⇨ Energy Stats and kernels

Energy statistics~ based on 'statistical potential energy between potential distributions'
- this 'potential energy' is different when distributions are different vs the same so it is a method of testing for distribution equality

Q: what are metric spaces of negative type?
key point – because generalized energy distances and kernel methods have been demonstrated equivalent, one can apply generalized energy distances to ML

Kernels with ML
- choosing the kernel = biggest challenge

K-means
- Lloyd's approach = computes the means of each cluster and assigns points according to nearness to cluster mean
    o fast
- Hartigan's approach = for each point assign it to a cluster so that a loss function is minimized
    o Can escape local minima

Kernel k groups ~ extending Hartigan's approach to kernel spaces

<u>Energy Statistics</u>
Given X, X' $\overset{iid}{\underset{\sim}{}}$ P, and Y, Y' $\overset{iid}{\underset{\sim}{}}$ Q
Then energy distance = $\varepsilon(Q, P) = 2E\|X - Y\| - E\|X - X'\| - E\|Y - Y'\|$

- energy distance characterizes the distributions equality
- $\varepsilon^{1/2}$ is a metric on the space of distributions
- raising each term in ε to α generalizes the equation
- both ε and its generalization are non negative

note: for α between 0 and 2 if $\varepsilon = 0$ then P = Q but for α = 2 ε = 0 just means the means of X and Y are equivalent

Q: what is a semimetric of negative type? Or even what is just a semimetric?
    -> there is an equation but I don't have great intuition about what it means graphically…

- what I did gather though is that this 'space of negative type' you can create by using a 'semimetric of negative type' enables further generalization of the energy distance
- it also allows use to map from the space of negative type to a Hilbert space which can be used in kernels

Key idea as far as I understand:
1. you can embed a probability measure into the RKHS

2. then you can 'measure' the distance between different probability measures using something called MMD (maximum mean discrepancy)
3. Then, based off of a relationship between symmetric positive definite kernels and semimetrics of negative type its possible to equate generalized energy distance and the inner product on the RKHS
4. To tie it up in a neat bow then there also exists a test statistic for the equality of distributions using this method

⇨ Clustering!