Consider the following dataset consisting of twelve training instances (ids 1–12, on the left) and seven test instances (ids 13–19, on the right), with four variables \( v_1 \)–\( v_4 \) and a class label \( y \) that can take one of three values.

<table>
<thead>
<tr>
<th>id</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>-2</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>5</td>
<td>-6</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>3</td>
<td>-6</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>●</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>●</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>●</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>●</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>id</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>7</td>
<td>4</td>
<td>-6</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>5</td>
<td>-2</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>●</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>3</td>
<td>-1</td>
<td>0</td>
<td>●</td>
</tr>
<tr>
<td>18</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>●</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>●</td>
</tr>
</tbody>
</table>

Problem 1 (Support Vector Machines). We trained a linear SVM for the binary classification of separating red instances from blue and yellow ones, after mapping the dataset to two dimensions as follows, where the first dimension is the sum of attributes \( v_1 + v_3 \), while attribute \( v_4 \) stands as the second dimension.

For the training dataset above, after mapping the variables and considering two classes (red vs. combined blue or yellow), we obtained the following vector of Lagrange multipliers:

\[
\mathbf{a} = (0, 57, 57, 0, 0, 0, 130, 0, 16, 0, 0)
\]

a) Compute the weights and bias defining the separating hyperplane (\( \mathbf{w} \) and \( b \)).

b) Compute the predicted labels for the test dataset.

c) Can you find a non-linear transformation that separates the red instances from the rest in one dimension?