PHYS 598 SDA RECITATION 2 - PROBLEM SET

1. Consider a p-dimensional vector $\boldsymbol{X} = \{X_1, \ldots, X_p\}$ and a q-dimensional vector $\boldsymbol{Y} = \{Y_1, \ldots, Y_q\}$. Then the **covariance matrix** $\Sigma(\boldsymbol{X})$ is defined by

$$\Sigma_{i,j}(\boldsymbol{X}) = \mathrm{cov}(X_i, X_j).$$

Similarly, define the *cross*-covariance matrix $\Sigma_{i,j}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{cov}(X_i, Y_j)$. Your task is to convince yourself of the following properties of Σ:

(a) $\Sigma(\boldsymbol{X}, \boldsymbol{Y}) = \Sigma(\boldsymbol{Y}, \boldsymbol{X})^T$. (C'mon, obviously!)

(b) Linearity, $\Sigma(\alpha\boldsymbol{X}, \boldsymbol{Y}) = \alpha\Sigma(\boldsymbol{X}, \boldsymbol{Y})$, where $\alpha \in \mathbb{R}$, etc.

(c) If $p = q$, then $\Sigma(\boldsymbol{X} + \boldsymbol{Y}) = \Sigma(\boldsymbol{X}) + \Sigma(\boldsymbol{X}, \boldsymbol{Y}) + \Sigma(\boldsymbol{Y}, \boldsymbol{X}) + \Sigma(\boldsymbol{Y})$. This should follow from (b).

(d) For a $q \times p$ matrix $\boldsymbol{A}$ and a q-dim vector $\boldsymbol{a}$, $\Sigma(\boldsymbol{AX} + \boldsymbol{a}) = \boldsymbol{A}\Sigma(\boldsymbol{X})\boldsymbol{A}^T$.

2. In this problem, we shall explore **Clustering**. On your laptop/tablet, open the link `http://www.cs.princeton.edu/courses/archive/spr08/cos424/slides/clustering-1.pdf`. The aim is to get through this document (until slide 23) and understand *k-means* and the *k-mediods* algorithms.

3. You are given a power grid, that is a square of side S ($\leq 10^9$). You are also given N special points ($N \leq 5000$). Each of the N special points has a different power R_i ($\leq 10^9$) associated with it. Each of these points can influence all points of the grid within a Manhattan distance of R_i (if the Manhattan distance between the special point and any other point is $\leq R_i$, it will be influenced). Find out the expected number of points that are influenced in the grid.