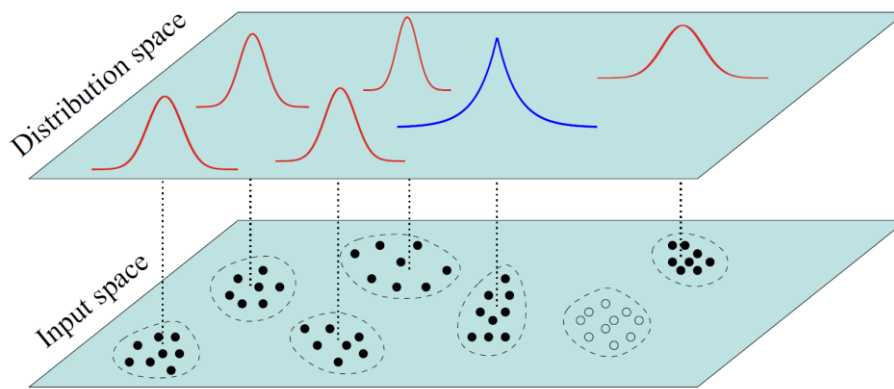


# One-Class Support Vector Machines for Group Anomaly Detection – Applications in eHealth and Bioinformatics –

**Project Description:** Anomaly detection is one of the most important tools in all data-driven scientific disciplines. Data that do not conform to the expected behaviors often have some interesting characteristics and can help domain experts better understand the problem at hand. And in the era of big data, the anomaly may appear not only in the data themselves, but also as a result of their interactions. In this project, we are interested in studying group anomaly detection, by treating it as detecting anomalous distributions based on their empirical samples. The intuition here is to compute a hyperplane separating the origin from the data in a kernelized feature space; which results in a nonlinear decision boundary encompassing the original data.



**Target dataset.** Applications in computational biology will turn around detecting a cluster from single RNAseq datasets to see if it is a real cluster or a noise-cluster (aberrant cluster).

**Learning Opportunities.** Students will enhance their programming skills in Python and acquire new knowledge in support-vector machines; data distributions; anomaly detection; etc.

**Duties/Activities:** The interns will run and test different instances of the machine learning code on real data. Some code in python will be provided.

**Required Skills:** Python and/or Matlab

**Preferred Intern Academic Level:** B.Sc. 3<sup>rd</sup> year, and/or B.Sc. 4<sup>th</sup> year.

**Expected Team Size:** 2 students

**Mentors:** Dr. Abdelkader Baggag <[abaggag@hbku.edu.qa](mailto:abaggag@hbku.edu.qa)>  
Dr. Halima Bensmail <[hbensmail@hbku.edu.qa](mailto:hbensmail@hbku.edu.qa)>