Machine Learning

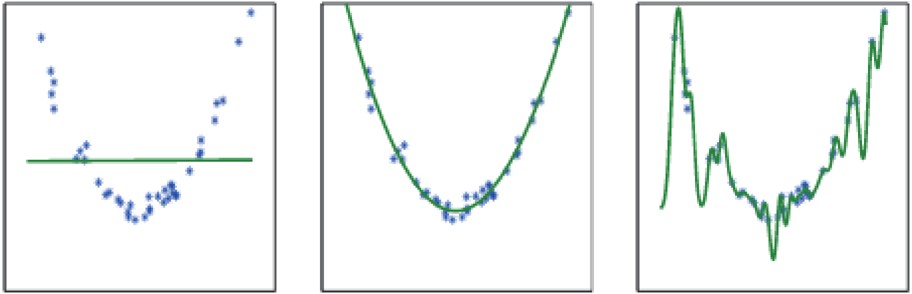
Anomalies, Music, Time Series

Jeff Abrahamson

octobre 2020

Review

Underfitting, overfitting



Linear Regression

Logistic Regression

SVM

CART

Random Forests

PCA

Handwriting Recognition

Clustering

Anomaly Detection (not time)

Introduction to Anomaly Detection

- Supervised
- Unsupervised

Introduction to Anomaly Detection

Supervised anomaly detection:

- Training data: normal, abnormal
- Train a classifier

So reduced to existing problem of supervised classification.

Introduction to Anomaly Detection

Unsupervised anomaly detection:

- Mostly, this is clustering
- Increasingly, this is neural networks in advanced applications

Introduction to Anomaly Detection

Applications:

- Intrusion detection (physical or electronic)
- Fraud detection
- Health monitoring (people, animals, machines)

Introduction to Anomaly Detection

Techniques:

- Density: kNN, local outlier factor
- SVM
- Clustering: *k*-Means

kNN techniques and variations

- Voronoi diagrams
- aNN

k-Means

Local Outlier Factor (LOF)

- Measure average density using kNN
- Points with low local density are suspect outliers
- There is no good thresholding technique

Local Outlier Factor (LOF)

Let a be an object (point) in the set of samples.

Let $N_k(a)$ be the set of k nearest neighbours to a.

Define the k-distance from a:

$$d_k(a) = \max_{p \in N_k(A)} d(a, p)$$

Local Outlier Factor (LOF)

Define now the reachability distance:

$$r_k(a, b) = \max(d_k(a), d(a, b))$$

In otherwords, r_k is the distance between two points, but is no less than the k-distance.

So all the points in $N_k(a)$ are considered equally r_k distant from a.

Math note: r_k is not a true distance function.

Local Outlier Factor (LOF)

Define the *local reachability density* of object *a* by

$$\operatorname{lrd}(a) = \frac{1}{\left(\frac{\sum_{b \in N_k(a)} r_k(a,b)}{|N_k(A)|}\right)} = \left(\frac{|N_k(A)|}{\sum_{b \in N_k(a)} r_k(a,b)}\right)$$

This is the (inverse of the) average reachability distance of the *k* nearest neighbours.

Local Outlier Factor (LOF)

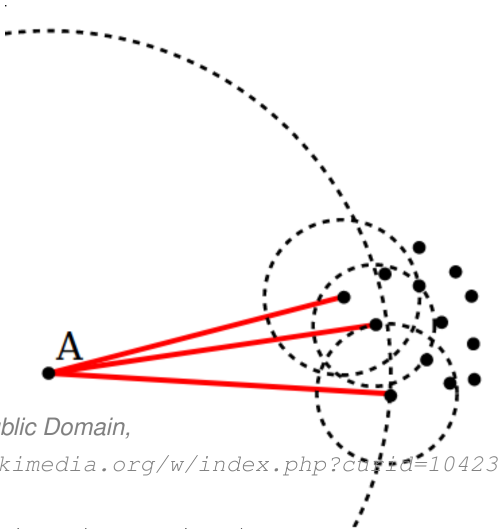
The

$$\mathrm{LOF}_k(a) = \left(\frac{\sum_{b \in N_k(a)} \frac{\mathrm{lrd}(b)}{\mathrm{lrd}(a)}}{|N_k(a)|} \right) = \left(\frac{\sum_{b \in N_k(a)} \mathrm{lrd}(b)}{|N_k(a)| \quad \mathrm{lrd}(a)} \right)$$

Interpretation:

- 1 indicates a point is comparable to its neighbours
- < 1 indicates more densely packed than its neighbours
- > 1 indicates more sparsely packed than its neighbours

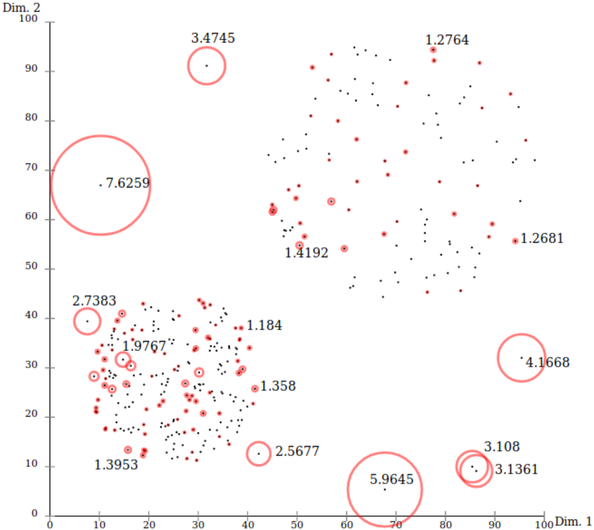
Local Outlier Factor (LOF)



By Chire - Own work, Public Domain,

https://commons.wikimedia.org/w/index.php?curid=10423954

Local Outlier Factor (LOF)



By Chire - Own work, Public Domain

Local Outlier Factor (LOF)

Advantages: intuitive, often works well (e.g., intrusion detection)

Disadvantages: fails at higher dimension (curse of dimensionality), hard to interpret

Examples

ping times

httpd response times

Examples

single/multiple host access abuse (DOS/DDOS)

Examples

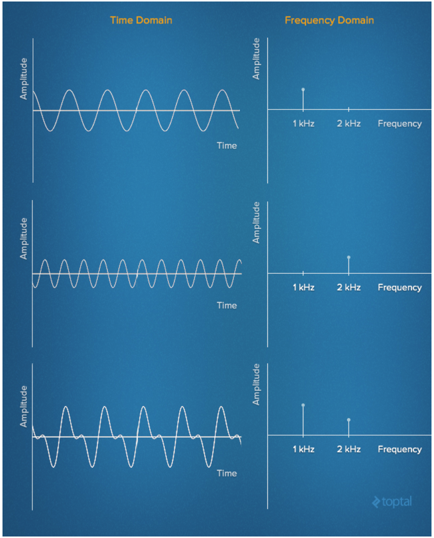
bank card fraud

Examples

spam

questions?

Music



http://www.toptal.com/algorithms/
shazam-it-music-processing-fingerprinting-and-recognition

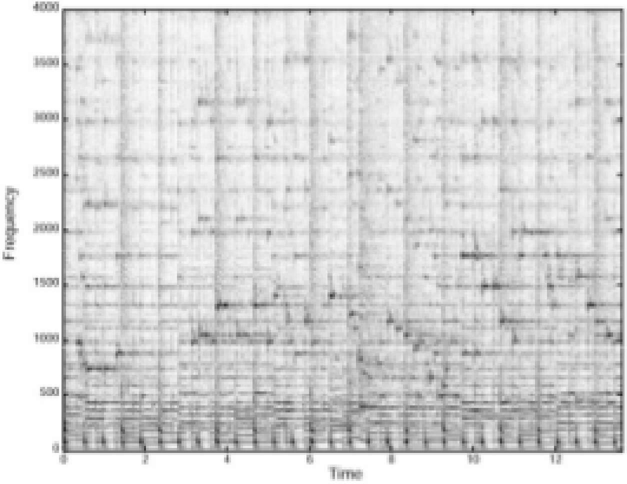


Fig. 1A - Spectrogram

https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf

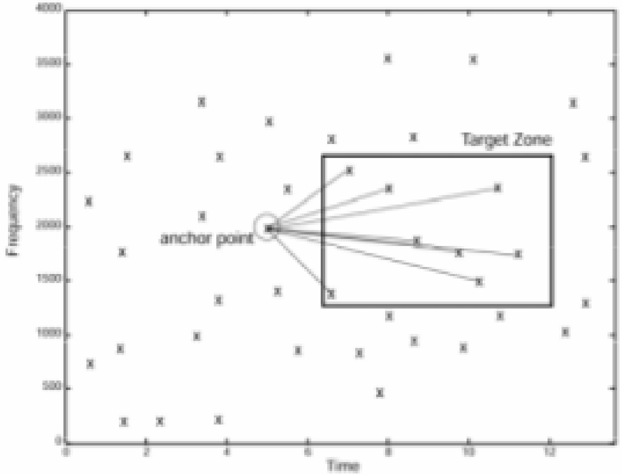


Fig. 1C - Combinatorial Hash Generation

https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf

questions?

Time Series

Introduction to time series

This is hard, but it depends on your goals. And on context.

Introduction to time series

Definition (discrete time series):

$$\{s_t \mid t \in \mathbb{R}^+ \land s \in \mathbb{R}\}$$

(though s in any vector space is fine)

Introduction to time series

Examples domains:

- Weather
- Economics
- Industry (e.g., factories)
- Medicine
- Web
- Biological processes

Introduction to time series

Why?

- Predict
- Control
- Understand
- Describe

Introduction to time series

Some strategies:

- Differencing:

$$y'_t = y_t - y_{t-1}$$

- Second-order differencing:

$$y''_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

Introduction to time series

Some strategies:

- Clustering
- Hidden Markov Models (HMM)
- Recurrent neural networks (RNN)
- Autoregressive integrated moving average (ARIMA)
  - Generalisation of autoregressive moving average (ARMA) model
  - Regress on series' own lag

One model:

$$s_t = g(t) + \phi_t$$

where

$g(t)$ is deterministic: signal (or trend)

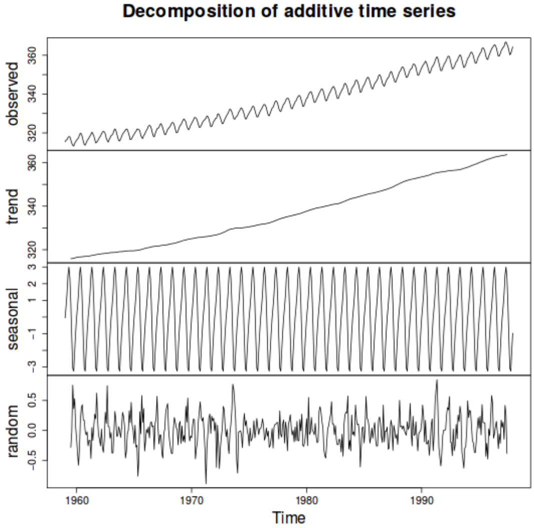
ϕ_t is stochastic noise

Introduction to time series

Variation types:

- Trend (g)
- Seasonal effect (g)
- Irregular fluctuation (residuals: ϕ)

Introduction to time series



http://www.ulb.ac.be/di/map/gbonte/ftp/time_ser.pdf

Some easy things to try

- Introduce features to break out seasonality
- Introduce lags as features
- Some domain-specific transformation

HMM

“simplest dynamic Bayesian network”

Markov Chains

A **Discrete time Markov chain (DTMC)** is a random process that undergoes state transitions**.**

Markov Chains

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & xnn \end{bmatrix} \begin{bmatrix} v_1^{(i)} \\ v_2^{(i)} \\ \vdots \\ v_n^{(i)} \end{bmatrix} = \begin{bmatrix} v_1^{(i+1)} \\ v_2^{(i+1)} \\ \vdots \\ v_n^{(i+1)} \end{bmatrix}$$

Markov Chains

$$Xv_i = v_{i+1}$$

Markov Chains

Examples:

- Random walks
- Weather (first approximation in many places)
- Thermodynamics
- Queuing theory (so also telecommunications)
- Spam

Markov Chains

Properties:

- Stochastic process
- Memoryless (“Markov property”)

HMM's

- State is not visible
- Output of state is visible

Examples: noisy sensor, medical diagnosis

HMM's

What we have:

- State space $S = \{s_1, \ldots, s_n\}$
- Observation space $O = \{o_1, \ldots, o_k\}$
- Transition matrix A of size $n \times n$
- Emission matrix B of size $n \times k$
- Initial state probabilities $\pi = \{\pi_1, \ldots, \pi_n\}$
- A sequence of observations $X = \{x_1, \ldots x_T\}$

Here

- $y_t = i \iff$ observation at time t is o_i
- $\Pr(x_1 = s_i) = \pi_i$

We want the sequence of states $X = \{x_1, \ldots, x_T\}$.

HMM's

Some pointers to learn more about HMM:

- Forward-Backward Algorith
- Viterbi Decoding
- Baum-Welch Algorithm


questions?