MINI PROJECT
MORTALITY PREDICTION IN ICU

Arushi Jain (2012130)
Shweta Sood (2012164)
Srishty Saha (2012107)

1. Aim

As a part of Xerox Mortality Prediction Challenge 2015 [3], we tried to predict the risk of death (mortality) in patients admitted to ICU within hospitals.

2. Data Description

- The data consists of 5990 (simulated) patient records where each patient record has the following variables:
    1. ID: a unique identifier for each patient
    2. Age
    3. 6 Vitals
    4. 25 Labs
    5. Timestamps: measurement time relative to first measurement for patient
    6. ICU flag: indicates whether a patient is in ICU or not at a given time
    7. Mortality label: indicates whether a patient survived or died (the label or outcome variable)
- Vitals and Labs are time-series variables associated with timestamps
    1. Each vital or lab variable may be measured multiple times (at different timestamps) for a patient
    2. Not all variables are measured at all timestamps. Some variables may not be measured at all in a patient record
- Measurements that are not made at a time stamp has the value 'NA'. The first timestamp for a patient is always 0
- At each timestamp the ICU flag indicates whether the patient is inside an ICU or outside (in the hospital)

3. Data Preprocessing

Data were processed to remove extreme outliers and obvious errors using a filter of normal values for each parameter. Missing information (NA) was replaced with normal values of each feature (25 labs and 6 Vitals).

4. Feature Extraction

1. For the purpose of feature extraction we computed the mean, minimum, maximum, last data point value and the trend estimation (the measurements were fitted to a linear line, and the slope of the line was used for the trend estimation) for all variables except the static variable of age, ICU type
2. Next, we eliminated the variables that were not varying much for all patients or were mostly NAs originally
3. To figure out relevant variables, we referred [1], which was a paper on similar lines that used similar data for mortality prediction. We also, used SAPS - II calculator [2] to figure out which values amongst mean, minimum, maximum etc are relevant for a particular feature
4. This reduced our number of features from 155 to 79
5. This also helped in consolidating data across multiple timestamps for a patient to just a single row feature vector represented by the aforementioned statistics

5. Classification

1. In order to perform, classification, we zeroed in on Artificial Neural Network approach after initial probes at regression techniques, as the ANN gave a higher classification accuracy and the papers referred confirmed the same [1]
2. The ANN was run on a 79 features comprising the various statistics discussed.
3. Two-layer network with 30 neurons in the first and second hidden layers respectively was used for training as this particular selection gave best accuracy, among various permutations of number of layers and neurons tried
4. The final_prediction either 1 or 0 was calculated using feature set
   a. final_prediction = 1 signified patient had died
   b. final_prediction =0 signified patient had survived

   where, Prediction was made for each patient only when the patient is in ICU i.e, ICU flag =1 and if the sequence of predictions for the patients contains only zeros, then the final prediction is 0, otherwise 1.

6. Results

- We performed testing on the dataset of 1198 patients provided.
  Following metrics were used for calculating the final score as per the specifications of the challenge [3].
  1. Sensitivity_score : TP/(TP+FN) = 0.3048
  2. Specificity_score : (Specificity - 0.99)*100={TN/(TN+FP) -0.99}*100= 0.104
  3. Median_pred_time_score : Median_pred_time_clipped_72/72
     a. Prediction time is only defined for patients whose final prediction is 1. It is the difference between the last timestamp (i.e. the last measurement made for the patient) and first timestamp with a prediction of 1. We found its median
     b. If Median_pred_time< 72 hrs:
        Median_pred_time_clipped_at_72 = Median_pred_time
        else:
        Median_pred_time_clipped_at_72 = 72 hrs

  Median_pred_time_score : 72/72 = 1 (Median_pred_time >72 for all timestamps)

- The output of the neural network is a score between 0 and 1. We used a fuzzy threshold found at the time of training to discriminate between the 2 classes
- Final_score = 0.75*Sensitivity_score + 0.2*Median_pred_time_score + 0.05*Specificity_score
- Final_score = 0.75*0.3048 + 0.2 *1 + 0.05*0.104

Final_score = 0.4338 ~ 0.43

This score is comparable to the highest score in the challenge which was 0.57

7. References

[1] Henian Xia, Brian J Daley, Adam Petrie, Xiaopeng Zhao. A Neural Network Model for Mortality Prediction in ICU. Computing in Cardiology 2012; 39:261-264

[2] SAPS-II calculator : http://clincalc.com/IcuMortality/?example

[3] Xerox Mortality Prediction Challenge : https://www.hackerrank.com/contests/xerox-research-innovation-challenge-2015/challenges/xerox-predict-mortality