Many time series classification problems involve irregularly sampled data wherein observed variables across data cases do not align in time or number due to inherent unpredictability in the sampling process. These time series are difficult to handle because few classical learning algorithms are robust to variable-length features. We motivate a direct approach to perform classification on irregularly sampled time series using neural networks without preprocessing our data set using feature extraction methods. Our results indicate promising performance at classification of noisy and irregularly sampled time series.

We evaluate our model using an ECG classification data set that is fully observed [Olszewski, 2001]. Each time series is an electrocardiogram (ECG) waveform for an individual patient. Data cases are labeled abnormal and normal. In our experimental protocol, we subsample the observed points to form artificial irregularly sampled data for performing controlled experiments.

We test our model on the UC Riverside ECG data set using a three-layer neural network (L=2). We subsample time series of length M=25, 50 percent of the original data set to demonstrate the model’s capacity to learn from sparse datasets. We vary the number of RBF centers in the first (RBN) hidden layer, but fix 2 hidden neurons in each hidden layer. Learning is performed using the LBFGS algorithm in Scipy (Python).

Fig 1: Subsampling of 100 examples from full ECG data set. Positive cases are indicated in red, negative in blue.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>79.6</td>
<td>67.2</td>
<td>92.4</td>
<td>76.4</td>
<td>60.4</td>
<td>89.8</td>
<td>75.2</td>
<td>77.8</td>
</tr>
</tbody>
</table>

We record training and test set classification accuracy across each trial run for each subsample (M = 25, 50) respectively, also varying B.

We plot the learned RBF function for M=50, K1 = 2, 10, fixing B=10:

Additionally, we can visualize the input space for the logistic output unit over the second layer of hidden units (M=50, K1 = 2, 10, training set):

Acknowledgments: This work is partially supported by the Research Experiences for Undergraduates (REU) program of the National Science Foundation (NSF) under award number CNS-1062749. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the NSF.

Works cited: