Locating Hyperplanes for Multiclass Classification

Víctor Blanco[1], **Alberto Japón**[2] y Justo Puerto[2]

Universidad de Granada[1]
Universidad de Sevilla[2]

Contents

Data

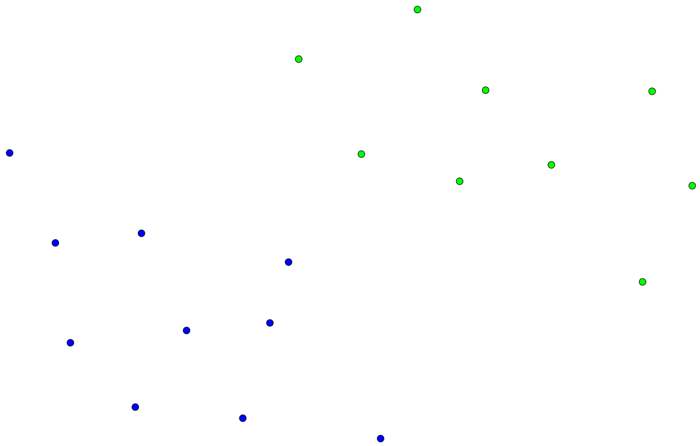
- Random sample of size n
- p predictor variables

$$x_i \in \mathbb{R}^p, \; i = 1, \ldots, n$$

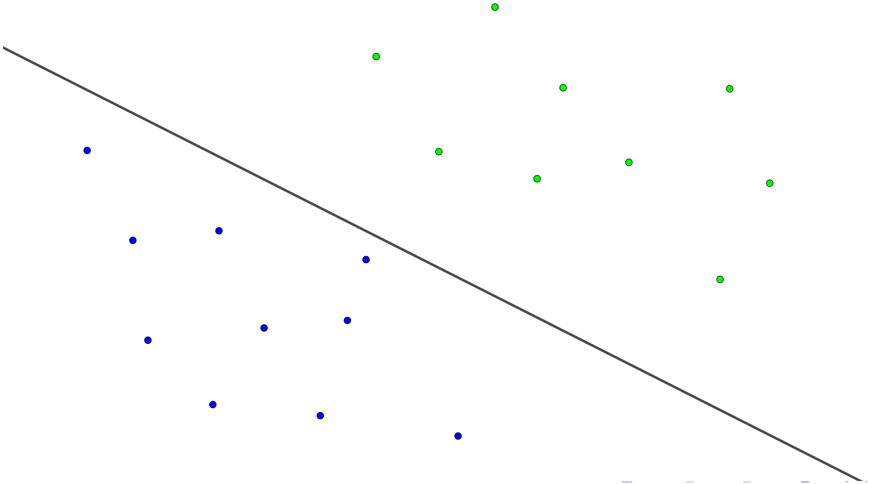
- One target variable with k classes

$$y_i \in \{y_{i1}, \ldots, y_{ik}\}, \; i = 1, \ldots, n$$

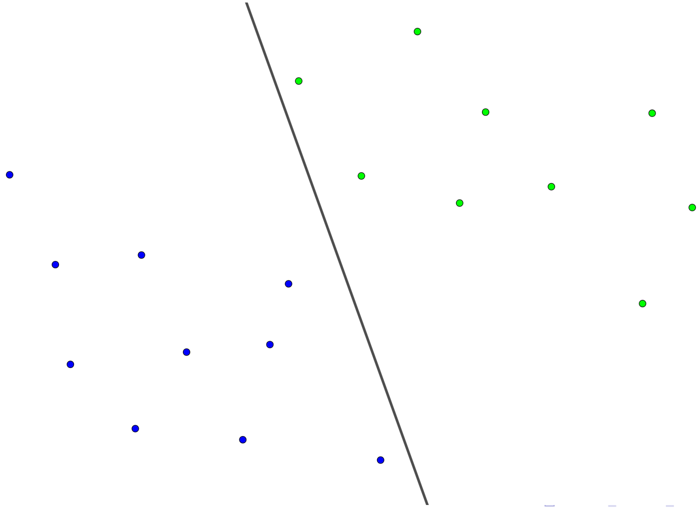
SVM



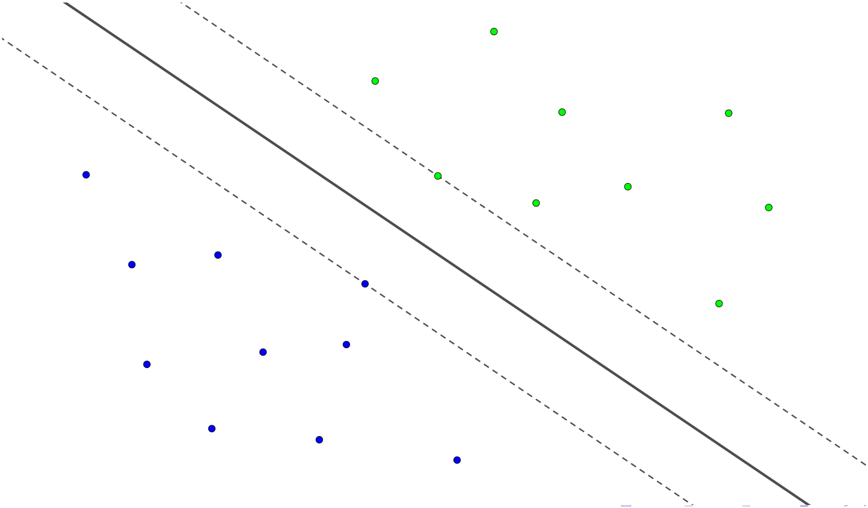
SVM



SVM



SVM



Basic SVM formulation

$$\min \frac{1}{2}\|\omega\|_2^2$$

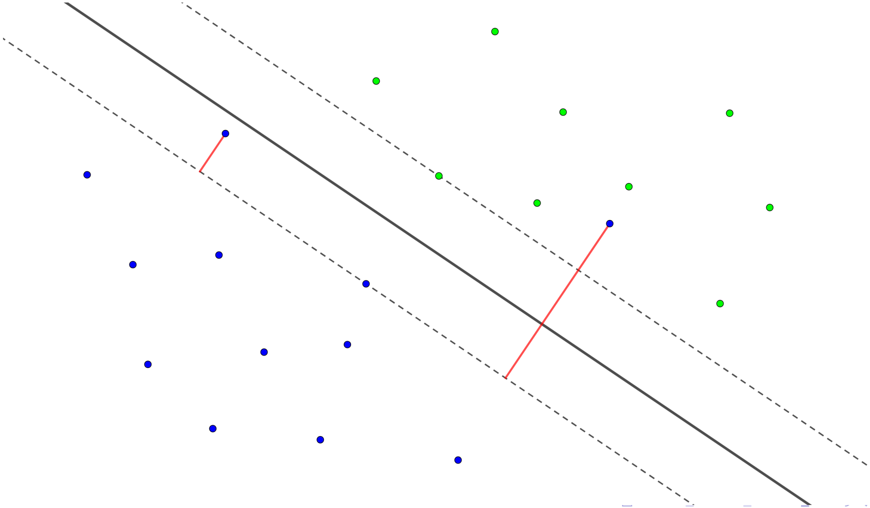
$$\begin{aligned} \text{s.t:} \quad & y_i(\omega' x_i + \omega_0) \geq 1 \qquad && \forall i = 1, \ldots, n \\ & \omega \in \mathbb{R}^p, \; \omega_0 \in \mathbb{R} \end{aligned}$$

General SVM formulation

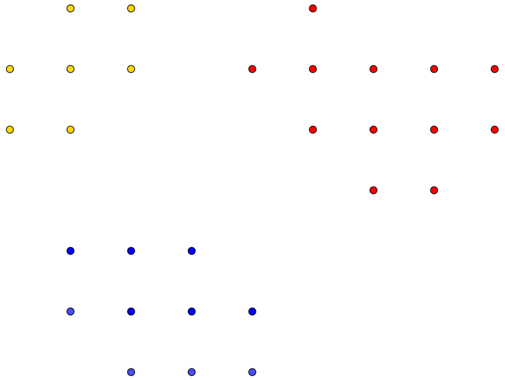
$$\min\left\{\frac{1}{2}\|\omega\|_2^2 + C\sum_{i=1}^{n} d_i\right\}$$

$$\begin{aligned}
\text{s.t:}\quad & y_i(\omega' x_i + \omega_0) \geq 1 - d_i && \forall i = 1,\dots,n\\
& \omega \in \mathbb{R}^p,\ \ \omega_0 \in \mathbb{R} && \\
& d_i \in \mathbb{R}^+ && \forall i = 1,\dots,n
\end{aligned}$$

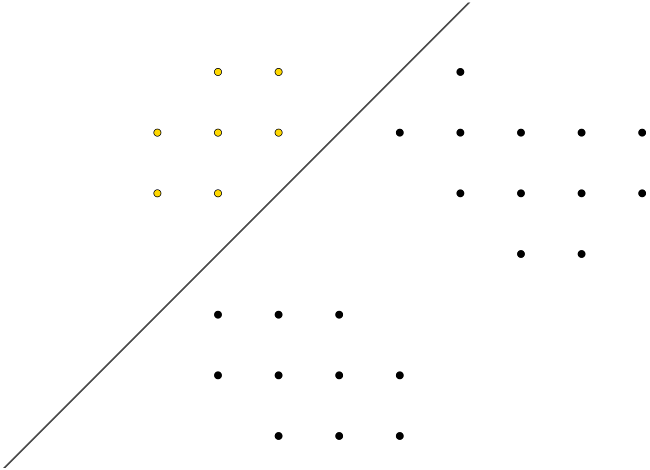
SVM



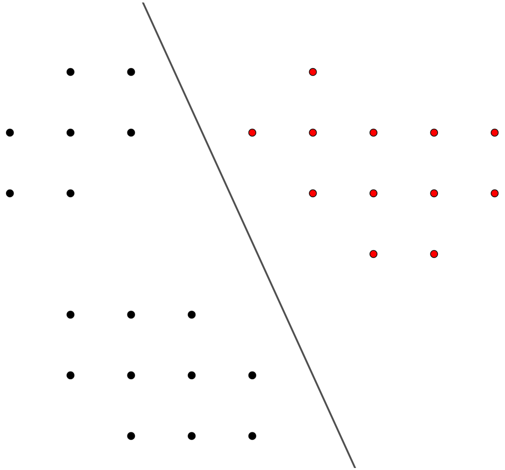
Multiclass problem



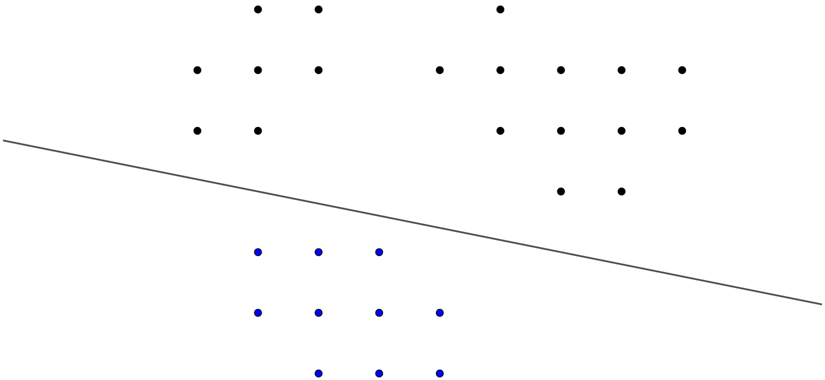
One Vs All



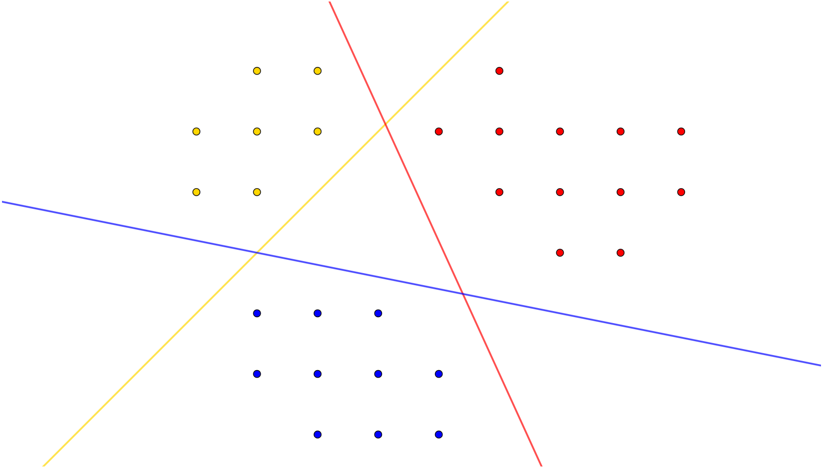
One Vs All



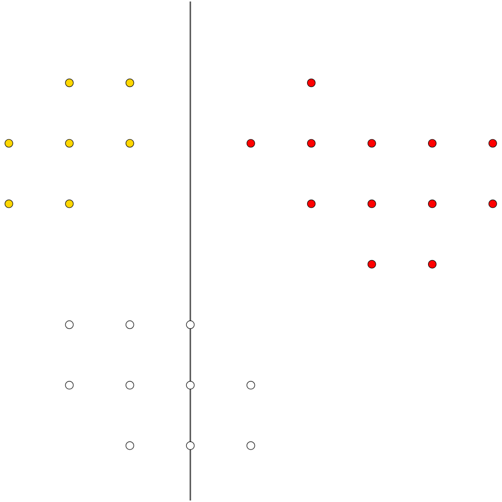
One Vs All



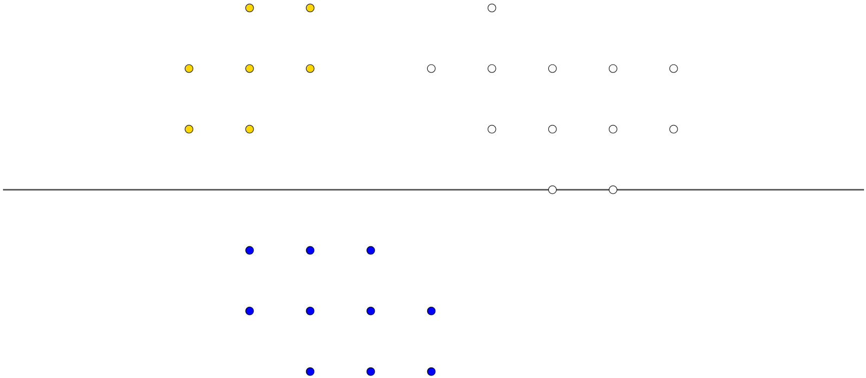
One Vs All



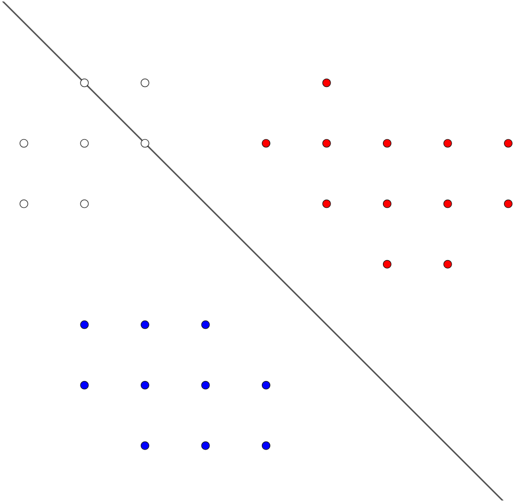
One Vs One



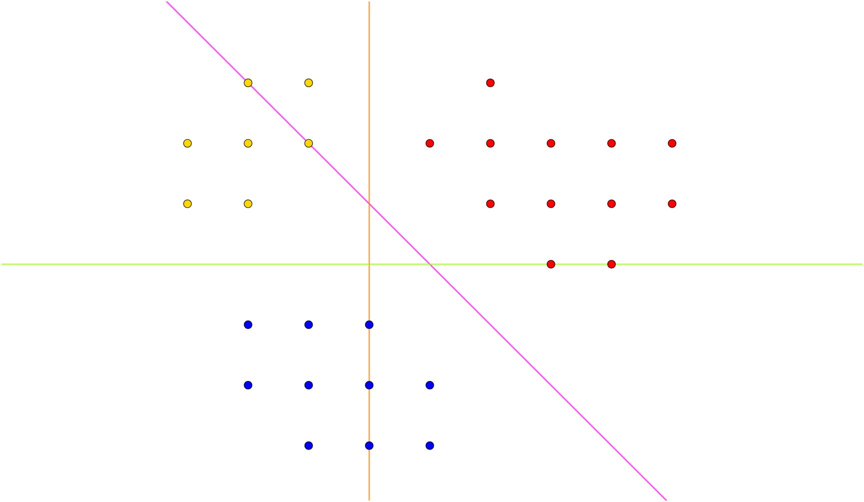
One Vs One



One Vs One



One Vs One



1. Support Vector Machine (SVM)

2. Multiclass methods for SVM
   - Sequential methods
   - Global methods

3. MCSVM
   - The model
   - Results

Weston-Watkins

$$\min \left\{ \sum_{r=1}^{k} \frac{\omega_r' \omega_r}{2} + C \sum_{i=1}^{n} \sum_{j \neq y_i} \xi_i^j \right\}$$

$$\begin{aligned}
\text{s.t:} \quad & \omega_{y_i}' x_i + \omega_{y_i 0} \geq \omega_j' x_i + \omega_{j0} + 2 - \xi_i^j && \forall i = 1, \ldots, n,\ j \in \{1, \ldots, k\} \setminus y_i \\
& \xi_i^j \geq 0 && \forall i = 1, \ldots, n,\ j \in \{1, \ldots, k\} \setminus y_i \\
& \omega_r \in \mathbb{R}^p, \omega_{r0} \in \mathbb{R} && \forall r = 1, \ldots, k
\end{aligned}$$

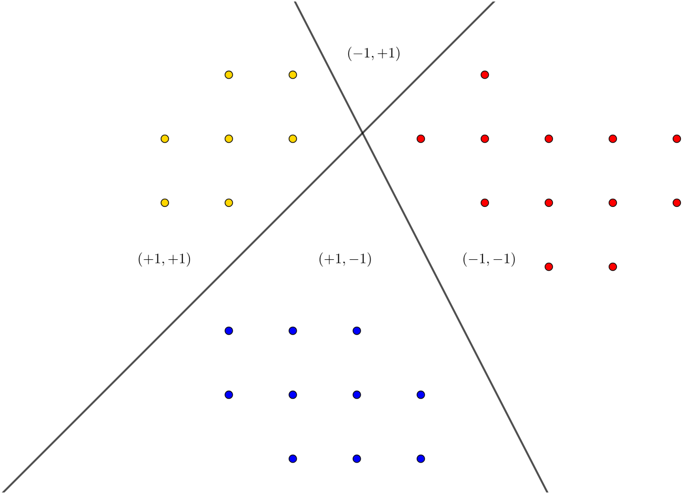
Crammer-Singer

$$\min \left\{ \sum_{r=1}^{k} \frac{\omega_r' \omega_r}{2} + C \sum_{i=1}^{n} \xi_i \right\}$$

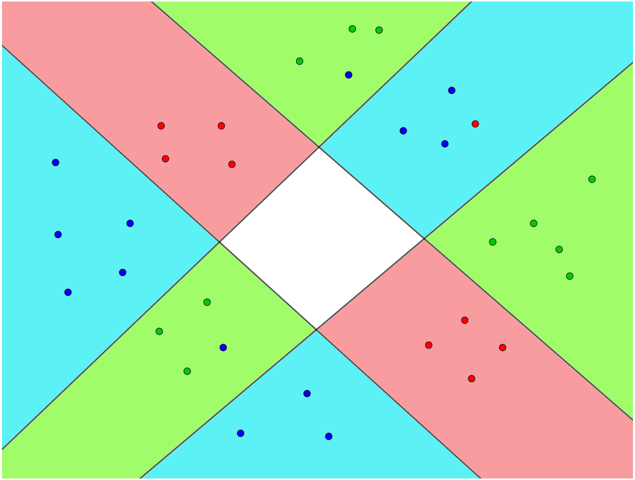
$$\begin{aligned}
\text{s.t:}\quad & \omega_{y_i}' x_i + \delta_{y_i j} - \omega_j' x_i \geq 1 - \xi_i^j && \forall i = 1, \dots, n,\ j \in \{1, \dots, k\} \\
& \xi_i \geq 0 && \forall i = 1, \dots, n \\
& \omega_r \in \mathbb{R}^p, \omega_{r0} \in \mathbb{R} && \forall r = 1, \dots, k \\
& \delta_{y_i j} \in \{0, 1\} && \forall y_i, j \in \{1, \dots, k\}
\end{aligned}$$

1. Support Vector Machine (SVM)
2. Multiclass methods for SVM
   - Sequential methods
   - Global methods
3. MCSVM
   - The model
   - Results

MCSVM



Class assignment through cells



Separation between classes

$$\max \min \left\{ \frac{2}{\|\omega_1\|_2}, \ldots, \frac{2}{\|\omega_m\|_2} \right\}$$

$$\min \max \left\{ \frac{1}{2}\|\omega_1\|_2^2, \ldots, \frac{1}{2}\|\omega_m\|_2^2 \right\}$$

Separation between classes

$$\max \min \left\{ \frac{2}{\|\omega_1\|_2}, \ldots, \frac{2}{\|\omega_m\|_2} \right\}$$

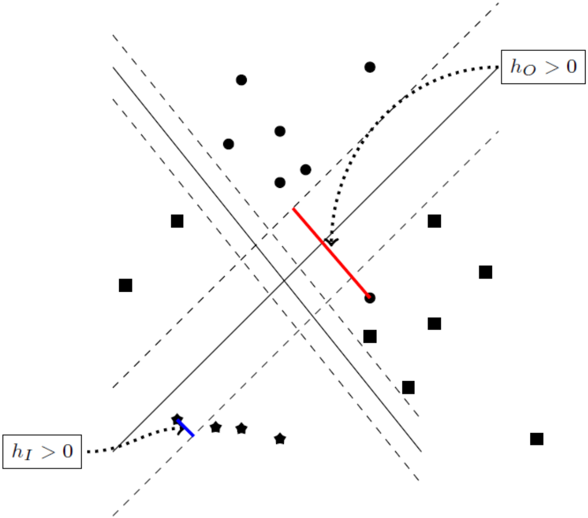
$$\min \max \left\{ \frac{1}{2}\|\omega_1\|_2^2, \ldots, \frac{1}{2}\|\omega_m\|_2^2 \right\}$$

Error functions

$$h_I\left(x,y,\mathscr{H}\right)=\begin{cases}\max\{0,\min\{1,1-\mathrm{s}_r(x)(\omega_r^t x+\omega_{r0})\}\} & \text{if } x \text{ is well classified with respect to } \mathscr{H}\\ 0 & \text{otherwise}\end{cases}$$

$$h_O\left(x,y,\mathscr{H}\right)=\begin{cases}1-\mathrm{s}(x)_r(\omega_r^t x+\omega_{r0}) & \text{if } x \text{ is wrong classified with respect to } \mathscr{H}\\ 0 & \text{otherwise}\end{cases}$$

Error functions



Continuous variables

$$\omega_r \in \mathbb{R}^p, \;\; \omega_{r0} \in \mathbb{R}, \qquad r = 1, \ldots m$$

$$e_{ir} \geq 0, \qquad i = 1, \ldots, n, \;\; r = 1, \ldots, m$$

$$d_{ir} \geq 0, \qquad i = 1, \ldots, n, \;\; r = 1, \ldots, m$$

Binary variables

- $t_{ir} = \begin{cases} 1 & \text{if } \omega_r^t x_i + \omega_{r0} \geq 0 \\ 0 & \text{otherwise} \end{cases}, \; i = 1, \ldots, n, \; r = 1, \ldots, m$

- $z_{is} = \begin{cases} 1 & \text{if } i \text{ is assigned to class } s \\ 0 & \text{otherwise} \end{cases}, \; i = 1, \ldots, n, \; s = 1, \ldots, k$

Binary variables

- $$\xi_i = \begin{cases} 0 & \text{if the class assigned to } i \\ & \text{coincides with } y_i \\ 1 & \text{otherwise} \end{cases}, \ i = 1, \ldots, n$$

- $$h_{ij} = \begin{cases} 1 & \text{if } x_j \text{ is well classified and} \\ & \text{is the representative of } x_i \\ 0 & \text{otherwise} \end{cases}, \ i,j = 1, \ldots, n \ (y_i = y_j)$$

MCSVM formulation

$$\min \quad \frac{1}{2}\|\omega_1\|_2^2 + C_1 \sum_{i=1}^{n}\sum_{r=1}^{m} e_{ir} + C_2 \sum_{i=1}^{n}\sum_{r=1}^{m} d_{ir}$$

$$\begin{aligned}
\text{s.a:} \quad & \frac{1}{2}\|\omega_1\|_2^2 \geq \frac{1}{2}\|\omega_r\|_2^2 && \forall r = 2,\ldots,m && (1)\\
& \omega_{ir}^t x_i + w_{r_0} \geq -T(1 - t_{ir}) && \forall i \in N, r \in M && (2)\\
& \omega_{ir}^t x_i + w_{r_0} \leq T t_{ir} && \forall i \in N, r \in M && (3)\\
& \sum_{s=1}^{k} z_{is} = 1 && \forall i \in N && (4)
\end{aligned}$$

MCSVM formulation

$$\|z_i - z_j\|_1 \leq 2\|t_i - t_j\|_1 \qquad \forall i,j \in N \qquad (5)$$

$$\xi_i = \|z_i - \delta_i\|_1 \qquad \forall i \in N \qquad (6)$$

$$\sum_{\substack{j \in N: \\ y_i = y_j}} h_{ij} = 1 \qquad \forall i \in N \qquad (7)$$

$$\xi_j + h_{ij} \leq 1 \qquad \forall i,j \in N(y_i = y_j) \qquad (8)$$

$$h_{ii} = 1 - \xi_i \qquad \forall i \in N \qquad (9)$$

MCSVM formulation

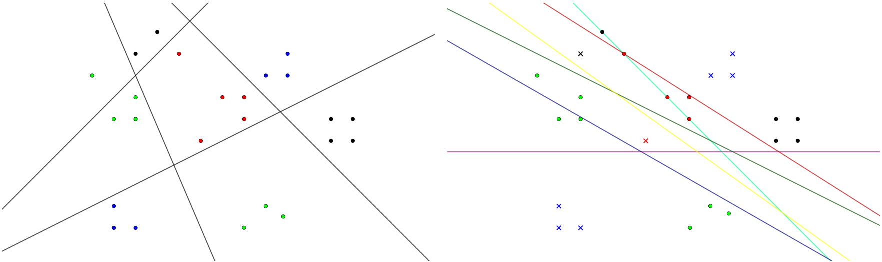
$$\omega_r^t x_i + \omega_{r0} \geq 1 - e_{ir} - T\left(3 - t_{ir} - t_{jr} - h_{ij}\right) \quad \forall i,j \in N, r \in M \tag{10}$$

$$\omega_r^t x_i + \omega_{r0} \leq -1 + e_{ir} + T\left(1 + t_{ir} + t_{jr} - h_{ij}\right) \quad \forall i,j \in N, r \in M \tag{11}$$

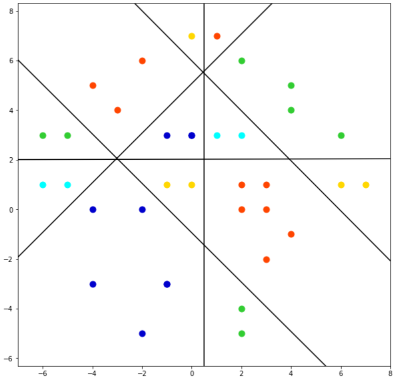
$$d_{ir} \geq 1 - \omega_r^t x_i - \omega_{r0} - T\left(2 + t_{ir} - t_{jr} - h_{ij}\right) \quad \forall i,j \in N, r \in M \tag{12}$$

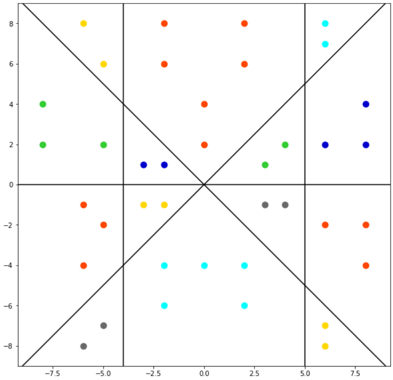
$$d_{ir} \geq 1 + \omega_r^t x_i + \omega_{r0} - T\left(2 - t_{ir} + t_{jr} - h_{ij}\right) \quad \forall i,j \in N, r \in M \tag{13}$$

MCSVM (left) and OVO (right)



Some MCSVM examples





Computational experiments

Description

Dataset	n_{Tr}	n_{Te}	p	k	m	m_{OVO}
Forest	75	448	28	4	3	6
Glass	75	139	10	6	6	15
Iris	75	75	4	3	2	3
Seeds	75	135	7	3	2	3
Wine	75	103	13	3	2	3
Zoo	75	26	17	7	4	21

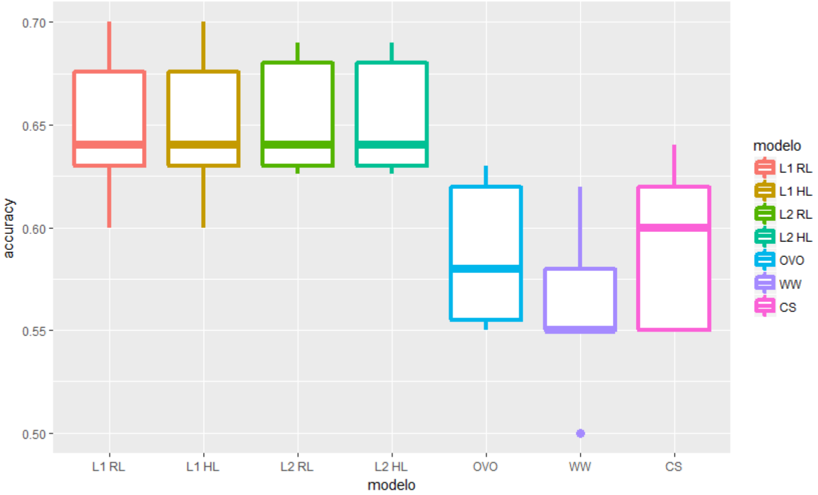
Computational experiments

Average accuracy results

Dataset	ℓ_1 RL	ℓ_1 HL	ℓ_2 RL	ℓ_2 HL	OVO	WW	CS
Forest	80.66	80.12	**82.30**	81.62	82.10	78.40	78.60
Glass	64.92	64.92	**65.32**	**65.32**	58.76	56.25	59.26
Iris	95.08	95.40	96.44	**96.66**	93.80	96.44	96.44
Seeds	**93.66**	**93.66**	93.52	93.52	91.02	93.52	93.52
Wine	95.20	95.20	**96.82**	**96.82**	96.34	96.09	96.17
Zoo	**89.75**	**89.75**	**89.75**	**89.75**	87.44	87.68	87.68

Computational experiments

Glass detailed experiment



References

1. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297 (1995)
2. Weston, J., Watkins, C.: Support vector machines for multi-class pattern recognition. In: European Symposium on Artificial Neural Networks, pp. 219–224 (1999)
3. Crammer, K., Singer, Y.: On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research **2**, 265–292 (2001)
4. Blanco, V., Japón, A. and Puerto, J. (2018). Optimal arrangements of hyperplanes for multiclass classification. Preprint available at https://arxiv.org/abs/1810.09167.

Conclusions

- We have developed a multiclass model that extends the idea of binary SVM.
- Our model obtains better or compared results on accuracy with respect to traditional SVM multiclass methods.
- We have proven that kernel tools are compatible with our model.