Principal Component Analysis as an Integral Part of Data Mining in Health Informatics

Abstract

keywords:

1. Introduction

dead
2. Background

2.1. Mathematical Notation

2.2. Linear Regression

\[ y = b_0 + \sum_{k=1}^{m} b_k x_k \]

2.3. Principle Component Analysis

\[ y = a + bx \]
Definition 1.
\[ \lambda \quad \text{non-zero} \quad x \quad x \quad \lambda \cdot x, \]
\[ \lambda \quad x \]

Definition 2.
\[ x \quad y \quad \sigma \quad y, \quad y \quad \sigma \cdot x, \quad \sigma \]

Note 1. \( \lambda \).

Note 2. \( \sigma \)

\[ x \quad m \quad n \]

\[ \text{symmetric} \]

Note 3.
\[ u \quad u \quad u \]
\[ u \]
\[ \text{get's rid of the asymmetrical ordering in favor of a right handed system of representation} \]
2.4. Logistic Regression

Precision = \frac{TP}{TP + FP}
Recall = \frac{TP}{TP + FN}

\[ F_1 = \frac{2PR}{P + R} \text{ or } F_1 = \frac{2TP}{2TP + FP + FN} \]

3. Hybrid Model
algorithms replace the existing traditional algorithms.

Here is the hybrid algorithm in 2D, it can make use of the following two methods.

Improved linear Regression

\[
\begin{align*}
\text{Input} & \quad x, y \\
\text{Output} & \quad y = a + bx \\
\text{Method} & \\
\text{Traditional} & \quad \sum_{i=1}^{n} (y_i - a - bx_i)^2 \\
\text{New} & \quad \sum_{i=1}^{n} \frac{(y_i - a - bx_i)^2}{\sqrt{(1 + b^2)}}
\end{align*}
\]

Improved Non-linear Logistic Regression

\[
\begin{align*}
\text{Input} & \quad x, y \\
\text{Output} & \\
\text{Method} & \\
\text{Improved} & \quad y \rightarrow \log_e\left(\frac{y}{1-y}\right) \\
\text{New approach} & \quad y = a + bx \\
\text{Improved} & \quad y \rightarrow \frac{e^y}{1 + e^y}
\end{align*}
\]

4. Experimental Analysis

<table>
<thead>
<tr>
<th>Procedure/Condition</th>
<th># of Cases</th>
<th>Hospital Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSHPD</td>
<td>6000</td>
<td>1</td>
</tr>
<tr>
<td>OSHPD</td>
<td>6000</td>
<td>2</td>
</tr>
<tr>
<td>OSHPD</td>
<td>6000</td>
<td>3</td>
</tr>
<tr>
<td>OSHPD</td>
<td>6000</td>
<td>4</td>
</tr>
<tr>
<td>OSHPD</td>
<td>6000</td>
<td>5</td>
</tr>
</tbody>
</table>

We use the OSHPD dataset for our experiments. Our experiments are focused on both the hospital location and the number of cases per hospital. The data is published by California’s Office of Statewide Health Planning and Development (OSHPD) for our purposes. Our experiments are not limited to data mining per se, but to show the feasibility of improved regression line and logistic regression. Improved linear Regression and PCA are adapted. The additional advantage occurs when we adapt the regression line. The additional advantage occurs when we adapt the regression line. This is where the human reviewer comes in.
5. Conclusion

6. References

CIO
Pattern Recognition and Machine Learning
Business Insider

In AMIA Annual Symposium Proceedings,

Journal of General Internal Medicine
Harward School of Public Health Magazine

IMS Institute Reports

Forbes Tech Linear Algebra

Social Computing, Behavioral-
Cultural Modeling and Prediction
In Proceedings of the 2012 international workshop on Smart health and wellbeing