Sixth week practicals in Machine learning 1 – 2022 – Paper 1

1 Kernel outlier Detection (October)

Consider the picture in Figure 2. The dots represent data items. Our task is to derive an algorithm that will detect the outliers (in this example there are 2 of them). To that end, we draw a circle rooted at location $\boldsymbol{a}$ and with radius R. All data-cases that fall outside the circle are detected as outliers

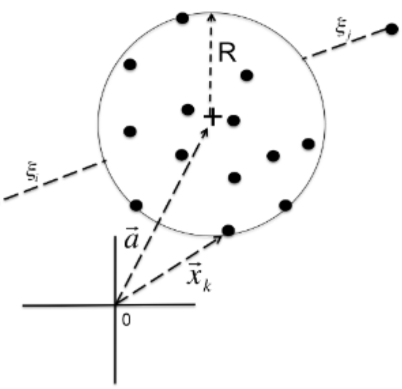

Figure 2: Kernel Outlier Detection

We will now write down the primal program that will find such a circle:

$$\min_{\boldsymbol{a},R,\boldsymbol{\xi}} R^2 + C\sum_{i=1}^{N} \xi_i$$

$$s.t.\ \forall i : \|\boldsymbol{x_i} - \boldsymbol{a}\|^2 \leq R^2 + \xi_i,\ \ \xi_i \geq 0$$

In words: we want to minimize the radius of the circle subject to the constraint that most data cases should lay inside it. Outliers are allowed to stay outside, but they pay a price proportional to their distance from the circle boundary and C. Answer the following questions:

(a) Introduce Lagrange multipliers for the constraints and write down the primal Lagrangian. Use the following notation: $\{\alpha_i\}$ are the Lagrange multipliers for the first constraint and $\{\mu_i\}$ for the second constraint.

(b) Write down all stationary and KKT conditions. (Hint: take the derivative w.r.t. R^2 instead of R).

(c) Use these conditions to derive which data-cases $\boldsymbol{x_i}$ will have $\alpha_i > 0$ and which ones will have $\mu_i > 0$.

(d) Derive the dual Lagrangian and specify the dual optimization problem. Kernelize the problem, i.e. write the dual program only in terms of kernel entries and Lagrange multipliers.

(e) The dual program will return optimal values for $\{\alpha_i\}$. Assume that at least one of these is such that $0 < \alpha_i < C$. In terms of the optimal values for α_i, compute the optimal values for the other dual variables $\{\mu_i\}$.

Then, solve the primal variables $\{\boldsymbol{a}, R, \boldsymbol{\xi}\}$ (in that order) in terms of the dual variables $\{\mu_i, \alpha_i\}$. Note that you do not need to know the dual optimization program to solve this question. You only need the KKT conditions.

(f) Assume we have solved the dual program. We now want to apply it to new test cases. Describe a test in the dual space (i.e. in terms if kernels and Lagrange multipliers) that could serve to detect outliers. (Students who got stuck along the way may describe the test in primal space).

(g) What kind of solution do you expect if we use $C = 0$. And what solution if we use $C = \infty$?