2 = \left(\sum_{i=1}^{d}(x_i - y_i)^2\right).$$

Remember, the *scalar product* of $x, y \in \mathbb{R}^d$ is defined by

$$\langle x, y\rangle = \sum_{i=1}^{d} x_i \cdot y_i$$

and has the following properties:

(1) $\langle x, y\rangle = \langle y, x\rangle$

(2) given $z \in \mathbb{R}^d$, $\langle x+z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$

(3) given $a \in \mathbb{R}$, $\langle ax, y\rangle = a\,\langle x, y\rangle$

**Exercise 1:**
Let $X \subseteq \mathbb{R}^d$. We denote the mean of set $X$ by $\mu(X) = \frac{1}{n}\sum_{x\in X} x$.

(a) Show that for each $m \in \mathbb{R}^d$

$$\sum_{x\in X}\|x-m\|_2^2 = \sum_{x\in X}\|x-\mu(X)\|_2^2 + |X|\cdot\|\mu(X)-m\|_2^2\ .$$

(b) Which $m \in \mathbb{R}^d$ minimizes $\sum_{x\in X}\|x-m\|_2^2$?

*Hint:* $\|x\|_2^2 = \langle x, x\rangle$.

**Exercise 2:**
The squared euclidean distance does not satisfy the the triangle inequality, but a relaxed version of the said inequality. Prove that for each $p, q, r \in \mathbb{R}^d$ we have

$$\|p-q\|_2^2 \le 2\|p-r\|_2^2 + 2\|r-q\|_2^2.$$

*Hint:* Use the triangle inequality.

**Exercise 3:**

a) Show that $D_{l_2^2}$ is a reflexive distance function.

b) Give a counterexample showing that $D_{l_2^2}$ does not satisfy the triangle inequality.