DATA MINING (CSCI-B 565)
Assignment 1
Data Science
Fall 2017
Indiana University Bloomington

Abhishek Rapelli
arapelli@iu.edu

October 23, 2017

All the work herein is solely mine

1 Given that $d(x, y) = |x - y|$, we have to prove that this is a metric. In order to prove that it is a metric, we shall have to prove the following from the definition for a equation to be called a metric. Let X be a finite, non-empty set and $\forall x, y, z \in X$ and $d : X^2 \to \mathbb{R}$ where $\mathbb{R}$ is the set of real numbers.

$$d(x, y) \geq 0$$
$$d(x, y) = 0 \leftrightarrow x = y$$
$$d(x, y) = d(y, x)$$
$$d(x, z) \leq d(x, y) + d(y, z)$$

If we prove the above conditions, it is called a **metric**. Given that, $d(x, y) = |x - y|$
a) As the absolute value of (x - y) is a non-negative value,it is alwaysgreater than or equal to zero.
b) If x = y then $|x - y| = 0$

c)

$$
\begin{aligned}
&d(x,y) = |x-y|,\\
&d(y,x) = |y-x|\\
&\quad = |-(x-y)|\\
&\quad = |x-y|\\
&\therefore d(x,y) = d(y,x)
\end{aligned}
$$

d)

$$
\begin{aligned}
&d(x,z) \leq d(x,y) + d(y,z)\\
&\text{Case-1}\\
&|x-z| \leq |x-y| + |y-z|\\
&+(x-z) \leq +(x-y) + (+(y-z))\\
&x - z \leq x - y + y - z\\
&\text{By simplification,}\\
&x - z \leq x - z\\
&\text{Case-2}\\
&|x-z| \leq |x-y| + |y-z|\\
&-(x-z) \geq -(x-y) + (-(y-z))\\
&z - x \geq y - x + z - y\\
&\text{By simplification,}\\
&z - x \geq z - x\\
&\therefore d(x,z) \leq d(x,y) + d(y,z)
\end{aligned}
$$

Hence we proved that $|x-y|$ is a *metric*.

2 Intra-block distances (ℓ^2 – norm or Euclidean distances) between all the unique pairs formed by the given four points.

$$d((x_1,x_2),(y_1,y_2)) \;=\; [(x_1-y_1)^2 + (x_2-y_2)^2]^{\frac{1}{2}}$$

Partition	Total Intra-block Distance
$\{\{(0,1),(2,1)\},\{(3,1),(5,5)\}\}$	$2+2\sqrt{5}$ (Given)
$\{\{(0,1),(3,1)\},\{(2,1),(5,5)\}\}$	$3+5=8$
$\{\{(0,1),(5,5)\},\{(3,1),(2,1)\}\}$	$1+\sqrt{41}$

3 Output of the given R code

```
x <-seq(1,50,by=2)
y <- 2*x - 30
png("Abhishekplot.png")
plot(x,y)
dev.off()
```

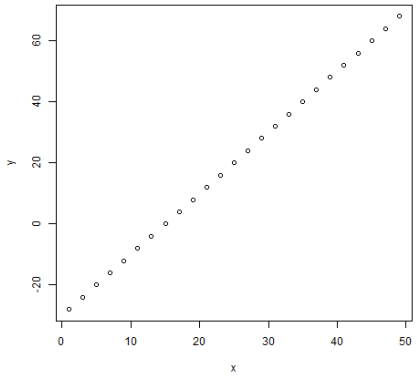

Figure 1: My first graph on Latex

References

- Lectures and notes of Prof.M.M. Dalkilic
- "Introduction to Data Mining" by Tan, Steinbach and Kumar