Modelli MPHero

alberto.ceselli

December 2022

1 SVR-style models

Data:

- a feature space J
- a set I of points $x^i = (x^i_1 \dots x^i_{|J|})$, and corresponding response values y^i
- a violation cost C
- a confidence region parameter ϵ
- a target number of features to select k_0

Variables:

- hyperplane slopes $w = (w_1 \dots w_{|J|})$ and intercept z values
- feature binary selection decisions f_j (1 if feature j is selected, 0 otherwise)

Optimize:

$$
\begin{aligned}
\min\ & \frac{1}{2}\sum_{j\in J} w_j^2 + C \cdot \sum_{i \in I} p_i^+ + p_i^- && && (1)\\
\text{s.t. } & (\sum_{j\in J} w_j \cdot x_j^i) + z - y^i \leq \epsilon + p_i^+ && \forall i \in I && (2)\\
& -(\sum_{j\in J} w_j \cdot x_j^i) - z + y^i \leq \epsilon + p_i^- && \forall i \in I && (3)\\
& w_j \leq f_j W_j^U && \forall j \in J && (4)\\
& w_j \geq f_j W_j^L && \forall j \in J && (5)\\
& \sum_{j \in J} f_j \leq k_0 && && (6)\\
& p_i^+ \geq 0, p_i^- \geq 0 && \forall i \in I && (7)\\
& f_j \in \{0,1\} && \forall j \in J && (8)
\end{aligned}
$$

Where p_i^+, p_i^- are measurement errors on point i.

1.1 Outliers

Optimize:

$$\min \frac{1}{2}\sum_{j\in J} w_j^2 + C\cdot\sum_{i\in I}(p_i^+ + p_i^-)s_i \tag{9}$$

$$\text{s.t. } (\sum_{j\in J} w_j \cdot x_j^i) + z - y^i \le \epsilon + p_i^+ \qquad \forall i \in I \tag{10}$$

$$-(\sum_{j\in J} w_j \cdot x_j^i) - z + y^i \le \epsilon + p_i^- \qquad \forall i \in I \tag{11}$$

$$w_j \le f_j W_j^U \qquad \forall j \in J \tag{12}$$

$$w_j \ge f_j W_j^L \qquad \forall j \in J \tag{13}$$

$$\sum_{j\in J} f_j \le k_0 \tag{14}$$

$$\sum_{i\in I} s_i \ge s_0 \tag{15}$$

$$p_i^+ \ge 0, p_i^- \ge 0 \qquad \forall i \in I \tag{16}$$

$$f_j \in \{0,1\} \qquad \forall j \in J \tag{17}$$

$$s_i \in \{0,1\} \qquad \forall i \in I \tag{18}$$

The terms $(p_i^+ + p_i^-)s_i$ need to be linearized in the fashion of our previous regression models.

McCormick's linearization. We first perform a standard McCormick's linearization. Note that the linearization of products $p_i^+ s_i$ and $p_i^- s_i$ can be performed in the same way. Then, for the sake of brevity, we denote by $p_i^\pm s_i$ either product. Given a valid upper-bound $R_i^\pm$ on $p_i^\pm$, and since $s_i \in \{0,1\}$, it is well-known [**?**] that for every i we can replace $p_i^\pm s_i$ by an additional variable $t_i \ge 0$ and adding the following 3 sets of constraints for every i:

$$t_i^\pm \le p_i^\pm$$
$$t_i^\pm \le s_i R_i^\pm$$
$$t_i^\pm \ge p_i^\pm - (1 - s_i)R_i^\pm$$

we obtain a MILP whose solutions corresponds to those of (9)–(18). Hence a first linearization of (9)–(18) is:

$$\min \frac{1}{2}\sum_{j\in J} w_j^2 + C\cdot\sum_{i\in I}(t_i^+ + t_i^-) \tag{19}$$

$$\text{s.t. } (\sum_{j\in J} w_j \cdot x_j^i) + z - y^i \le \epsilon + p_i^+ \qquad \forall i \in I \tag{20}$$

$$-(\sum_{j\in J} w_j \cdot x_j^i) - z + y^i \le \epsilon + p_i^- \qquad \forall i \in I \tag{21}$$

$$t_i^+ \le p_i^+ \qquad \forall i \in I \tag{22}$$

$$t_i^+ \le s_i R_i^+ \qquad \forall i \in I \tag{23}$$

$$t_i^+ \ge p_i^+ - (1-s_i)R_i^+ \qquad \forall i \in I \tag{24}$$

$$t_i^- \le p_i^- \qquad \forall i \in I \tag{25}$$

$$t_i^- \le s_i R_i^- \qquad \forall i \in I \tag{26}$$

$$t_i^- \ge p_i^- - (1-s_i)R_i^- \qquad \forall i \in I \tag{27}$$

$$w_j \le f_j W_j^U \qquad \forall j \in J \tag{28}$$

$$w_j \ge f_j W_j^L \qquad \forall j \in J \tag{29}$$

$$\sum_{j\in J} f_j \le k_0 \tag{30}$$

$$\sum_{i\in I} s_i \ge s_0 \tag{31}$$

$$p_i^+ \ge 0, p_i^- \ge 0 \qquad \forall i \in I \tag{32}$$

$$f_j \in \{0,1\} \qquad \forall j \in J \tag{33}$$

$$s_i \in \{0,1\} \qquad \forall i \in I \tag{34}$$

$$t_i^+, t_i^- \ge 0 \qquad \forall i \in I \tag{35}$$

Disjunctive linearization (Projected McCormick's linearization). We observe that there exists an optimal solution to (19)–(35) satisfying (24) with equality for every $i \in I$. Indeed, if $s_i = 0$ then, given the validity of $R_i^{\pm}$, we may always set $p_i^{\pm} = R_i^{\pm}$ in an optimal solution (whose value depends only on the $t_i^{\pm}$ values); when $s_i = 1$ constraint (24) boils down to $t_i^{\pm} >= p_i^{\pm}$, which together with (25), gives $t_i^{\pm} = p_i^{\pm}$. So, we can restrict the McCormick's linearization to the set of solutions satisfying (24) with equality. This allows a projection of the $t_i^{\pm}$'s variables. After substitution we obtain the following valid MILP for the starting problem:

$$\min \frac{1}{2}\sum_{j\in J} w_j^2 + C\cdot\sum_{i\in I}(p_i^+ - (1-s_i)R_i^+) + C\cdot\sum_{i\in I}(p_i^- - (1-s_i)R_i^-) \tag{36}$$

$$\text{s.t. } (\sum_{j\in J} w_j\cdot x_j^i) + z - y^i \le \epsilon + p_i^+ \qquad \forall i\in I \tag{37}$$

$$-(\sum_{j\in J} w_j\cdot x_j^i) - z + y^i \le \epsilon + p_i^- \qquad \forall i\in I \tag{38}$$

$$R_i^+(1-s_i) \le p_i^+ \le R_i^+ \qquad \forall i\in I \tag{39}$$

$$R_i^-(1-s_i) \le p_i^- \le R_i^- \qquad \forall i\in I \tag{40}$$

$$w_j \le f_j W_j^U \qquad \forall j\in J \tag{41}$$

$$w_j \ge f_j W_j^L \qquad \forall j\in J \tag{42}$$

$$\sum_{j\in J} f_j \le k_0 \tag{43}$$

$$\sum_{i\in I} s_i \ge s_0 \tag{44}$$

$$p_i^+ \ge 0, p_i^- \ge 0 \qquad \forall i\in I \tag{45}$$

$$f_j \in \{0,1\} \qquad \forall j\in J \tag{46}$$

$$s_i \in \{0,1\} \qquad \forall i\in I \tag{47}$$

1.2 The Lagrangian Dual of the Disjunctive Linearization Model

In order to easily describe the dual formulation we introduce and fix some notations:

- v represents a generic variables (that is, any of $w_j, z, p_i^{\pm}, f_j, s_i$) and $\mathbf{v}$ is their vector (following the ordering $w_j, z, p_i^{\pm}, f_j, s_i$).
- $\Gamma\mathbf{v} \le \gamma$ is the constraint matrix of the available generic cuts obtained from the ILP. The column corresponding to variable v is indicated with Γ^v. Then Γ_h^v is the h-th entry of that column. We assume that the matrix Γ has $m \ge 0$ row.

- The x^i's are column vectors in $\mathbb{R}^d$, so that x^i_j is the j-th entry of x^i.

We define the duals of the disjunctive-based formulation.

- α^+ (resp. α^-) is the dual vector of constraints (37) (resp. (38))
- π^+ and ψ^+ (resp. π^- and ψ^-) are the dual vectors of upper- and lower-bound constraints (39) (resp. (40))
- λ^U and λ^L are the dual vectors of constraints (41) and (42) respectively
- β_1 and β_2 are the duals of constraints (43) and (44)
- η^+ and η^- are the dual vectors of constraints (45)
- φ (resp. σ) are the dual vectors of upper- and lower bound constraints $f_j \leq 1$ and $f_j \geq 0$ (resp. $s_i \leq 1$ and $s_i \geq 0$).

We have that the dual of the disjunctive-based formulation is:

$$\min \frac{1}{2}\sum_{j\in J}\left(\sum_{i\in I} x^i_j(\alpha^-_i - \alpha^+_i)\right)^2 + \sum_{j\in J}\left(\left(\lambda^L_j - \lambda^U_j - \sum_{h=1}^m \xi_h \Gamma_h^{\omega_j}\right)\sum_{i\in I} x^i_j(\alpha^-_i - \alpha^+_i)\right) + \frac{1}{2}\sum_{j\in J}\left(\lambda^L_j - \lambda^U_j - \sum_{h=1}^m \xi_h \Gamma_h^{\omega_j}\right)^2$$
$$- \sum_{i\in I}(\alpha^+_i(y_i+\epsilon) - \alpha^-_i(y_i - \epsilon)) - \sum_{i\in I}\left((\pi^+_i - \psi^+_i)R^+_i + (\pi^-_i - \psi^-_i)R^-_i\right) - k_0\beta_1 + s_0\beta_2$$
$$- \sum_{j\in J}\varphi_j - \sum_{i\in I}\sigma_i - \sum_{h=1}^m \xi_h\gamma_h \tag{48}$$

s.t.

$$\sum_{i\in I}(\alpha^+_i - \alpha^-_i) + \sum_{h=1}^m \xi_h \Gamma_h^z = 0 \tag{∂z}$$
$$\psi^+_i + \alpha^+_i - \pi^+_i - \sum_{h=1}^m \xi_h \Gamma_h^{p^+_i} \leq C \quad \forall i \in I \tag{∂p^+_i}$$
$$\psi^-_i + \alpha^-_i - \pi^-_i - \sum_{h=1}^m \xi_h \Gamma_h^{p^-_i} \leq C \quad \forall i \in I \tag{∂p^-_i}$$
$$\lambda^L_j W^L_j - \lambda^U_j W^U_j + \varphi_j + \beta_1 + \sum_{h=1}^m \xi_h \Gamma_h^{f_j} \geq 0 \quad \forall j \in J \tag{∂f_j}$$
$$\psi^+_i R^+_i + \psi^-_i R^-_i + \beta_2 - \sigma_i - \sum_{h=1}^m \xi_h \Gamma_h^{s_i} \leq C(R^+_i + R^-_i) \quad \forall i \in I \tag{∂s_i}$$
$$\alpha^\pm, \pi^\pm, \psi^\pm, \lambda^U, \lambda^L, \beta_1, \beta_2, \varphi, \sigma, \xi \geq \mathbf{0} \tag{49}$$