Artificial Intelligence: Project 3

Health Monitoring for the Elderly

Agathe Benichou, Wassim Gharbi, Greg Shindel, Nick Turney, Thanh Vu

Problem statement

- Walgreens has developed a new watch-like technology that records vitals from older patients in order to monitor their health and address diseases before they develop.

- Can break down the problem into four subsections:
    1. Better inform the doctor, patients and their families.
    2. Determine the best times to administer medication.
    3. Determine the efficacy of medications.
    4. Determine if a patient is actually taking medication.

Parameter Highlights

Parameter Highlights

Input

- This watch will be **initialized** by the doctor.
- This watch will collect vital signs such as x, y, z.

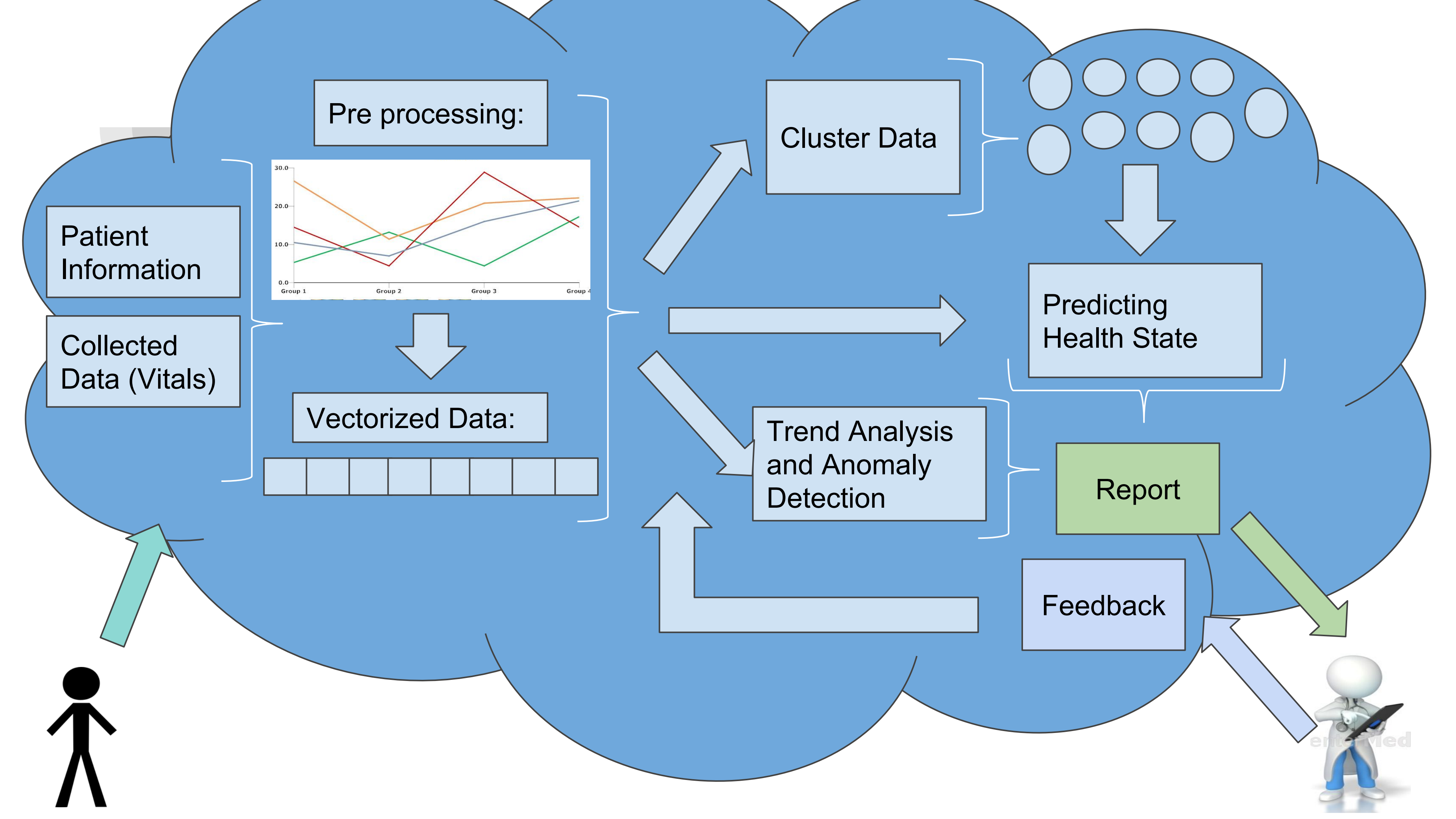
Assumptions

- The watch is charged by body heat.
- Patients are 50+ years old and have doctors appointments at least once a month.
- After an appointment, the doctor will update the watch with any developments.

Output

- An analysis of the data based on trends or patterns which the doctor can use.
- Information regarding medication that patient is taking.

Solution Overview


Pre processing:
Patient Information
Collected Data (Vitals)
30.0
20.0
10.0
0.0
Group 1
Group 2
Group 3
Vectorized Data:
Cluster Data
Predicting Health State
Trend Analysis and Anomaly Detection
Report
Feedback

Preprocessing Steps

Preprocessing Step 1 - Gathering Data

- What data are we collecting?
  - **Any vital which can be measured numerically and non-invasively**
  - Ex: Temperature, Blood Pressure, Heart Rate, Glucose Level
- How often is this data collected?
  - Continuous measurements for optimal precision
- Where is this data sent/stored?

Wrist Band

- Single day continuous curve
- Linear approximations for prev. days

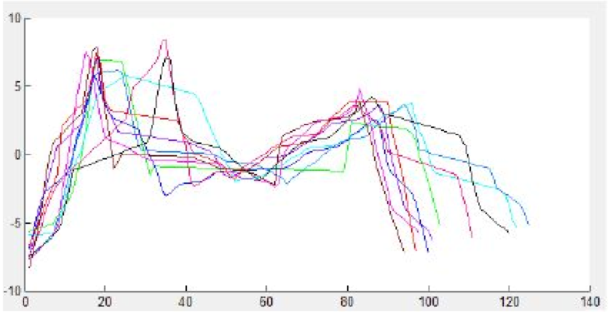
External Database (User Profile)

- Dynamic Linear Approximations and Data Summarization
- Patient Metadata

Preprocessing Step 2 - Simplifying Data - Piecewise Linear Approximation

Daily:

Continuous Curve → Piecewise Linear Approximation → K * {Slope, Intercept, Magnitude}



Append to end

Over D days:

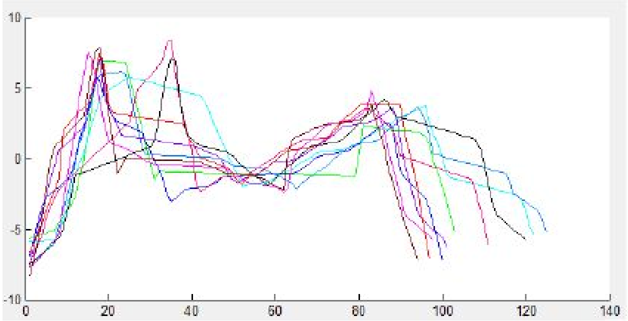
K * D lines ordered by time

Doctor's Visit:

Offload Data → External Server

Preprocessing Step 2 - Simplifying Data - Piecewise Linear Approximation

- For Doctor Visit V:
  - Rerun approximations on past vitals:
    - For Visit **(V-X),** Use **(K-X)** line segments
  - Reduces **significance** and **size** of historical data
- Resultant Multidimensional Vector:

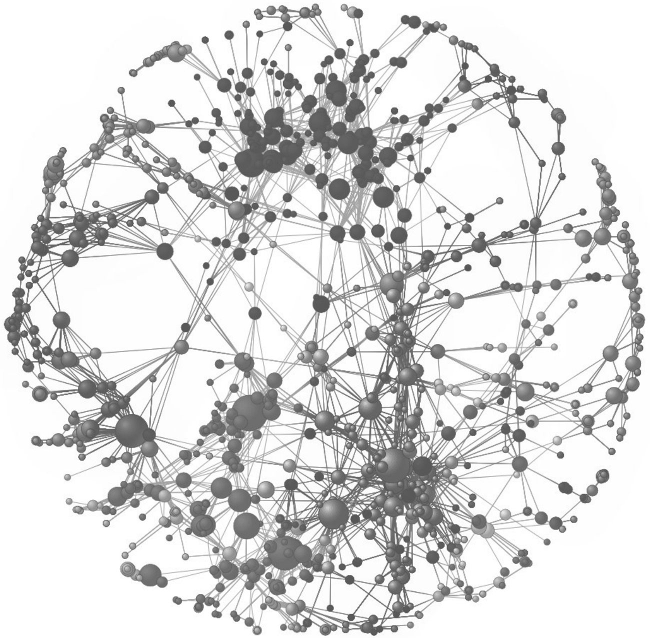


Metadata	Data since last visit: K * D entries {Slope, Int, Len}	Visit V-1: (K-1) * D entries	...	Visit V-K: 0 entries

Main Algorithms

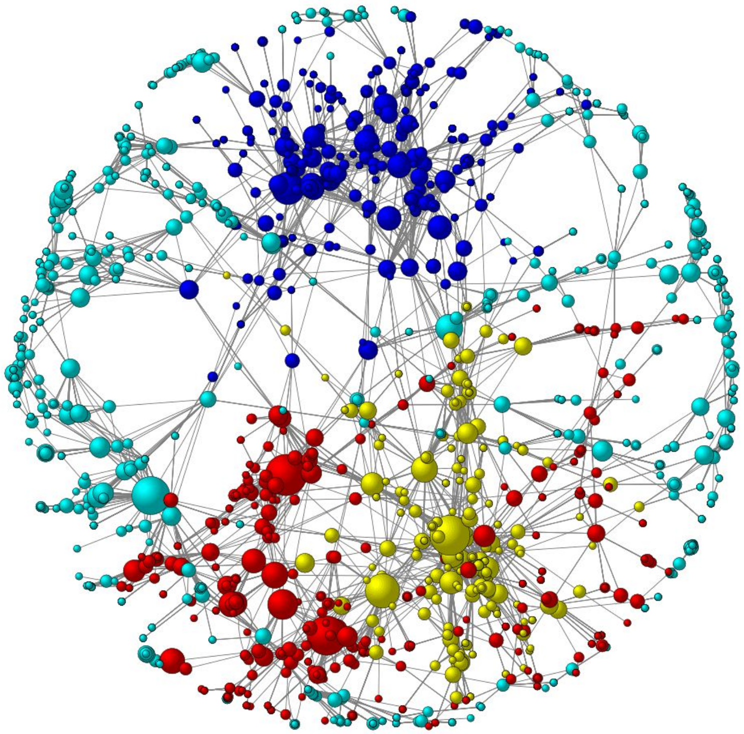
Subproblem 1: **Clustering data**

- Input:
  - Demographic data: sex, birth, ethnicity, blood-type, ...
  - Medical history: diseases, injuries, food allergies, ...
- Output:
  - Clusters of patients with similar background



Subproblem 1: **Clustering data**

- Algorithm:
  - Dimension-based Partitioning and Merging (DPM) [1]
- Motivation
  - Large scale high dimensional datasets
  - Unknown number of clusters
  - Speed and Scalability



Subproblem 1: **Clustering data**

- Step 1: Data partitioning
    - Partition patients into dense volumes by considering the histogram of each dimension

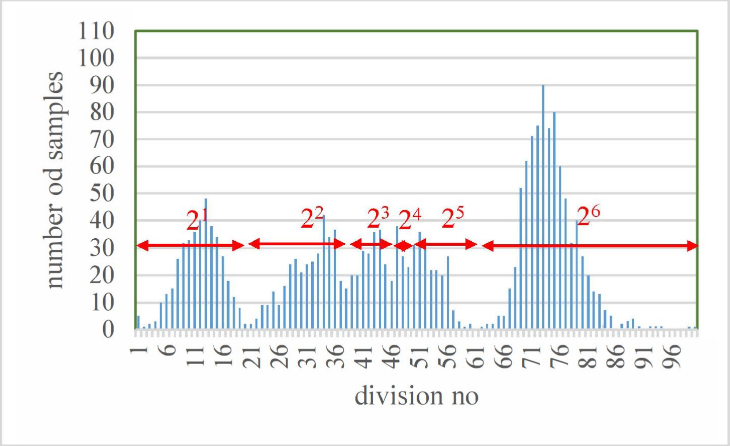


Figure 1: histogram for data dimension no 1, t = 100

Subproblem 1: **Clustering data**

- Step 1: Data partitioning
  - Partition patients into dense volumes by considering the histogram of each dimension
- Step 2: Noise detection
  - Filter bad partitions by evaluating dimensional densities

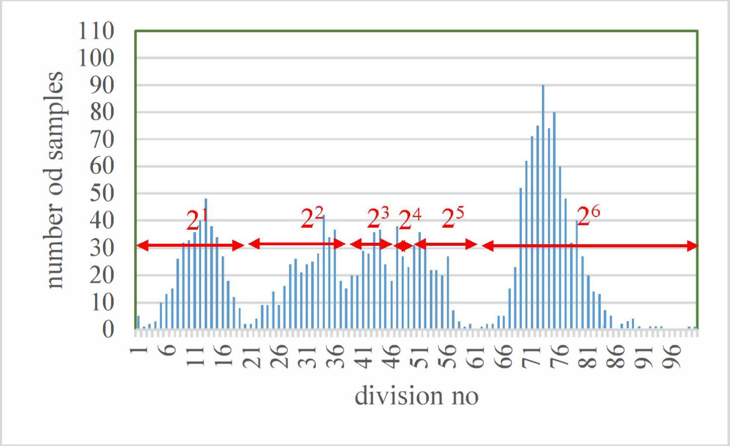


Figure 1: histogram for data dimension no 1, t = 100

Subproblem 1: **Clustering data**

- Step 1: Data partitioning
  - Partition patients into dense volumes by considering the histogram of each dimension
- Step 2: Noise detection
  - Filter bad partitions by evaluating dimensional densities
- Step 3: Cluster Formation
  - Merge dense partition into clusters through nearest neighbor test

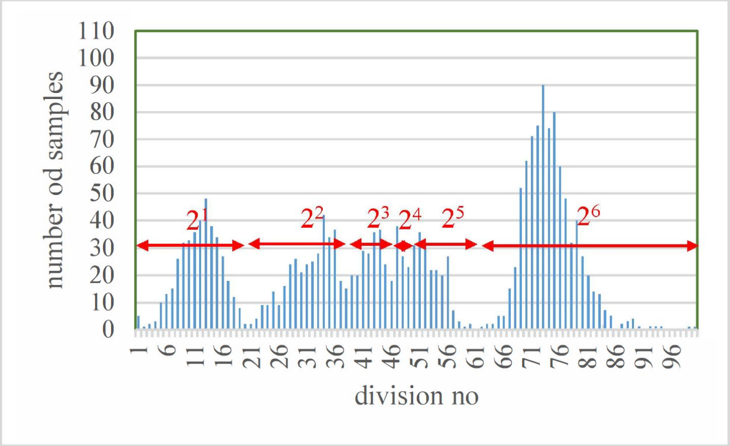
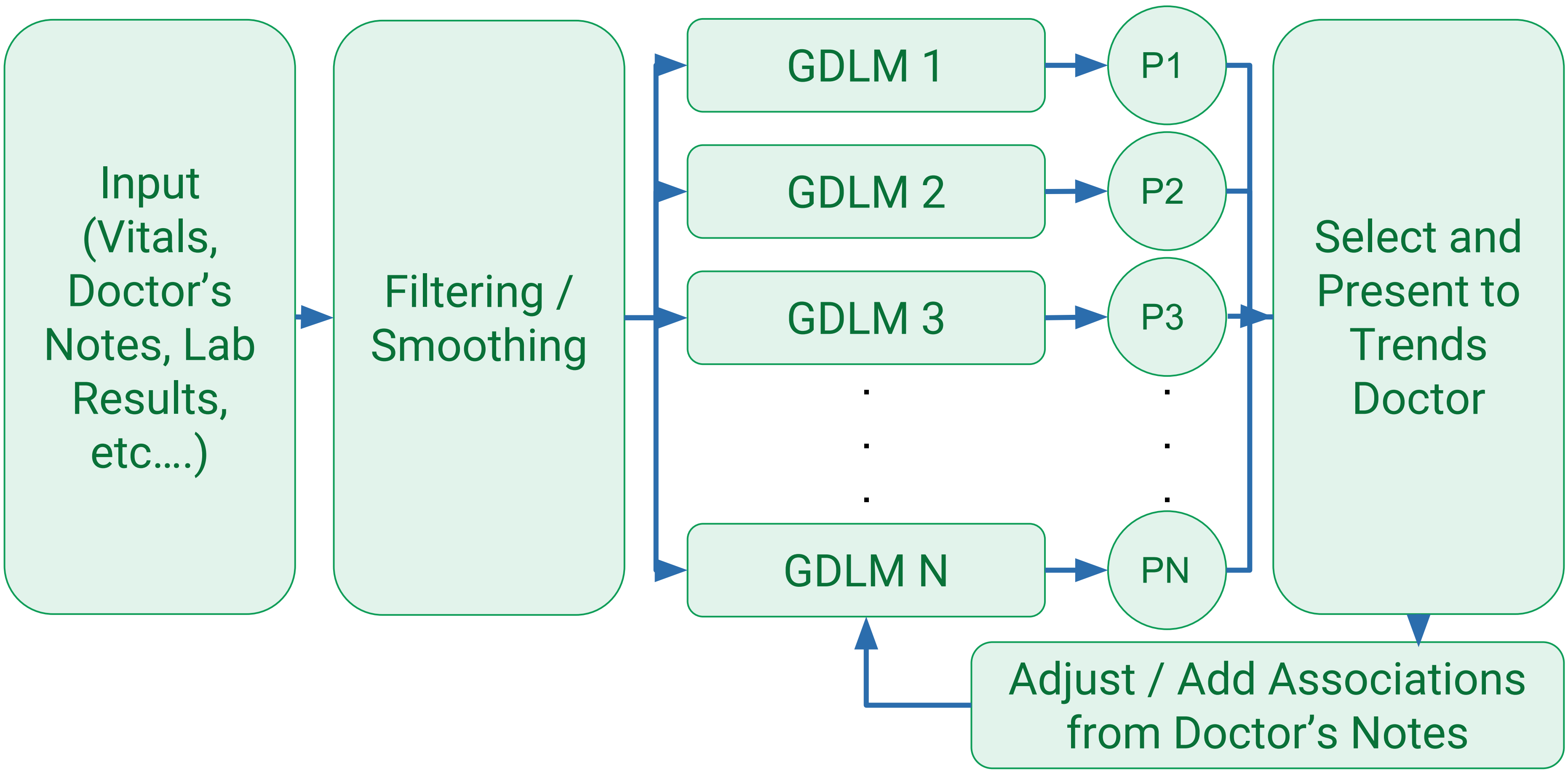


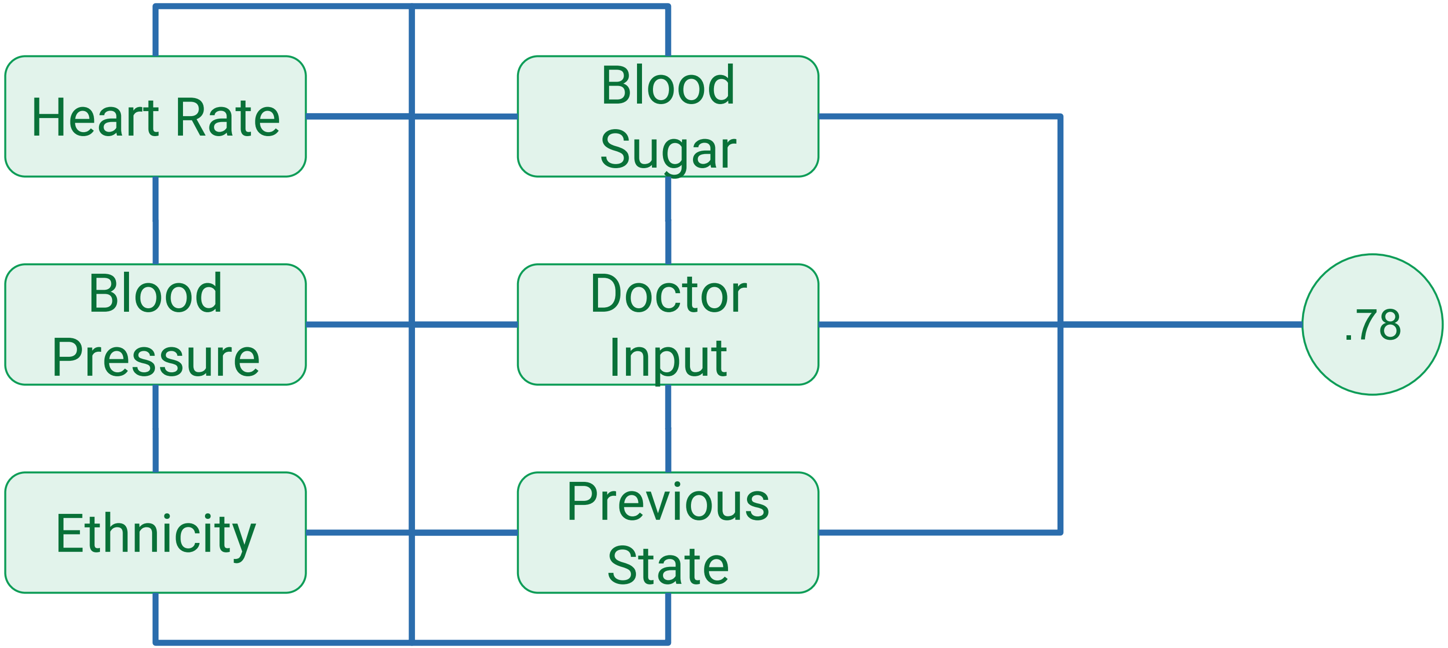
Figure 1: histogram for data dimension no 1, t = 100

$$dist(x, y) = \max_{1 \leq i \leq d} |\beta_i(x_i - y_i)|$$

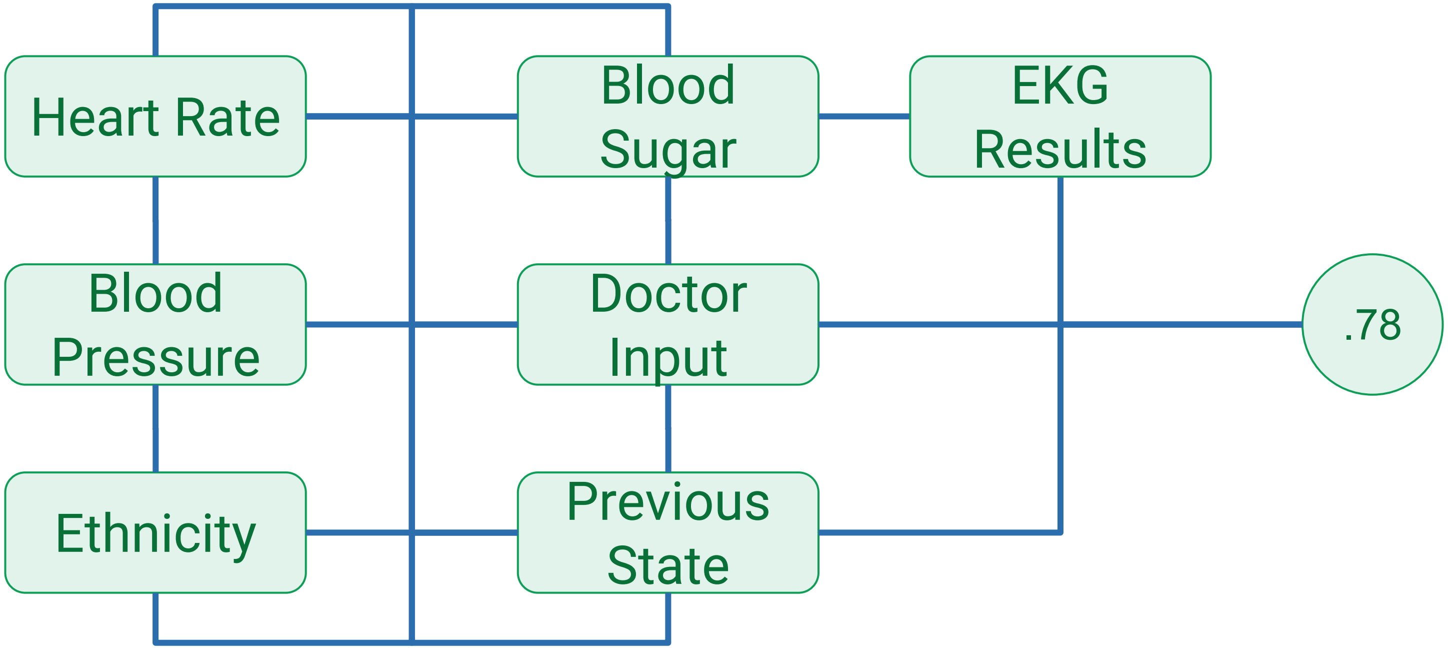
Subproblem 2: **Predicting Health State**



Subproblem 2: **General Dynamic Linear Model**



Subproblem 2: **General Dynamic Linear Model**



Subproblem 3: **Trend Analysis and Anomaly Detection**

(3) gather data about possible long term changes (e.g., early onset of diabetes)
(4) gather data about the efficacy of medications (e.g., how long does the effect of insulin persist, are the medications adversely affecting each other)
(5) monitor the ability of the patient to ingest medications (e.g., how often do they forget to take the medication)

http://www.anodot.com/blog/close...

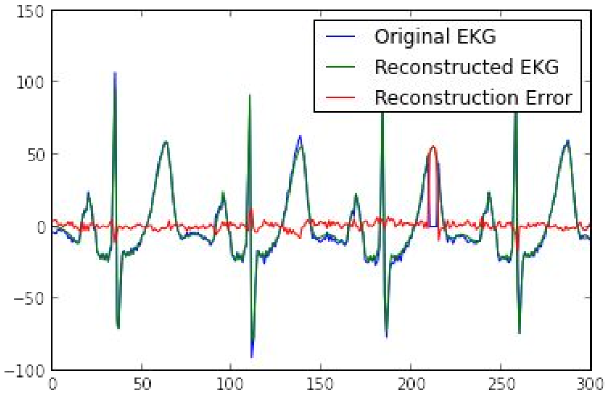
Subproblem 3: **Trend Analysis and Anomaly Detection**

We identify the sensor measurements as **deterministic** time series with **additive seasonal patterns.**

Deterministic: The effects of the shocks/anomalies revert to the trend in long run (ex. eating/fasting)
Seasonal patterns: Regular, predictable, short-term fluctuations (ex. Taking medicine)
Additive: The effects of shocks/anomalies are added to the regular pattern

http://www.anodot.com/blog/close...

Subproblem 3: **Trend Analysis and Anomaly Detection**



We identify two types of anomalies: **shock** and **ramp**

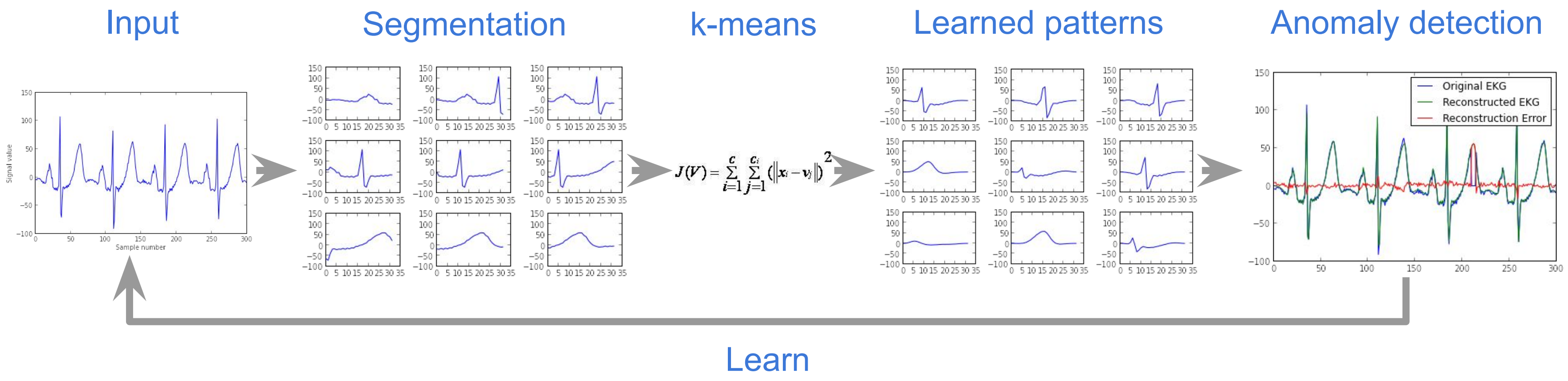
Shock : instantaneous sudden change

Ramp: slow long-term change

http://amid.fish/anomaly-detection-with-k-means-clustering

Shock detection

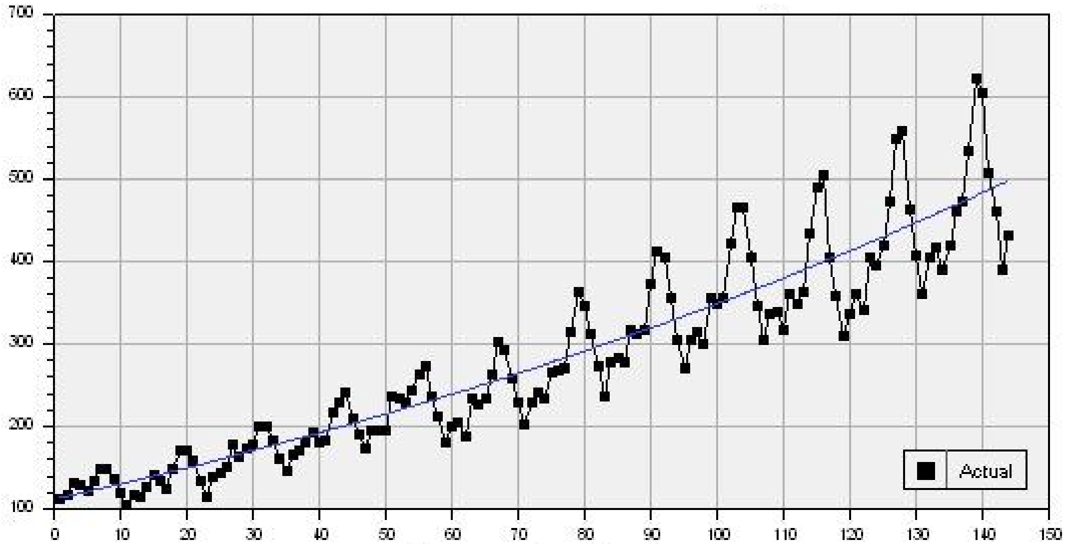
1. Perform **overlapping segmentation** on the time series
2. Use **k-means clustering** on the segmented data to obtain the basic patterns that form the signal
3. **Reconstruct** the signal using the predicted patterns
4. Find anomalies by looking at the "**errors**" (difference between input signal and predicted signal)
5. **Learn** new patterns



http://www.cs.cmu.edu/~./awm/tutorials/biosurv01.pdf

Ramp detection (trend)

1. Perform dynamic summarization of the time series (more recent data points have greater effect)
2. Perform **linear regression** (least squares) on the time series
3. Obtain a starting value and a slope indicating whether the data is increasing/decreasing and at what rate.



http://www.cs.cmu.edu/~./awm/tutorials/biosurv01.pdf

Subproblem 3: **Trend Analysis and Anomaly Detection**

(3) gather data about possible long term changes (e.g., early onset of diabetes)
→ Long-term ramp-up in blood sugar
(4) gather data about the efficacy of medications (e.g., how long does the effect of insulin persist, are the medications adversely affecting each other)
→ Time period of daily recurring “anomalies” (trained with data before prescription)
→ High-amplitude anomalies in vital signs affected by one or both medicines
(5) monitor the ability of the patient to ingest medications (e.g., how often do they forget to take the medication)
→ Frequency of daily anomalies (absence of shocks in this case)

http://www.anodot.com/blog/close...

Questions?

References

[1] Ghanem, T. F., Elkilani, W. S., Abdelkader, H. M., & Hadhoud, M. M. (2015). Fast Dimension-based Partitioning and Merging clustering algorithm. Applied Soft Computing, 36, 143-151. doi:10.1016/j.asoc.2015.05.049

[2] http://www.cs.cmu.edu/~./awm/tutorials/biosurv01.pdf

[3] http://amid.fish/anomaly-detection-with-k-means-clustering