Mixed-integer linear programs for the SVR with feature and outlier detection

- **Input**: $(x^1, y_1), (x^2, y_2), \ldots, (x^k, y_k) \in \mathbb{R}^d \times \mathbb{R}$
- **User-chosen**: d_0 and k_0, nonzero naturals
- **Pre-processed** (from the input):
  - $R_1^U, R_2^U, \ldots R_k^U$
  - $W_1^L, W_2^L, \ldots, W_d^L$
  - $W_1^U, W_2^U, \ldots, W_d^U$

Basic model

$$
\begin{aligned}
\min \sum_{i=1}^{k} q_i & & & (1)\\
q_i + y_i - ax^i - z - t_i \geq 0 & & \forall i = 1, 2, \ldots, k & \quad (2)\\
-q_i + y_i - ax^i - z - t_i \leq 0 & & \forall i = 1, 2, \ldots, k & \quad (3)\\
t_i \geq -(1 - s_i) R_i^U & & \forall i = 1, 2, \ldots, k & \quad (4)\\
t_i \leq (1 - s_i) R_i^U & & \forall i = 1, 2, \ldots, k & \quad (5)\\
a_j \geq W_j^L f_j & & \forall j = 1, 2, \ldots, k & \quad (6)\\
a_j \leq W_j^U f_j & & \forall j = 1, 2, \ldots, k & \quad (7)\\
\sum_{j=1}^{d_0} f_j \leq d_0 & & & (8)\\
\sum_{i=1}^{k_0} s_j \geq k_0 & & & (9)\\
a_j \in \mathbb{R} & & \forall j = 1, 2, \ldots, d & \quad (10)\\
z \in \mathbb{R} & & & (11)\\
f_j \in \{0, 1\} & & \forall j = 1, 2, \ldots, d & \quad (12)\\
q_i \geq 0 & & \forall i = 1, 2, \ldots, k & \quad (13)\\
t_i \in \mathbb{R} & & \forall i = 1, 2, \ldots, k & \quad (14)\\
s_i \in \{0, 1\} & & \forall i = 1, 2, \ldots, k & \quad (15)
\end{aligned}
$$

Explanation.

- The pre-processed values R^U, W^L and W^U are such that an optimal solution $(a^\star, z^\star)$ satisfies
  - $|y_i - a^\star x^i - z^\star| \leq R_i^U$ for every $i = 1, 2, \ldots, k$
  - $W_j^L \leq a_j^\star \leq W_j^U$ for every $j = 1, 2, \ldots, d$
  - they are difficult to compute exactly, so we will use *heuristics* (in this first phase I will provide them)
- $s_i = 0$ if and only if point (x^i, y_i) is to be considered an outlier
- constraints (2) and (3) together imply that $-q_i \leq y_i - ax^i - z - t_i \leq q_i$, which is equivalent to $|y_i - ax^i - z - t_i| \leq q_i$. That is, q_i is an upper-bound on the prediction error made by our solution (a, z) on the point (x^i, y_i).
- from constraints (4) and (5) we obtain that $t_i = 0$ if and only if $s_i = 1$
- putting the last two observations together we obtain that if (x^i, y_i) is an outlier, then $s_i = 0$, then t_i can be nonzero and it can be used to "cancel" the error (using $t_i = y_i - ax^i - z$), thus $q_i \geq 0$. Since we minimize $\sum_{i=1}^k q_i$, $q_i = 0$ in an optimal solution for all outliers (it is never convenient to have a positive term when it can be zero!). So in an optimal solution we sum only the errors of the non-outliers.
- constraints (8) and (9) guarantee that no more than d_0 features are chosen, and no more than k_0 features are chosen.

Stronger model

$$\min \sum_{i=1}^{k}(p_i - (1-s_i)R_i^U) \tag{16}$$

$$p_i \geq (1-s_i)R_i^U \qquad \forall i = 1,2,\ldots,k \tag{17}$$

$$p_i \leq R_i^U \qquad \forall i = 1,2,\ldots,k \tag{18}$$

$$p_i + y_i - ax^i - z \geq 0 \qquad \forall i = 1,2,\ldots,k \tag{19}$$

$$-p_i + y_i - ax^i - z \leq 0 \qquad \forall i = 1,2,\ldots,k \tag{20}$$

$$a_j \geq W_j^L f_j \qquad \forall j = 1,2,\ldots,k \tag{21}$$

$$a_j \leq W_j^U f_j \qquad \forall j = 1,2,\ldots,k \tag{22}$$

$$\sum_{j=1}^{d_0} f_j \leq d_0 \tag{23}$$

$$\sum_{i=1}^{k_0} s_j \geq k_0 \tag{24}$$

$$a_j \in \mathbb{R} \qquad \forall j = 1,2,\ldots,d \tag{25}$$

$$z \in \mathbb{R} \tag{26}$$

$$f_j \in \{0,1\} \qquad \forall j = 1,2,\ldots,d \tag{27}$$

$$q_i \geq 0 \qquad \forall i = 1,2,\ldots,k \tag{28}$$

$$t_i \in \mathbb{R} \qquad \forall i = 1,2,\ldots,k \tag{29}$$

$$s_i \in \{0,1\} \qquad \forall i = 1,2,\ldots,k \tag{30}$$