Method: Linear Programming SVM (Liknon) feature selection combined with state of the art classification rules.


Code/engine:


Implementation strategy: The dataset first is divided into 10 parts (10 fold crossvalidation). The training set of the single fold is further subdivided into balanced training and unbalanced (the rest samples) monitoring sets. A number of the subdivisions is arbitrary, we did 31. In every subdivision a number of LP SVM models/discriminants is trained on the balanced training set and the BER - balanced error rate - of each is estimated on the monitoring set. The number of the models depends on the set of the chosen values of regularization parameter C. The data, number of features and available computational time determine the range of C values. The model with smallest monitoring BER is selected in every subdivision. As a result, in the single fold, we have an ensemble of the linear discriminants and a feature profile possibly to be investigated with the other classification rules.

Preprocessing: none.

Feature selection: Feature selection relies on the property of LP SVM to produce sparse solutions. The identities of the features, corresponding to non zero weights of the discriminants, are included in the profile. Every fold reveals slightly different feature profiles. However relevant feature identities consistently appear in many subdivisions. First it was noticed in the experiments with synthetic data and then in the experiments with microarrays.

Classification:

1) ensemble of linear discriminants;
2) the rules tested on the derived feature profile included linear and nonlinear, all available in the PRTools: fisher classifier, linear logistic classifier, subspace classifier, nearest neighbors, decision tree, nearest mean classifier, quadratic classifier.
3) NO TRANSDUCTION.

Results:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Entry name</th>
<th>Entry ID</th>
<th>Test BER</th>
<th>Test AUC</th>
<th>Score</th>
<th>Track</th>
</tr>
</thead>
</table>

Table 1: Our methods best results
Table 2: Winning entries of the AlvsPK challenge

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Entrant name</th>
<th>Entry name</th>
<th>Entry ID</th>
<th>Test BER</th>
<th>Test AUC</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADA</td>
<td>Roman Lutz</td>
<td>LogitBoost with trees</td>
<td>13, 18</td>
<td>0.166</td>
<td>0.9168</td>
<td>0.002</td>
</tr>
<tr>
<td>GINA</td>
<td>Roman Lutz</td>
<td>LogitBoost/Doubleboost</td>
<td>892, 893</td>
<td>0.0339</td>
<td>0.9668</td>
<td>0.2308</td>
</tr>
<tr>
<td>HIVA</td>
<td>Vojtech Franc</td>
<td>RBF SVM</td>
<td>734, 933, 934</td>
<td>0.2827</td>
<td>0.7707</td>
<td>0.0763</td>
</tr>
<tr>
<td>NOVA</td>
<td>Mehreen Saeed</td>
<td>Submit E final</td>
<td>1038</td>
<td>0.0456</td>
<td>0.9552</td>
<td>0.0385</td>
</tr>
<tr>
<td>SYLVA</td>
<td>Roman Lutz</td>
<td>LogitBoost with trees</td>
<td>892</td>
<td>0.0062</td>
<td>0.9938</td>
<td>0.0302</td>
</tr>
<tr>
<td>Overall</td>
<td>Roman Lutz</td>
<td>LogitBoost with trees</td>
<td>892</td>
<td>0.1117</td>
<td>0.8892</td>
<td>0.1431</td>
</tr>
</tbody>
</table>

Quantitative advantages: Simplicity and interpretability of the results in terms of feature identities, important for discrimination. The stability of the discovered feature identities in different folds suggests that the feature selection via LP SVM is robust to the sample bias. In case of high dimensional data, the discovered features provide a reduced representation of the data for testing other classifiers. A success of the suggested approach was apparent for GINA dataset. DISADVANTAGE- high computational burden.

Qualitative advantages: The sequence of values of the regularization parameter determines the increasing number of features given by the increasing number of non-zero weights of the linear discriminant. This can be considered as a feature selection structure. The sequence of the LP SVM solutions- linear discriminants of increasing complexity- forms a nested structure, where the principles of the structural risk minimization may apply.

Keywords:
- Preprocessing or feature construction: none.
- Feature selection approach: embedded feature selection.
- Feature selection engine: SVM.
- Feature selection search: feature ranking, ordered FS (ordered feature selection)
  Feature selection criterion: monitoring error
- Classifier: nearest neighbors, tree classifier, L1 norm regularization, ensemble
  method, bagging.
- Other: sample bias.