Human Activity Recognition using Wearable Devices Sensor Data

Shutong Zhang, Jason Wu, Chenying Zhang

Problem Formulation

Our problem includes two parts:

1. Extract useful features

\[ F = \{ F_1, F_2, ..., F_m \} \]

from Raw sensor data

\[ R = \{ R_1, R_2, ..., R_m \} \]

for training and test examples

\[ X = \{ x_1, x_2, ..., x_m \} \]

2. Use the features space \( F \) to predict the labels

\[ y_{predict} = \{ y_1', y_2', ..., y_m' \} \]

for test data \( X_{test} \).

Feature Selection

Principal components analysis (PCA) is performed on the train and test matrix \( X \) to make feature selections. 561 features are projected to a 30-dimension space and a 50-dimension space, with total percentage of eigenvalues 88.18% and 92.47% respectively.

Data Preprocessing

The data set provided by [1] includes 561 features, in Section 1 we used PCA to select 30 features from those features and achieved a correctness rate of 40.8%, which is less than satisfactory. Therefore, we performed our own feature extraction process to extract a same amount of features and achieved a correctness rate of 84.93%.

Classification Methods

1. Naive Bayes

Assigns a class label \( \hat{y} = C_k \) for some \( k \) as follows:

\[ \hat{y} = \arg\max_{k \in \{1,...,K\}} p(x|C_k) \]

2. KNN

KNN estimates the conditional probability for class \( C_k \) as the fraction of points in \( N_0 \) whose response values equal \( j \):

\[ P(y = C_k|X = x_0) = \frac{1}{K} \sum_{i=1}^K I(y_i = C_k) \]

3. SVM

The objective function for SVM is:

\[ \min_{\gamma, w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \]

s.t. \[ y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, ..., m \]

\[ \xi_i \geq 0, i = 1, ..., m \]

4. Softmax

The probability that an input vector \( x \) is a member of a class \( i \) is:

\[ P(\hat{y} = C_k|X = x, W, b) = \frac{e^{x_i w_k}}{\sum_{j=1}^K e^{x_j w_j}} \]

The prediction label \( \hat{y} \) should be:

\[ \hat{y} = \arg\max_{\gamma, w, b} P(\hat{y} = C_k|X, W, b) \]

5. Multi-Layer Perceptron

Results

Table 2: Classification Accuracy Percentages for different classification models.

<table>
<thead>
<tr>
<th>Model</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes</td>
<td>0.7802</td>
<td>0.7433</td>
<td>0.7391</td>
<td>0.7378</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>0.8264</td>
<td>0.8931</td>
<td>0.8966</td>
<td>0.9042</td>
<td>0.8960</td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>0.9269</td>
<td>0.9371</td>
<td>0.9377</td>
<td>0.9450</td>
<td>0.9456</td>
<td></td>
</tr>
<tr>
<td>Softmax</td>
<td>0.8863</td>
<td>0.9150</td>
<td>0.9237</td>
<td>0.9167</td>
<td>0.9297</td>
<td></td>
</tr>
<tr>
<td>MLP</td>
<td>0.8873</td>
<td>0.9206</td>
<td>0.9291</td>
<td>0.9212</td>
<td>0.8953</td>
<td>0.9342</td>
</tr>
</tbody>
</table>

References