Feature selection via Mixed Integer Linear Programming

Martine Labbé   Antonio M. Rodríguez-Chía

Dept. Computer Science. Université Libre de Bruxelles
Dept. Statistics and Operations Research. Universidad de Cádiz

Contents:

Contents:

- Introduction to classification problems:
  - Classification problems via Mixed Integer Linear Programming.
  - Support Vector Machine.
  - Feature selection.
- A classification model with feature selection.
- Analysis of the model.

In today's world, six billion digital sources generate several zettabytes (10^{21} bytes) of data every year, and the number of sources is expected to grow to fifty billion by 2020. Computers, cell phones, graphic tablets, emails, social networks, remote sensors, and retail transactions all provide information that, properly extracted and analyzed, can reveal hidden patterns that can help to a classification of them. However, the technology required to advantageously process large data sets, often endowed with complex structures needs a deep analysis. When high-volume streams of data arrive, traditional methods for managing the information flow are inadequate and a further research to provide efficient classification schemes is needed. According to F. Jahanian, head of the Canadian National Science Foundation's, "Foundational research in data management and data analytics promises breakthrough discoveries and innovations across all disciplines." Many major software companies, including IBM, Microsoft, Oracle, SAP, SAS Institute, and Google, as well as a growing number of start-ups, are focusing on the challenges of Big Data computing.

Supervised classification problem (discriminant analysis)

Given:

- Ω: set of objects.
- $\mathcal{C}$: predefined set of classes.

The aim is to build a classification rule that predicts the class membership of an object, $u \in \Omega$ into one of the class of $\mathcal{C}$ by means of its predictor $x^u \in \mathbb{R}^n$.

Supervised classification problem (discriminant analysis)

Given:

- Ω: set of objects.
- $\mathcal{C}$: predefined set of classes.

The aim is to build a classification rule that predicts the class membership of an object, $u \in \Omega$ into one of the class of $\mathcal{C}$ by means of its predictor $x^u \in \mathbb{R}^n$.

These problems are studied from different perspectives:

- Artificial Intelligence
- Machine learning
- Statistical Pattern Recognition
- Mathematical Programming.

Supervised classification problem (discriminant analysis)

Given:

- Ω: set of objects.
- $\mathcal{C}$: predefined set of classes.

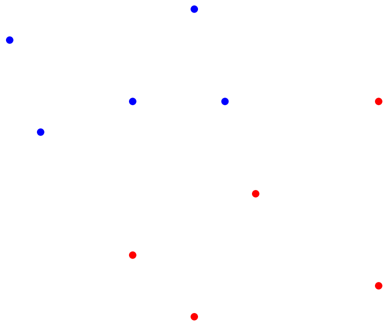
The aim is to build a classification rule that predicts the class membership of an object, $u \in \Omega$ into one of the class of $\mathcal{C}$ by means of its predictor $x^u \in \mathbb{R}^n$.

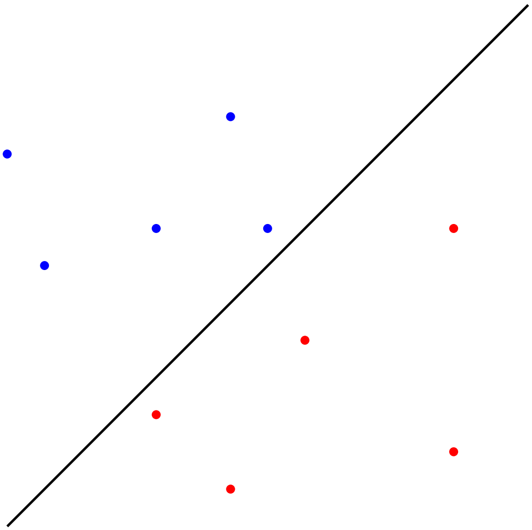
Applications:

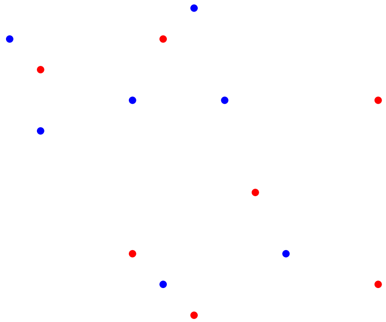
- Biology and medicine: Classification of gene expression data, homology detection, cancer diagnosis.
- Agriculture.
- Business: credit scoring, fraud detection, bankruptcy.

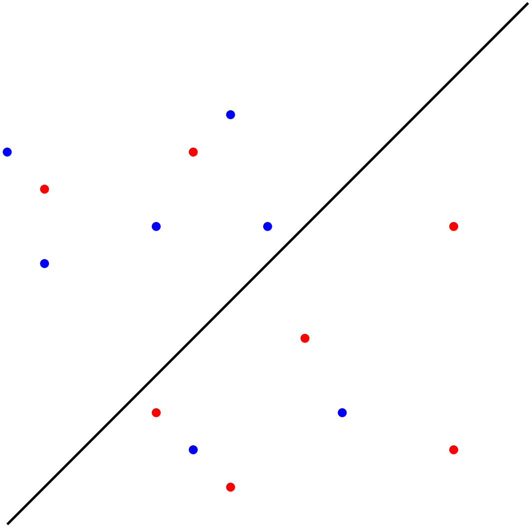
Notation:

- Entity or observation: is essentially a data point.
- Training set: Set of entities with known classification that is used to develop classification rules.
  Each entity is associated a pair (x, y), with $x \in \mathbb{R}^n$ (the predictor vector) and $y \in \mathcal{C}$ (the class).









- Cross validation: The training set is partitioned so that some entities are withheld during the model-development process. The withheld entities form a test set that is used to determine the validity of the model.
- *m*-fold cross-validation: the data with known classification is partitioned into m folds of approximately equal size.
- $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
- $$AUC = \frac{sensitivity + specificity}{2}$$

TP=True positive, TN= True negative

FP= False positive, FN= False negative

$sensitivity = \frac{TP}{TP+FN}$

$specificity = \frac{TN}{TN+FP}$

Classification via Mathematical Programming

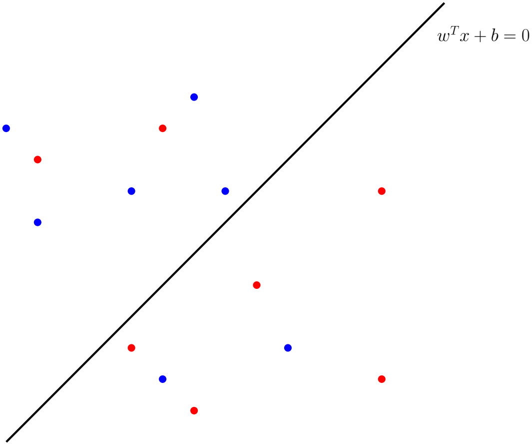
Mathematical Programming Approaches:

- It does not require assumption about the distribution.
- Mainly determine coefficients of linear discriminant functions or with support vector machines

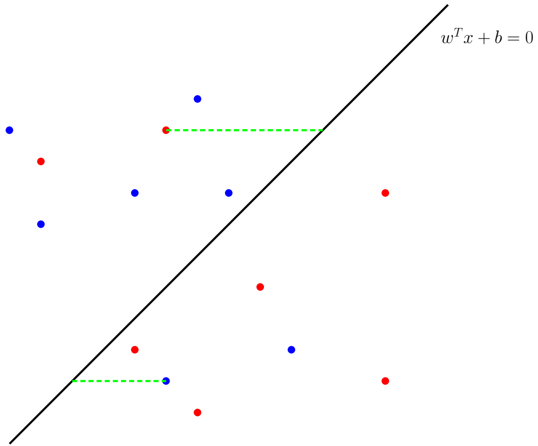
Linear programming classification models:

- Minimizing the sum of the distances to the separating hyperplane.
- Minimizing the maximum distance of an observation to the hyperplane.
- Maximizing measures of goodness of fit

Minimizing the sum of the distances to the separating hyperplane:



Minimizing the sum of the distances to the separating hyperplane:



Linear programming classification models

Minimizing the sum of the distances to the separating hyperplane:

$$
\begin{aligned}
MSD \quad \min \quad & \sum_i d_i \\
s.t. \quad & b + \sum_j x_{ij} w_j - d_i \leq 0, \quad \forall i \in G_1 \\
& b + \sum_j x_{ij} w_j + d_i \geq 0, \quad \forall i \in G_2 \\
& b, w_j, \quad \text{urs } \forall j \\
& d_i \geq 0 \forall i
\end{aligned}
$$

Linear programming classification models

Minimizing the maximum distance of an observation to the hyperplane:

$$
\begin{array}{lll}
MMD & \min & d \\
& s.t. & w_0 + \sum_j x_{ij} w_j - d \leq 0, \quad \forall i \in G_1 \\
& & w_0 + \sum_j x_{ij} w_j + d \geq 0, \quad \forall i \in G_2 \\
& & w_j, \quad \text{urs } \forall j \\
& & d_i \geq 0 \forall i
\end{array}
$$

Mixed-Integer programming classification models

Wherea LP offers a polynomial-time computational guarantee, MIP allows more flexible in (among other things) modeling misclassified observations and/or misclassification costs.

Mixed-Integer programming classification models

Wherea LP offers a polynomial-time computational guarantee, MIP allows more flexible in (among other things) modeling misclassified observations and/or misclassification costs.

MM: Minimizing the number of misclassifications:

$$
\begin{aligned}
MM \quad \min \quad & \sum_i z_i \\
s.t. \quad & w_0 + \sum_j x_{ij} w_j \leq M z_i, \quad \forall i \in G_1 \\
& w_0 + \sum_j x_{ij} w_j \geq -M z_i, \quad \forall i \in G_1 \\
& w_j \text{ urs } \forall j, \\
& z_i \in \{0,1\}, \ \forall i
\end{aligned}
$$

Nonlinear programming classification models

Minimizing the sum of the distances to the separating hyperplane:

$$
\begin{aligned}
MSD \quad \min \quad & (\sum_i d_i^p)^{\frac{1}{p}} \\
s.t. \quad & w_0 + \sum_j x_{ij} w_j - d_i \leq 0, \quad \forall i \in G_1 \\
& w_0 + \sum_j x_{ij} w_j + d_i \geq 0, \quad \forall i \in G_2 \\
& w_j, \quad \text{urs } \forall j \\
& d_i \geq 0 \forall i
\end{aligned}
$$

Feature Selection

Associated with discriminant analysis as a statistical tool are the tasks of determining the features that best discriminant analysis as a statistical tool.

$$x = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1n} \\ \vdots & \ddots & & & \vdots \\ x_{i1} & & x_{ij} & & x_{in} \\ \vdots & & & \ddots & \vdots \\ x_{m1} & \dots & x_{mj} & \dots & x_{mn} \end{pmatrix}$$

Usual case: $n >>> m$.

- Colorectal: 60×2000.
- Brain-Tumor1: 71×7131.
- Brain-Tumor2: 50×12627.
- Cancer I: 214×31099.
- Cancer II: 340×54675.

LP classification models+Feature selection

Minimizing the sum of the distances to the separating hyperplane

$$
\begin{aligned}
MSD \qquad & \min && \sum_i d_i \\
& s.t. && w_0 + \sum_j x_{ij} w_j - d_i \leq 0, \quad \forall i \in G_1 \\
& && w_0 + \sum_j x_{ij} w_j + d_i \geq 0, \quad \forall i \in G_2 \\
& && w_j, \quad \text{urs } \forall j \\
& && d_i \geq 0, \quad \forall i.
\end{aligned}
$$

LP classification models+Feature selection

Minimizing the sum of the distances to the separating hyperplane + Feature selection:

$$
\begin{aligned}
MSD + FS \quad \min \quad & (1-\lambda)\sum_i d_i + \lambda \sum_j |w_j|_* \\
s.t. \quad & w_0 + \sum_j x_{ij} w_j - d_i \leq 0, \quad \forall i \in G_1 \\
& w_0 + \sum_j x_{ij} w_j + d_i \geq 0, \quad \forall i \in G_2 \\
& w_j, \quad \text{urs } \forall j \\
& d_i \geq 0, \quad \forall i.
\end{aligned}
$$

where

$$
|w|_* = \begin{cases} 1, & \text{if } |w_j| > 0 \\ 0, & \text{otherwise.} \end{cases}
$$

LP classification models+Feature selection

$$
\begin{array}{lll}
MSD + FS & \min & (1-\lambda)\sum_i d_i + \lambda \sum_j (v_j)_* \\
& s.t. & w_0 + \sum_j x_{ij} w_j - d_i \leq 0, \quad \forall i \in G_1 \\
& & w_0 + \sum_j x_{ij} w_j + d_i \geq 0, \quad \forall i \in G_2 \\
& & -v_j \leq w_j \leq v_j, \quad \forall j \\
& & d_i \geq 0, \quad \forall i.
\end{array}
$$

where

$$
(v_j)_* = \begin{cases} 1, & \text{if } v_j > 0 \\ 0, & \text{otherwise.} \end{cases}
$$

Feature selection+LP classification models

Because of discontinuity of the step function term $\sum_j (v_j)_*$, it is approximated by a concave exponential on the nonnegative real line. (Mangasarian 1996)

$$(v_j)_* \sim 1 - \varepsilon^{-\alpha v_j}, \quad \alpha > 0.$$

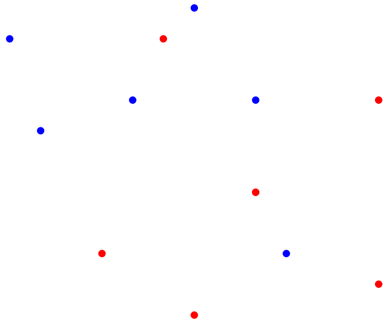
$$
\begin{aligned}
MSD \quad \min \quad & (1-\lambda)\sum_i d_i + \lambda \sum_j (1 - \varepsilon^{-\alpha v_j}) \\
s.t. \quad & w_0 + \sum_j x_{ij} w_j - d_i \leq 0, \quad \forall i \in G_1 \\
& w_0 + \sum_j x_{ij} w_j + d_i \geq 0, \quad \forall i \in G_2 \\
& -v_j \leq w_j \leq v_j, \quad \forall j \\
& d_i \geq 0, \quad \forall i.
\end{aligned}
$$

Minimization of a concave objective function over a polyhedral set.

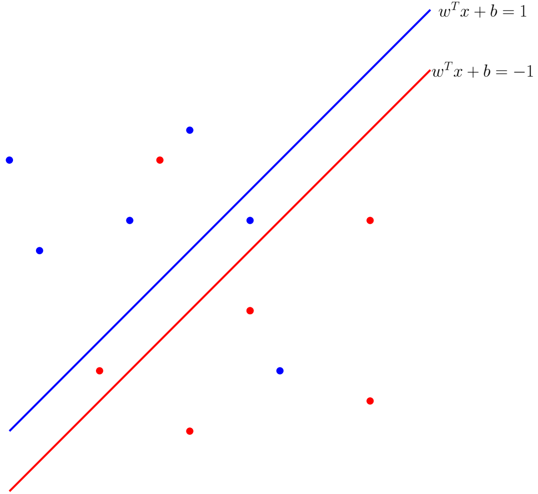
Support Vector Machine (Cortes and Vapnik (1995))

- $x_j \in \mathbb{R}^n$ training sample $j = 1, \dots, m$.
- $y_j \in \{-1, +1\}$ their labels.
- $f(x) = w^T x + b$ a hyperplane to optimally separate the training sample.
- The SVM formulatation balances:
  - Structural risk or Margin: minimization of $\|w\|$.
  - Empirical risk or deviations: misclassification errors ξ.

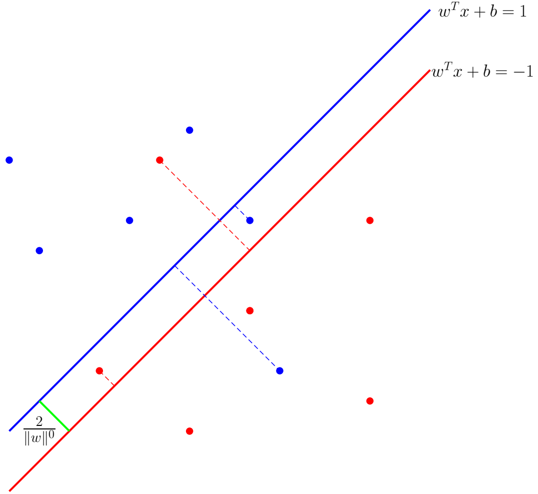
Support Vector Machine (Cortes and Vapnik (1995))



Support Vector Machine (Cortes and Vapnik (1995))



Support Vector Machine (Cortes and Vapnik (1995))



Support Vector Machine

$$\left.\begin{array}{ll} y_i = 1 & \Longrightarrow w^t x_i + b \geq 1 \\ y_i = -1 & \Longrightarrow w^t x_i + b \leq -1 \end{array}\right\} \Longrightarrow y_i(w^T x_i + b) \geq 1$$

ℓ_2 Support Vector Machine: Bradley and Mangasarian (1998)

$$\begin{array}{ll} \min\limits_{w,b,\xi} & \|w\|_2 + C \sum_{i=1}^{m} \xi_i \\ s.t. & y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \forall i = 1, \ldots, m, \\ & \xi_i \geq 0, \quad \forall i = 1, \ldots, m. \end{array}$$

Support Vector Machine

ℓ_1 Support Vector Machine (Bradley and Mangasarian (1998))

$$
\begin{aligned}
\min_{w,b,\xi} \quad & \|w\|_1 + C\sum_{i=1}^{m}\xi_i \\
s.t. \quad & y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \forall i = 1,\dots,m, \\
& \xi_i \geq 0, \quad \forall i = 1,\dots,m.
\end{aligned}
$$

Or equivalently,

$$
\begin{aligned}
\min_{w,b,\xi} \quad & \sum_{j=1}^{n} z_j + C\sum_{i=1}^{m}\xi_i \\
s.t. \quad & y_i\Big(\sum_{j=1}^{n} w_j x_{ij} + b\Big) \geq 1 - \xi_i, \quad \forall i = 1,\dots,m, \\
& -z_j \leq w_j \leq z_j, \quad \forall j = 1,\dots,n, \\
& \xi_i \geq 0, \quad \forall i = 1,\dots,m.
\end{aligned}
$$

Surveys: Classifiction & Mathematical Programming

- Bradley & Mangasarian. Feature selection via Mathematical Programming. IJOC, 1998.
- Carrizosa & Romero. Supervised classification and Mathematical Programming. COR. 2013.
- Lee & Wu. Classification and disease prediction via Mathematical Programming. Springer. 2009.

Support Vector Machine & Feature Selection

- Maldonado, Pérez, Labbé, Weber (2014)
- Feature selection via a budget constraint
- l_j/u_j correspond to a lower/upper bounds of the value of w_j, $\forall j$.

$$\begin{aligned}
\min_{w,b,\xi,v} \quad & \sum_{i=1}^{m} \xi_i \\
s.t. \quad & y_i\Big(\sum_{j=1}^{n} w_j x_{ij} + b\Big) \geq 1 - \xi_i, \quad \forall i = 1, \dots, m, \\
& l_j v_j \leq w_j \leq u_j v_j, \quad \forall j = 1, \dots, n, \\
& \sum_{j=1}^{n} v_j \leq B, \\
& v_j \in \{0, 1\}, \quad \forall j = 1, \dots, n, \\
& \xi_i \geq 0, \quad \forall i = 1, \dots, m.
\end{aligned}$$

Support Vector Machine & Feature Selection

- Maldonado, Pérez, Labbé, Weber (2014)
- Feature selection via a budget constraint
- l_j/u_j correspond to a lower/upper bounds of the value of w_j, $\forall j$.

$$
\begin{aligned}
\min_{w,b,\xi,v} \quad & \sum_{i=1}^{m} \xi_i \\
s.t. \quad & y_i\Big(\sum_{j=1}^{n} w_j x_{ij} + b\Big) \geq 1 - \xi_i, \quad \forall i = 1, \ldots, m, \\
& l_j v_j \leq w_j \leq u_j v_j, \quad \forall j = 1, \ldots, n, \\
& \sum_{j=1}^{n} v_j \leq B, \\
& v_j \in \{0, 1\}, \quad \forall j = 1, \ldots, n, \\
& \xi_i \geq 0, \quad \forall i = 1, \ldots, m.
\end{aligned}
$$

An important issue is the appropriate choice of these bounds.

The model

$$
\begin{aligned}
\min_{w,b,\xi,v,z} \quad & \sum_{i=1}^{m} \xi_i + C \sum_{j=1}^{n} z_j \\
s.t. \quad & y_i \Big(\sum_{j=1}^{n} w_j x_{ij} + b \Big) \geq 1 - \xi_i, \quad \forall i = 1, \ldots, m, \\
& l_j v_j \leq w_j \leq u_j v_j, \quad \forall j = 1, \ldots, n, \\
& -z_j \leq w_j \leq z_j, \quad \forall j = 1, \ldots, n, \\
& \sum_{j=1}^{n} v_j \leq B, \\
& v_j \in \{0, 1\}, \; w_j \text{ unrestricted}, \quad \forall j = 1, \ldots, n, \\
& \xi_i \geq 0, \quad \forall i = 1, \ldots, m.
\end{aligned}
$$

The model

$$
\begin{aligned}
\min_{w,b,\xi,v,z} \quad & \sum_{i=1}^{m} \xi_i + C \sum_{j=1}^{n} z_j \\
s.t. \quad & y_i \Big(\sum_{j=1}^{n} w_j x_{ij} + b \Big) \geq 1 - \xi_i, \quad \forall i = 1, \dots, m, \\
& l_j v_j \leq w_j \leq u_j v_j, \quad \forall j = 1, \dots, n, \\
& -z_j \leq w_j \leq z_j, \quad \forall j = 1, \dots, n, \\
& \sum_{j=1}^{n} v_j \leq B, \\
& v_j \in \{0, 1\}, \ w_j \text{ unrestricted}, \quad \forall j = 1, \dots, n, \\
& \xi_i \geq 0, \quad \forall i = 1, \dots, m.
\end{aligned}
$$

For simplicity: $l_j = -u_j \ \forall j = 1, \dots, n.$

Column Generation

$$\overbrace{y_i \sum_{j=1}^{n} x_{ij} w_j}^{A_1} + \overbrace{y_i b + \xi_i}^{A_2} \geq 1, \quad \forall i = 1, \ldots, m$$

$$\left.\begin{array}{ll} w_j \leq u_j v_j, & \forall j = 1, \ldots, n \\ -w_j \leq u_j v_j, & \forall j = 1, \ldots, n \\ w_j + z_j \geq 0, & \forall j = 1, \ldots, n \\ -w_j + z_j \geq 0, & \forall j = 1, \ldots, n \\ \sum_{j=1}^{n} v_j \leq B, & \\ v_j \in \{0, 1\}, & \end{array}\right\} X^1$$

$$\xi_i \geq 0, \forall i = 1, \ldots, m, \; b \text{ unrestricted} \quad \longrightarrow \quad X^0$$

Column Generation

$$\overbrace{y_i \sum_{j=1}^{n} x_{ij} w_j}^{A_1} + \overbrace{y_i b + \xi_i}^{A_2} \geq 1, \quad \forall i = 1, \ldots, m$$

$$\left.\begin{array}{ll} w_j \leq u_j v_j, & \forall j = 1, \ldots, n \\ -w_j \leq u_j v_j, & \forall j = 1, \ldots, n \\ w_j + z_j \geq 0, & \forall j = 1, \ldots, n \\ -w_j + z_j \geq 0, & \forall j = 1, \ldots, n \\ \sum_{j=1}^{n} v_j \leq B, & \\ v_j \in \{0, 1\}, & \end{array}\right\} X^1$$

$$\xi_i \geq 0, \forall i = 1, \ldots, m, \ b \text{ unrestricted} \quad \longrightarrow \quad X^0$$

Extreme points: $(v_j, w_j, z_j) = \{(0,0,0), (1, u_j, u_j), (1, -u_j, u_j)\}$

Column Generation

- $z_j = \sum_{t=1}^{T} \lambda^t z_j^t, \quad w_j = \sum_{t=1}^{T} \lambda^t w_j^t$

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} \xi_i + C \sum_{t=1}^{T} \sum_{j=1}^{n} \lambda^t z_j^t \\ s.t. \quad & y_i \sum_{t=1}^{T} \sum_{j=1}^{n} x_{ij} w_j^t \lambda^t + y_i b + \xi_i \geq 1, \quad i = 1, \dots, m \\ & \sum_{t=1}^{T} \lambda^t = 1, \\ & \lambda^t \geq 0, \quad t = 1, \dots, T \\ & \xi_i \geq 0, \quad \forall i = 1, \dots, m. \end{aligned}$$

Column Generation

Computation of the reduced cost:

$$
\begin{aligned}
\min \quad & C\sum_{j=1}^{n} z_j - \sum_{i=1}^{m}\alpha_i y_i \sum_{j=1}^{n} w_j x_{ij} - \beta \\
s.t. \quad & w_j \geq -u_j v_j, \quad \forall j = 1,\ldots,n \\
& w_j \leq u_j v_j, \quad \forall j = 1,\ldots,n \\
& w_j + z_j \geq 0, \quad \forall j = 1,\ldots,n \\
& -w_j + z_j \geq 0, \quad \forall j = 1,\ldots,n \\
& \sum_{j=1}^{n} v_j \leq B, \\
& v_j \in \{0,1\}, \quad \forall j = 1,\ldots,n.
\end{aligned}
$$

Objective function:

$$
\sum_{j=1}^{m} Cz_j - \left(\sum_{i=1}^{n}\alpha_i y_i x_{ij}\right) w_j \sim \sum_{j=1}^{n} C|w_j| - \left(\sum_{i=1}^{n}\alpha_i y_i x_{ij}\right) w_j.
$$

Column Generation

Derivative of objective function with respect to w_j is:

$$\begin{array}{ll} C - \sum_{i=1}^{n} \alpha_i y_i x_{ij} & \text{if } w_j \geq 0 \\ -C - \sum_{i=1}^{n} \alpha_i y_i x_{ij} & \text{if } w_j \leq 0 \end{array}$$

- If $C > \sum_{i=1}^{n} \alpha_i y_i x_{ij}$ and $w_j \geq 0$ then $w_j^* = 0$.
- If $C < \sum_{i=1}^{n} \alpha_i y_i x_{ij}$ and $w_j \geq 0$ then $w_j^* = u_j$.
- If $C \leq -\sum_{i=1}^{n} \alpha_i y_i x_{ij}$ and $w_j \leq 0$ then $w_j^* = -u_j$.
- If $C \geq -\sum_{i=1}^{n} \alpha_i y_i x_{ij}$ and $w_j \leq 0$ then $w_j^* = 0$.

$\forall j = 1, \ldots, n$. Consider:

$$\left(\min \left\{ 0, C|u_j| - \sum_{i=1}^{n} \alpha_i y_i x_{ij} u_j, C|u_j| + \sum_{i=1}^{n} \alpha_i y_i x_{ij} u_j \right\} \right)_{j=1}^{n} =$$

$$\left(\min \left\{ 0, C - \sum_{i=1}^{n} \alpha_i y_i x_{ij}, C + \sum_{i=1}^{n} \alpha_i y_i x_{ij} \right\} \right)_{j=1}^{n}.$$

Choose the B smallest values.

The model

Reformulation: Formulation II.

Considering $w_j = w_j^a - w_j^b$, where $w_j^a, w_j^b \geq 0$, $\forall j = 1, \ldots, n$.

$$
\begin{aligned}
\min_{w,b,\xi,v,z} \quad & \sum_{i=1}^{m} \xi_i + C \sum_{j=1}^{n} (w_j^a + w_j^b) \\
s.t. \quad & y_i \Big(\sum_{j=1}^{m} (w_j^a - w_j^b) x_{ij} + b \Big) \geq 1 - \xi_i, \quad \forall i = 1, \ldots, m, \\
& w_j^a + w_j^b \leq u_j v_j, \quad \forall j = 1, \ldots, n, \\
& \sum_{j=1}^{m} v_j \leq B, \\
& v_j \in \{0, 1\}, w_j^a, w_j^b \geq 0, \quad \forall j = 1, \ldots, n, \\
& \xi_i \geq 0, \quad \forall i = 1, \ldots, m.
\end{aligned}
$$

Comparative:

	Form.-1	Col. Generat.	Form.-2
Leukemia1=72x5329	43.92	9.63	1.66
Brain-Tumor1=60x7131	65.69	2.61	1.97
DLBCL=77x7131	78.98	13.02	2.16
Carcinoma=36x7459	47.72	12.97	0.98
Brain-Tumor2=50x12627	233.49	13.09	2.30
Prostate-Tumor=102x12606	618.61	36.89	6.66

Strategy for obtaining tightened values of u_j

Strategy I:

Step 1: Solve the LP problem.

$$\begin{aligned}
\min_{w,b,\xi,v,z} \quad & \sum_{i=1}^{m} \xi_i + C \sum_{j=1}^{n} (w_j^a + w_j^b) \\
s.t. \quad & y_i \Big(\sum_{j=1}^{m} (w_j^a - w_j^b) x_{ij} + b \Big) \geq 1 - \xi_i, \quad \forall i = 1, \dots, m \\
& w_j^a + w_j^b \leq u_j v_j, \quad \forall j = 1, \dots, n \\
& \sum_{j=1}^{m} v_j \leq B, \\
& 0 \leq v_j \leq 1, w_j^a, w_j^b \geq 0 \quad \forall j = 1, \dots, n \\
& \xi_i \geq 0, \quad \forall i = 1, \dots, m \\
& b \text{ unrestricted.}
\end{aligned}$$

Strategy for obtaining tightened values of u_j

Strategy I:

Step 1: Solve the LP problem.

Step 2: Let $\overline{w}_j^a$ and $\overline{w}_j^b$ be the optimal solution and set $J_1 = \{j : \overline{w}_j^a + \overline{w}_j^b > 0\}$. Solve the following $MILP(J_1)$:

$$
\begin{aligned}
\max_{w,b,\xi,v,z} \quad & \sum_{i=1}^{m} \xi_i + C \sum_{j \in J_1} (w_j^a + w_j^b) \\
s.t. \quad & y_i \Big(\sum_{j \in J_1} (w_j^a - w_j^b) x_{ij} + b \Big) \geq 1 - \xi_i, \quad \forall i = 1, \ldots, m \\
& w_j^a + w_j^b \leq u_j v_j, \quad \forall j \in J_1 \\
& \sum_{j \in J_1} v_j \leq B, \\
& v_j \in \{0,1\}, w_j^a, w_j^b \geq 0 \quad \forall j \in J_1, \\
& \xi_i \geq 0, \quad \forall i = 1, \ldots, m \\
& b \text{ unrestricted.}
\end{aligned}
$$

Strategy for obtaining tightened values of u_j

Strategy I:

Step 1: Solve the LR problem.

Step 2: Let $J_1 = \{j : \overline{w}_j^a + \overline{w}_j^b > 0\}$. Solve t $MILP(J_1)$:

Step 3: Let UB be the optimal value of the above problem.

$$\begin{aligned}
\max_{w,b,\xi,v} \quad & w_j^a + w_j^b \\
s.t. \quad & y_i\Big(\sum_{j=1}^{m}(w_j^a - w_j^b)x_{ij} + b\Big) \geq 1 - \xi_i, \ \forall i = 1, \ldots, m \\
& w_j^a + w_j^b \leq u_j v_j, \quad \forall j = 1, \ldots, n \\
& \sum_{j=1}^{m} v_j \leq B, \\
& \sum_{i=1}^{m} \xi_i + C \sum_{j=1}^{n}(w_j^a + w_j^b) \leq UB, \\
& 0 \leq v_j \leq 1, w_j^a, w_j^b \geq 0 \, \forall j, \ \xi_i \geq 0, \ \forall i, \ b \text{ unrestricted.}
\end{aligned}$$

Strategy for obtaining tightened values of u_j

Strategy I:

Step 1: Solve the LR problem.

Step 2: Let $\overline{w}_j^a$ and $\overline{w}_j^b$ be the optimal solution and set $J_1 = \{j \,:\, \overline{w}_j^a + \overline{w}_j^b > 0\}$. Solve the following $MILP(J_1)$:

Step 3: Let *obval* be the optimal value of the above problem.

Output: Let $\overline{w}_j^a + \overline{w}_j^b$ be the optimal value of the above problem for $j = 1, \ldots, n$. Update the values of u_j as follows: If $\overline{w}_j^a + \overline{w}_j^b < u_j$ then $u_j := \overline{w}_j^a + \overline{w}_j^b$ for $j = 1, \ldots, n$.

Strategy for obtaining tightened values of u_j

Strategy II:

Step 1: Solve the LP problem.

$$
\begin{aligned}
\min_{w,b,\xi,v,z} \quad & \sum_{i=1}^{m} \xi_i + C \sum_{j=1}^{n} (w_j^a + w_j^b) \\
s.t. \quad & y_i \Big(\sum_{j=1}^{m} (w_j^a - w_j^b) x_{ij} + b \Big) \geq 1 - \xi_i, \quad \forall i = 1, \ldots, m \\
& w_j^a + w_j^b \leq u_j v_j, \quad \forall j = 1, \ldots, n \\
& \sum_{j=1}^{m} v_j \leq B, \\
& 0 \leq v_j \leq 1, w_j^a, w_j^b \geq 0 \quad \forall j = 1, \ldots, n \\
& \xi_i \geq 0, \quad \forall i = 1, \ldots, m \\
& b \text{ unrestricted.}
\end{aligned}
$$

Strategy for obtaining tightened values of u_j

Strategy II:

Step 1: Consider linear relaxation.

Step 2: LB:the optimal value of this problem

$w_j^{a^*}$, $w_j^{b^*}$: the optimal solutions.

If $w_{j_0}^{a^*} = 0$ and $w_{j_0}^{b^*} = 0$:

i) The optimum value of the problem with $w_{j_0}^a = \overline{w}_{j_0}^a$ is:

$$LB + \overline{w}_{j_0}^a \Big(C - \sum_{i=1}^m \alpha_i y_i x_{ij_0} \Big)$$

$$\text{whenever } \sum_{j=1}^m v_j^* + \frac{\overline{w}_{j_0}^a}{u_{j_0}} \leq B.$$

Step 3: Let UB be an upper bound of the objective function:

i) $LB + \overline{w}_{j_0}^a (C - \sum_{i=1}^n \alpha_i y_i x_{ij_0}) \leq UB$ then

$$\overline{w}_{j_0}^a \leq \frac{UB - LB}{C - \sum_{i=1}^m \alpha_i y_i x_{ij_0}}.$$

Strategy for obtaining tightened values of u_j

Strategy I: $\lvert w \rvert = w^a + w^b$	Strategy II: w^a, w^b
u(3)=0,537901	ua(3):=0,947904; ub(3):=0,280248
u(4)=0,455854	ua(4):=0,517722; ub(4):=0,371513
u(5)=0,556773	ua(5):=0,345003; ub(5):=0,579808
u(6)=0,441695	ua(6):=0,413056; ub(6):=0,45408
u(7)=0,436505	ua(7):=0,437826; ub(7):=0,427492
u(8)=0,456681	ua(8):=0,457311; ub(8):=0,410419
u(9)=0,43807	ua(9):=0,435321; ub(9):=0,429909
u(10)=0,470935	ua(10):=0,473788; ub(10):=0,397997
u(11)=0,45484	ua(11):=0,355108; ub(11):=0,553345
u(12)=0,436303	ua(12):=0,427289; ub(12):=0,43804
u(13)=0,462439	ua(13):=0,376622; ub(13):=0,508116
u(14)=0,610984	ua(14):=0,813715; ub(14):=0,294612
u(15)=0,587497	ua(15):=0,770599; ub(15):=0,300703
u(16)=0,628137	ua(16):=0,841183; ub(16):=0,291169

Flow cover inequalities

$$X = \{(x,y) \in \mathbb{R}^n_+ \times B^n \ : \ \sum_{j\in N_1} x_j - \sum_{j \in N_2} x_j \le b,\ x_j \le a_j y_j \text{ for } j \in N_1 \cup N_2\}.$$

Let $C = C_1 \cup C_2$ with $C_1 \subseteq N_1$, $C_2 \subseteq N_2$:

$$\sum_{j\in C_1} a_j - \sum_{j\in C_2} a_j = b + \lambda, \quad \lambda > 0.$$

$$\sum_{j\in C_1} x_j + \sum_{j\in C_1} (a_j - \lambda)^+ (1 - y_j) \le b + \sum_{j\in C_2} a_j + \lambda \sum_{j \in L_2} y_j + \sum_{j\in N_2 \setminus (C_2 \cup L_2)} x_j.$$

A separation heuristic for the FCI by solving the following knapsack pb:

$$\begin{aligned} \min \quad & \sum_{j\in N_1} (1 - y_j^*) z_j - \sum_{j \in N_2} y_j^* z_j \\ s.t. \quad & \sum_{j\in N_1} a_j z_j - \sum_{j\in N_2} a_j z_j > b, \\ & z \in B^n. \end{aligned}$$

Work in Progress...

- Is our problem NP-hard?
- Solving MILP:
  - Branch and Price Algorithm.
  - Benders decomposition.
  - Alternative approaches.
- Application to real data and comparison with typical classification methods.