Background of the project:

Advances in wireless sensor networks and wearables have positively changed e-health by opening cheaper ways to perform patient situation management. Portable wearable devices that measure body vital signs are being developed but they have not penetrated into clinical practice, mainly due to insufficient research in “intelligent” analysis methods that are sufficiently robust to perform functions such as: time series analysis, monitoring displays, decision making, and generating warnings when threshold values are reached. We are not trying to replace doctors but creating a large-scale clinical tool to help medical professionals perform the previously mentioned functions.

Standard of Care:

The South African Triage Scale is existing standard of care used in most hospitals in South Africa. This is a manual early warning system (EWS) i.e. involve the clinician making a manual observation of a patient’s vital signs, applying univariate scoring criteria to each vital sign in turn (e.g., “score 3 if pulse exceeds 181 beats per minute”), and then escalating care to a higher level if any of the scores assigned to individual vital signs, or the sum of all such scores, exceeds some threshold [1]. This standard scale makes use of an EWS systems such as the one shown in Table 1, the scores 3, 2, 1 and 0 represent High Level, Medium Level, Low Level and Normal level respectively. Before a novel system is introduced to replace an existing system in healthcare, influential evidence must be acquired concerning the effectiveness of the existing standard of care. Evidence has motivated for an integrated system that learns from existing patient data [1].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Score** | **3** | **2** | **1** | **0** | **1** | **2** | **3** |
| **Systolic (mmHg)** | <= 59 | 60-79 | 80-99 | 100-130 | 131-160 | 161-200 | 201 >= |
| **Diastolic (mmHg)** | <= 44 | 45-49 | 50-59 | 60-85 | 86-90 | 91-110 | 111 >= |
| **SpO2 (%)** | <= 79 | 80-91 | 92-94 | 95-100 |  |  |  |
| **Pulse (per munite)** | <= 44 | 45-49 | 50-59 | 60-100 | 101-120 | 121-180 | 181 >= |

Table1: The EWS System used for the dataset in this project.

Project:

You have been provided with a scored dataset containing the following vital signs: Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Peripheral Oxygen Saturation (SPO2), Pulse and a risk score (Score). If you decided to only use the provided dataset please make assumptions for example: assume that (1) each row of the vital signs was acquired after every second and (2) the dataset is from different individuals. This dataset provided was derived from record number a42397n of the Physio Bank ATM. You can download time stamped unscored datasets from the link: <http://www.physionet.org/cgi-bin/atm/ATM> (see Appendix A for more information or see your tutors for assistance.)

Implement your algorithms to perform anomaly detection and report on your findings. [\*\*]Marks

Your report should include the following:

1. Give a background about your two algorithms for example see Table 2 and Table 3. (N.B. can also be written in paragraph form) [\*\*] Marks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Arithmetic method using lookup tables | Multivariate Linear Regression | K-means Clustering and Parzen Window (or Gaussian ) estimator | Support Vector Machines |
| **Type of Learning Algorithm** | **No learning involved**  **( simple arithmetic e.g. addition)** | Supervised learning – training data consists of **input/output pairs** | Unsupervised learning – no output values **and** learning task is to **gain understanding of the process** that generated the data.  Density estimation, clustering, learning the support of the distribution | Supervised learning – training data consists of **input/output pairs** |
| **Function** | **Linear function** – with predefined weights | **Linear function** – weights are calculated from the training set | **Gaussian Function** | Uses a hypothesis space of **Linear** **Functions** |
| **Output** | Real-valued outputs **(regression problem)** | Real-valued outputs (**regression problem**) | Estimated density | Real-valued outputs (**regression problem**) |

Table 2: The characteristics of the novel detection algorithms.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Multivariate Linear Regression | K-means Clustering and Parzen window estimator algorithm | Arithmetic method using lookup table 1 | Support Vector Machine |
| Use of the algorithm | Detecting patient deterioration.  Generates warnings when threshold values are reached.  Generates priority list from all patient data in order to perform patient prioritization. | Generates priority list from all patient data in order to perform patient prioritization. | Generates priority list from all patient data in order to perform patient prioritization. | Detecting patient deterioration.  Generates warnings when threshold values are reached.  Generates priority list from all patient data in order to perform patient prioritization. |
| Efficiency | **Time complexity is not affected** by the amount of **data**. | **Time complexity is affected** by the amount of **data**. | **Time complexity not affected** by amount of **data**. | **Time complexity not affected** by amount of **data** |
| Scoring | Uses the **EWS**  **system table 1** to score the data before training  Algorithm learns from the data. | A **patient status index** is calculated from the data provided; k clustering means are calculated from history data. | Uses **EWS system table 1** to score the patients (the algorithm uses expert knowledge to score the patients). | Uses the **EWS**  **system table 1** to score the data before training  Learns. |

Table 3: The basic characteristics of the algorithms observed during the experiments.

1. The results obtained from your implementations for example see Table 4. [\*\*]Marks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Multivariate Linear Regression** | **K-means Clustering and Parzen window estimator algorithm** | **Arithmetic method using lookup table 1** | **Support Vector Machine** |
| Time Complexity  (approximately) | seconds | seconds for only 10 clusters | seconds | seconds |
| Accuracy  (approximately) | 99.9996560811529 % | NaN | NaN | 80.67 % |

Table 4: The performance results in the form of time complexity and accuracy.

1. Evaluate your anomaly detection algorithms:
   1. If applicable plot graphical representation of effects of your EWS scoring methods on performance of your algorithms e.g. Figure 1 [\*\*] Extra Marks

Figure 1: The smaller the score values the more accurate the algorithm becomes.

* 1. Plot your Learning Curve e.g. see Figure 2

Also include its definition and interpret your learning curve [\*\*] Marks NB: Interpretation of your learning curve is very important.

Note this can be the Accuracy Rate (%)

Figure 2: The learning curve for the Multivariate Linear Regression by Gradient Descent.

* 1. Report on any other method you can use to evaluate your anomaly detection algorithm’s performance. [\*\*] Extra Marks

1. Write your conclusion stating: the best algorithm of the two and reason for your choice. You can also write other content that you deem fit to be included in the conclusion. [\*\*]Marks

NB: .All your findings and experimental outcomes of your speculations should be documented. Provide figures and information other the examples you have been provided. This is merely a guideline of the content expected. Your document should exhibit a very good appearance (you can write it in the form of a publication) and also provide pseudo code and equations. [\*\*] Marks

References

[1] D. Clifton, D. Wong, L. Clifton, S. Wilson, R. Way, R. Pullinger and L. Tarassenko, “A Large Scale Clinical Validation of an Integrated Monitoring System in the Emergency Department,” IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 835-842, July 2013.

Appendix A

Patient datasets can be downloaded from the link: <http://www.physionet.org/cgi-bin/atm/ATM>.You can download as many records as you want, choose the following settings:

1. Database: MIMIC II Waveform DB, v2 Numerics [deprecated, use v3] (mimic2db/numerics)
2. Length: to end
3. Tool box: Export signals as CSV