VODKA-PLSR, a family of PLS models based on the NIPALS algorithm

Jean Claude Boulet, Dominique Bertrand, Gerard Mazerolles, Jean-Michel Roger, Robert Sabatier

To cite this version:

Jean Claude Boulet, Dominique Bertrand, Gerard Mazerolles, Jean-Michel Roger, Robert Sabatier. VODKA-PLSR, a family of PLS models based on the NIPALS algorithm. CAC2010, Oct 2010, Anvers, Belgium. pp.1, 2010. hal-02750739


**Theory**

Including expert information into regression models using NIPALS

A re-writing of NIPALS puts forwards a new parameter, a vector r chosen by the operator. This vector allows the extraction of useful information from X

(1) NIPALS
Known properties:
\[ \mathbf{T} = \mathbf{X} \mathbf{W} (\mathbf{P'} \mathbf{W}^{-1}) \]
\[ \mathbf{b} = \mathbf{W} (\mathbf{P'} \mathbf{W}^{-1}) \mathbf{c} \]

(2) a new writing of NIPALS
New properties:
\[ \hat{\mathbf{y}} = (\mathbf{T} \mathbf{T}^{-1}) \mathbf{T} \mathbf{y} \]
\[ \mathbf{b} = \Sigma (\mathbf{P'} \Sigma \mathbf{P})^{-1} (\mathbf{T} \mathbf{T}^{-1}) \mathbf{T} \mathbf{y} \]
\[ \mathbf{b} = \Sigma (\mathbf{P'} \Sigma \mathbf{P})^{-1} \mathbf{P'} \Sigma \mathbf{x} \mathbf{y} \]
(simplified)

(3) Vector Orientation Decided through Knowledge Assessment: VODKA-PLSR

(1) \( r = x'y \) \( \Rightarrow \) NIPALS (postulate)

(2) \( r = \text{any vector of dimension } P \)

**Application**

Ethanol quantification in wines and musts

<table>
<thead>
<tr>
<th>Model</th>
<th>r choice</th>
<th>Notes</th>
<th>LV5</th>
<th>LV6</th>
<th>LV7</th>
<th>LV8</th>
<th>LV9</th>
<th>LV10</th>
<th>LV11</th>
<th>LV12</th>
<th>LV13</th>
<th>LV14</th>
</tr>
</thead>
<tbody>
<tr>
<td>m₁</td>
<td>₁pₚ</td>
<td></td>
<td>2.30</td>
<td>2.94</td>
<td>1.43</td>
<td>1.12</td>
<td>1.09</td>
<td>1.08</td>
<td>0.99</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>m₂</td>
<td>X₁₉</td>
<td>Mean of X spectra</td>
<td>2.22</td>
<td>2.50</td>
<td>2.23</td>
<td>1.46</td>
<td>0.94</td>
<td>0.93</td>
<td>1.02</td>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>m₃</td>
<td>X₉</td>
<td>NIPALS</td>
<td>1.24</td>
<td>1.04</td>
<td>1.03</td>
<td>1.34</td>
<td>1.02</td>
<td>1.38</td>
<td>1.19</td>
<td>1.08</td>
<td>1.19</td>
<td>1.18</td>
</tr>
<tr>
<td>m₄</td>
<td>k</td>
<td>Pure spectra</td>
<td>1.93</td>
<td>2.42</td>
<td>1.88</td>
<td>1.21</td>
<td>1.02</td>
<td>1.01</td>
<td>1.02</td>
<td>1.03</td>
<td>1.03</td>
<td>1.02</td>
</tr>
<tr>
<td>m₅</td>
<td>NAS</td>
<td>Net analyte signal</td>
<td>0.94</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>0.97</td>
<td>0.99</td>
<td>1.02</td>
<td>1.04</td>
<td>1.04</td>
<td>1.01</td>
</tr>
<tr>
<td>m₆</td>
<td>X₉, y₉</td>
<td>NIPALS centered</td>
<td>1.05</td>
<td>1.00</td>
<td>0.95</td>
<td>1.25</td>
<td>1.02</td>
<td>1.40</td>
<td>1.20</td>
<td>1.11</td>
<td>1.23</td>
<td>1.22</td>
</tr>
</tbody>
</table>

**Discussion and conclusion**

**Practical aspects**

- An infinity of regression models based on NIPALS
- Expert information (e.g. NAS) can be directly introduced into regression models through \( r \)
- NIPALS (\( r = x'y \)) isn't always the best choice

**Theoretical aspects**

- A more general model depending on the choices of \( P \) and \( \Sigma \)

**VODKA-PLSR synopsis**

Tool: an orthogonal projection

Tuning: a matrix

X

Experience \( (X,y) \)

Knowledge \( r \)

Useful information: a subspace basis

CAC 2010 - ANTWERP