Data Science Toolbox Question Sheet

04.1 Non-parametrics

Daniel Lawson

Block 4

Short Questions

1. Methods that perform data transformation are often said to be parameter free. Discuss this claim.
2. List and briefly describe 3 transforms that might be applied to discrete or continuous time-series data.
3. Explain kernel smoothing. What concerns arise when moving to multiple dimensions?
4. In the equation for kernel density estimation we write $K_H(y - x_i) = K(H^{-1}(y - x_i))/det(H)$. What is the 1-D equivalent? What is the role of H?
5. Given H being diagonal and a Uniform kernel, what shape is made?
6. A colleague claims that kernel density estimation makes no assumptions about the data. Discuss in what sense this is and isn't true.
7. A colleague claims that computing the nearest neighbour search is $O(N^2)$. Discuss the senses in which they are correct or not.
8. Given the equation for the k-NN density estimate around a point x in d dimensions and with k neighbours considered:

$$\hat{p}_{kNN}(x) = \frac{k}{N} \cdot \frac{1}{V_d R_k^d(x)}.$$

   Interpret the quantity $R_k^d(x)$.
9. In what sense is k-Nearest Neighbours parameter free? Is it making any assumptions about the data?
10. Describe the motivation of the kernel trick.
11. Consider $K(x_i, x_j) = |x_i - x_j|$ for x_i, $x_j \in \mathbb{R}$.
    - Compute the gram matrix for the dataset $D = (0, 1)$.
    - Compute the eigenvalues of the gram matrix. Is it positive semi-definite?

- Kernel spaces are closed under addition, multiplication of a scalar, kernel product, functional product, repeated application of a kernel, and matrix operations. What is the advantage of this?
- List three methods that can use the kernel trick.
- In what way is kernel PCA computed differently to regular PCA?