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

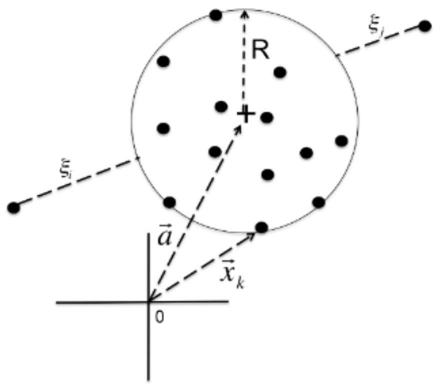

Figure 2: Kernel Outlier Detection

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.

Answer:

$$\mathcal{L}(\boldsymbol{a}, R, \xi, \alpha, \mu) = R^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i \left(R^2 + \xi_i - \|\boldsymbol{x}_i - \boldsymbol{a}\|^2\right) - \sum_{i=1}^{N} \mu_i \xi_i.$$

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

Answer: The first three conditions are obtained by setting the derivative of the primal Lagrangian to zero with respect to R^2, a and ξ_i.

$$\frac{\partial \mathcal{L}}{\partial R^2} = 1 - \sum_{i=1}^{N} \alpha_i$$

$$\frac{\partial \mathcal{L}}{\partial R^2} = 0 \implies \sum_{i=1}^{N} \alpha_i = 1.$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}} = -2 \sum_{i=1}^{N} \alpha_i (\boldsymbol{x}_i - \boldsymbol{a})$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{a}} = 0 \implies \sum_{i=1}^{N} \alpha_i \boldsymbol{x}_i = \boldsymbol{a} \sum_{i=1}^{N} \alpha_i \implies \boldsymbol{a} = \sum_{i=1}^{N} \alpha_i \boldsymbol{x}_i$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C - \alpha_i - \mu_i$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \implies C - \alpha_i - \mu_i = 0, \forall i$$

Note that the three conditions above are **not** KKT conditions. The KKT conditions for this model are given below:

- $R^2 + \xi_i - \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 \geq 0 \ \forall i.$
- $\xi_i \geq 0 \ \forall i.$
- $\alpha_i \geq 0 \ \forall i.$
- $\mu_i \geq 0 \ \forall i.$
- $\alpha_i \left(R^2 + \xi_i - \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 \right) = 0 \ \forall i.$
- $\mu_i \xi_i = 0 \ \forall i.$

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

Answer: The complementary slackness conditions are:

Inside the ball: $\xi_i = 0,\ R^2 + \xi_i - \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 > 0 \implies \alpha_i = 0 \implies \mu_i = C > 0$

Outside the ball: $\xi_i > 0,\ R^2 + \xi_i - \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 = 0 \implies \mu_i = 0 \implies \alpha_i = C > 0$

On the ball: $\xi_i = 0,\ R^2 + \xi_i - \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 = 0 \implies \mu_i \geq 0,\ \alpha_i \geq 0,\ \mu_i + \alpha_i = C$

(*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.

Answer: We use conditions a, b, and c to eliminate R^2, a, and ξ_i from the primal Lagrangian to obtain the dual representation.

$$\mathcal{L}(\mathbf{a}, R, \xi, \alpha, \mu) = R^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left(R^2 + \xi_i - \|\boldsymbol{x}_i - \boldsymbol{a}\|^2\right) - \sum_{i=1}^{N}\mu_i\xi_i$$

$$= R^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i^T\boldsymbol{x}_i - 2\sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i^T\boldsymbol{a} + \sum_{i=1}^{N}\alpha_i\boldsymbol{a}^T\boldsymbol{a}$$

$$-\sum_{i=1}^{N}\alpha_i R^2 - \sum_{i=1}^{N}\alpha_i\xi_i - \sum_{i=1}^{N}\mu_i\xi_i$$

$$= \left(R^2 - \sum_{i=1}^{N}\alpha_i R^2\right) + \sum_{i=1}^{N}(C - \alpha_i - \mu_i)\xi_i$$

$$-2\left(\sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i^T\right)\boldsymbol{a} + \left(\sum_{i=1}^{N}\alpha_i\right)\boldsymbol{a}^T\boldsymbol{a} + \sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i^T\boldsymbol{x}_i$$

$$= \sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i^T\boldsymbol{x}_i - \boldsymbol{a}^T\boldsymbol{a}$$

$$= \sum_{i=1}\alpha_i\|\boldsymbol{x}_i\|^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\boldsymbol{x}_i^T\boldsymbol{x}_j$$

In the first step, we expand the expression for the primal Lagrangian, in the second step we rearrange the terms, finally, we apply the conditions. Kernelize the problem:

$$\sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i^T\boldsymbol{x}_i - \boldsymbol{a}^T\boldsymbol{a} = \sum_{i=1}\alpha_i\|\boldsymbol{x}_i\|^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\boldsymbol{x}_i^T\boldsymbol{x}_j$$

$$= \sum_{i=1}^{N}\alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}_i) - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}_j)\alpha_j$$

The dual program is:

$$argmax_\alpha \sum_{i=1}^{N}\alpha_i K_{ii} - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K_{ij}$$

with $\alpha_i \in [0, C]$ $\forall i$.

(*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.

Answer:

Note that when $0 < \alpha_i < C$, then $\boldsymbol{x}_i$ must be on the ball. When $\alpha_i = 0$, then $\boldsymbol{x}_i$ is on or inside the ball, and when $\alpha_i = C$ then $\boldsymbol{x}_i$ is on or outside the ball.

$$\mu_i^* = C - \alpha_i^*$$

$$\boldsymbol{a}^* = \sum_{i=1}^{N} \alpha_i^* \boldsymbol{x}_i$$

$$R^{*2} = \|\boldsymbol{x}_i - \boldsymbol{a}^*\|^2 \text{ (when } 0 < \alpha_i^* < C)$$

$$\xi_i^* = \begin{cases} \|\boldsymbol{x}_i - \boldsymbol{a}^*\|^2 - R^{*2} \text{ (when } \alpha_i^* = C) \\ 0 \text{ (when } \alpha_i^* = 0) \end{cases}$$

(*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).

Answer: A new test case $\boldsymbol{x}_t$ is an outlier when:

$$\|\boldsymbol{x}_t - \boldsymbol{a}^*\|^2 > R^{*2}$$

$$\boldsymbol{x}_t^T \boldsymbol{x}_t - 2\boldsymbol{x}_t^T \boldsymbol{a}^* + \boldsymbol{a}^{*T} \boldsymbol{a}^* > R^{*2}$$

$$K(\boldsymbol{x}_t, \boldsymbol{x}_t) - 2\sum_{i=1}^{N} \alpha_i^* \boldsymbol{x}_i^T \boldsymbol{x}_t + \sum_{i,j} \alpha_i \alpha_j \boldsymbol{x}_i^T \boldsymbol{x}_j > R^{*2}$$

$$K(\boldsymbol{x}_t, \boldsymbol{x}_t) - 2\sum_{i=1}^{N} \alpha_i^* K(\boldsymbol{x}_i, \boldsymbol{x}_t) + \sum_{i,j} \alpha_i \alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) > R^{*2}$$

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

Answer: When $C \to 0$ we expect $R \to 0$, and when $C \to \infty$ we expect R to be such that all data-cases are inside the ball.
