Using Zoom for Lectures

- **Sign in using:**

– your name

- **Please mute both:**
  - – your video cameras for the entire lecture
  - – your audio/mics unless asking or answering a question

- **Asking/answering a question, option 1:**
  - – click on Participants
  - – use the hand icon to raise your hand
  - – I will call on you and ask you to unmute yourself

- **Asking/answering a question, option 2:**
  - – click on Chat
  - – type your question, and I will answer it

Today: Outline

- **Anomaly Detection**

- **Reminders:** PS4 self score, due Apr 3

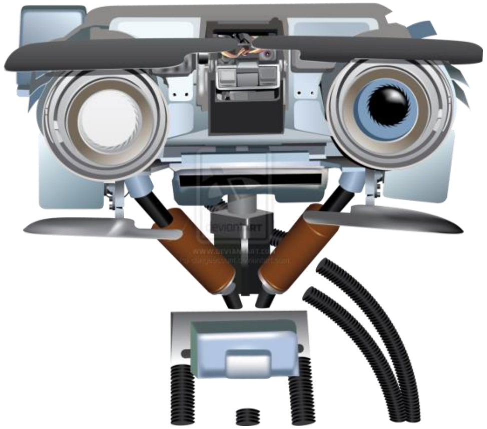
  Class Challenge is posted, due Apr 24

  (3-week challenge)

  Midterm Exam, Apr 15 during class time

  (covering material up to and including Apr 3)

  Practice Problems, will post later today



Unsupervised Learning III: Anomaly Detection

Machine Learning

Anomaly detection

- What is anomaly detection?
- Methods:
  - Density estimation
  - Detection by reconstruction
  - One-class SVM

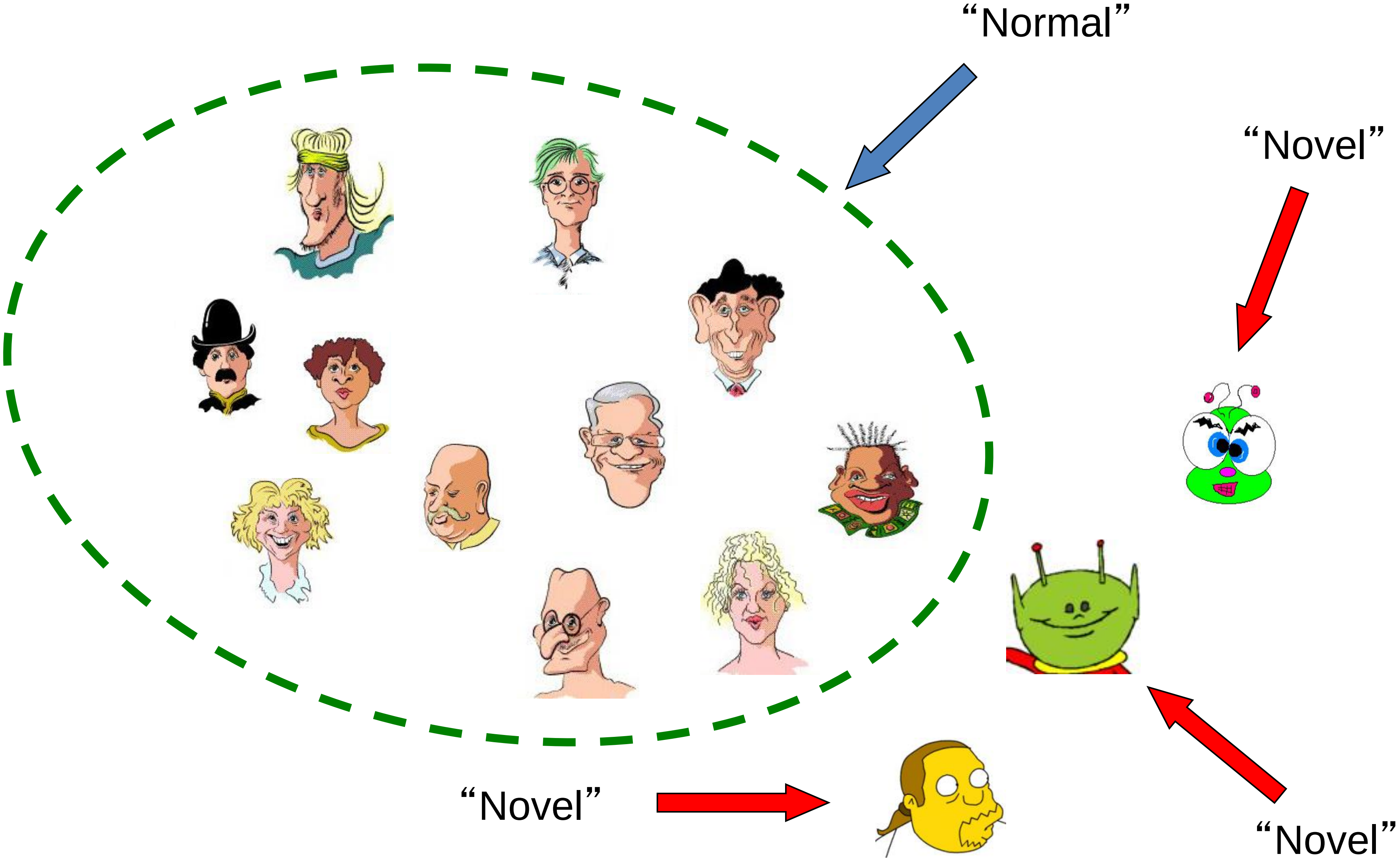
What is an anomaly?

Anomaly Detection is

- An unsupervised learning problem (data unlabeled)

- About the identification of new or unknown data or signal that a machine learning system is not aware of during training

Example 1



“Normal”

“Novel”

“Novel”

“Novel”

So what seems to be the problem?

It’s a 2-Class problem.
“Normal” vs. “Novel”

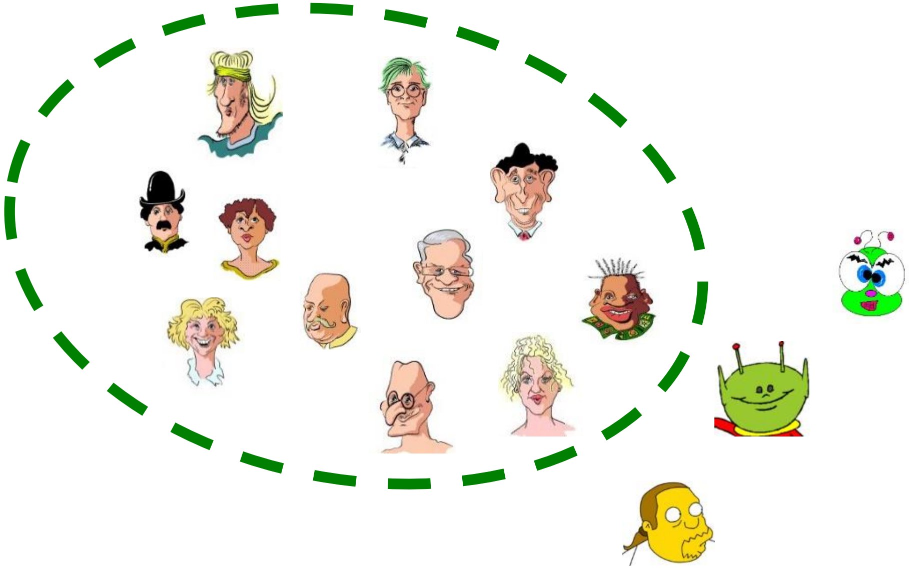
So what seems to be the problem?

It’s a 2-Class problem.
“Normal” vs. “Novel”

Wrong!

The Problem is

That “All positive examples are alike but each negative example is negative in its own way”.



One-class recognition

- Suppose we want to build a classifier that recognizes anomalous activities in an airport

- How can we collect a training data?
  - We easily assemble videos of normal airport activities like walking, checking in, etc., as positive examples.

- What about negative examples ?
  - The negative examples are… all other activities!!

- So the negative examples come from an unknown # of negative classes.

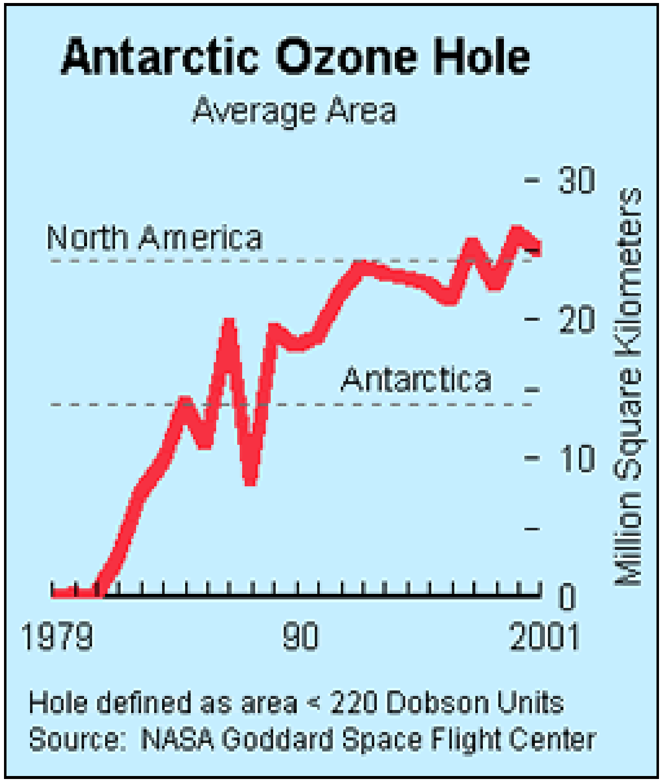
Importance of Anomaly Detection

Ozone Depletion History

- In 1985 three researchers (Farman, Gardinar and Shanklin) were puzzled by data gathered by the British Antarctic Survey showing that ozone levels for Antarctica had dropped 10% below normal levels

- Why did the Nimbus 7 satellite, which had instruments aboard for recording ozone levels, not record similarly low ozone concentrations?

- The ozone concentrations recorded by the satellite were so low they were being treated as outliers by a computer program and discarded!



Sources:
http://exploringdata.cqu.edu.au/ozone.html
http://www.epa.gov/ozone/science/hole/size.html

Real World Anomalies

- Credit Card Fraud
  - An abnormally high purchase made on a credit card

- Cyber Intrusions
  - A web server involved in *ftp* traffic

Fraud Detection

- Fraud detection refers to detection of criminal activities occurring in commercial organizations
  - Malicious users might be the actual customers of the organization or might be posing as a customer (also known as identity theft).
- Types of fraud
  - Credit card fraud
  - Insurance claim fraud
  - Mobile / cell phone fraud
  - Insider trading

- Challenges
  - Fast and accurate real-time detection
  - Misclassification cost is very high

Healthcare Informatics

- Detect anomalous patient records
  - Indicate disease outbreaks, instrumentation errors, etc.
- Key Challenges
  - Only normal labels available
  - Misclassification cost is very high
  - Data can be complex: spatio-temporal

outbreaks from 2006 to today preventable by vaccinations Article

Canada
Hudson Bay
AB
MB
NL
PE
NS
NH
MA
RI
DE
NJ
MD
MT
WY
NV
OK
MS
AL
SC
LA
Gulf of California
Mexico
Gulf of Mexico
Cuba
Nicaragua

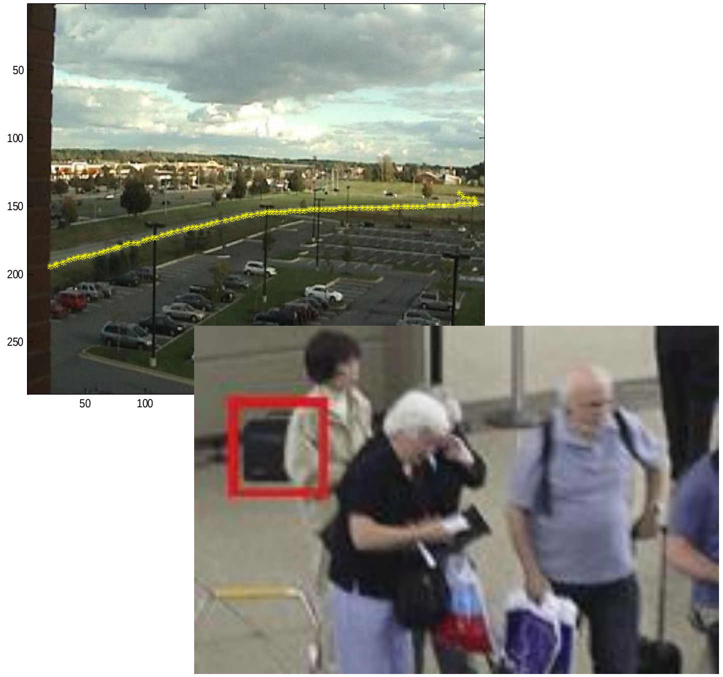
Industrial Damage Detection

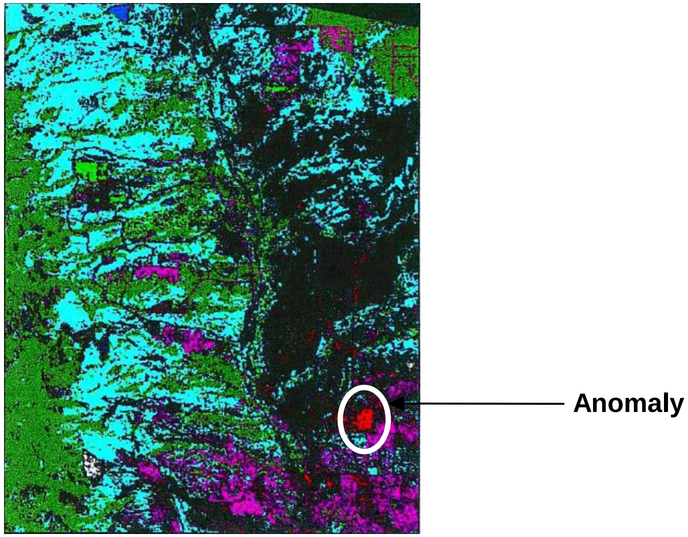
- Industrial damage detection refers to detection of different faults and failures in complex industrial systems, structural damages, intrusions in electronic security systems, suspicious events in video surveillance, abnormal energy consumption, etc.
  - Example: Aircraft Safety
    - Anomalies in engine combustion data

- Key Challenges
  - Data is extremely huge, noisy and unlabelled
  - Most of applications exhibit temporal behavior
  - Detecting anomalous events typically require immediate intervention

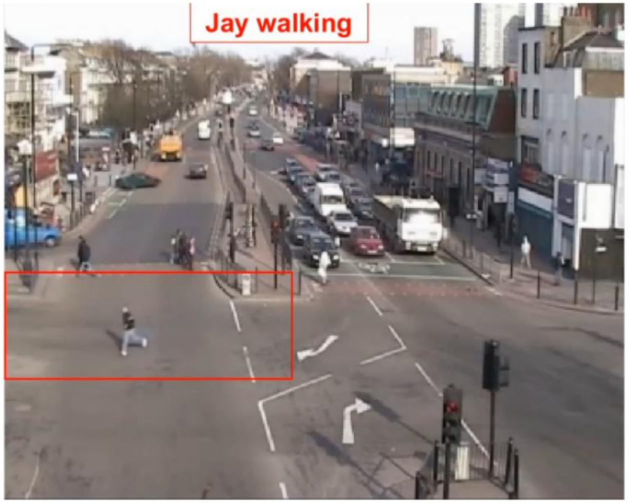
Image Processing

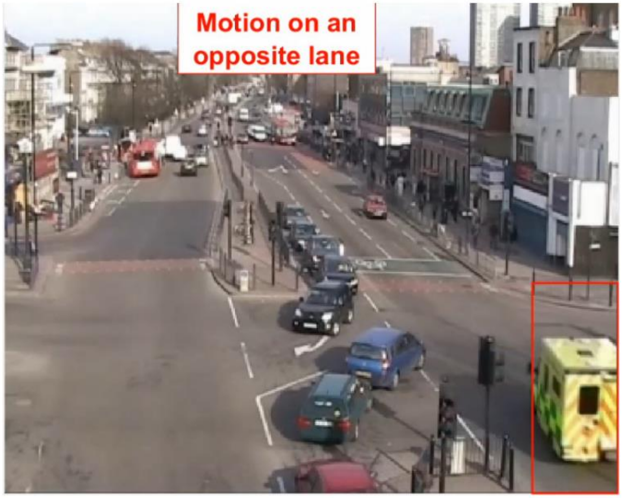
- Detecting outliers in a image monitored over time
- Detecting anomalous regions within an image
- Used in
  - mammography image analysis
  - video surveillance
  - satellite image analysis
- Key Challenges
  - Detecting collective anomalies
  - Data sets are very large

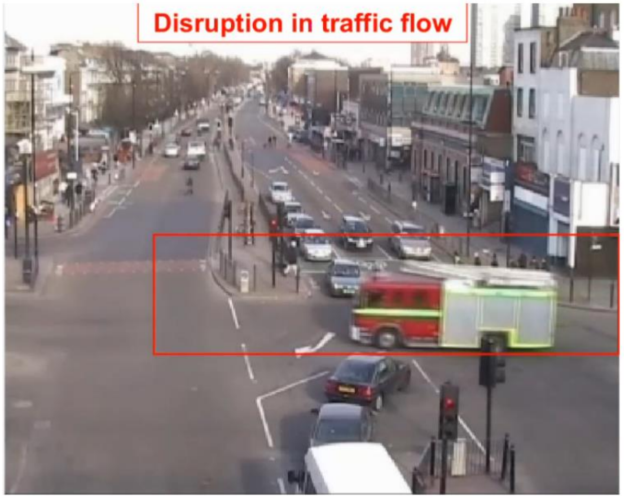




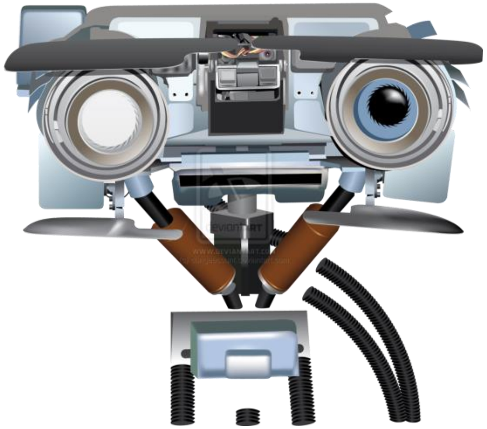
Video Surveillance







https://arxiv.org/pdf/1606.08455.pdf



Density Estimation Method

Anomaly Detection

Anomaly detection example

Aircraft engine features:

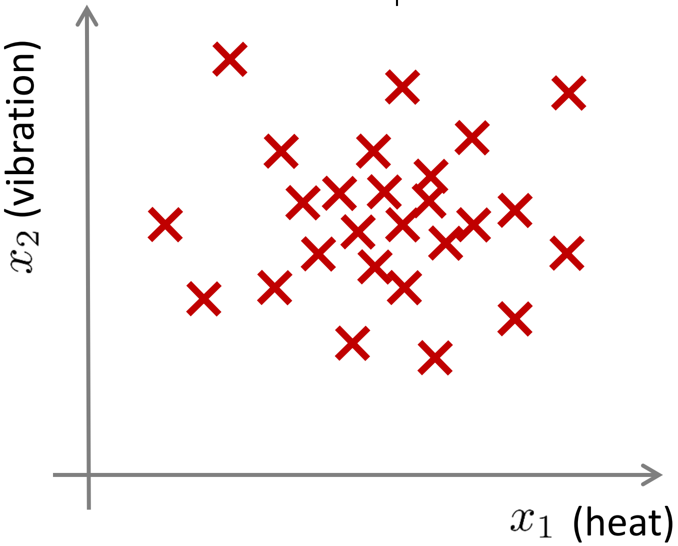
x_1 = heat generated

x_2 = vibration intensity

…

Dataset: $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

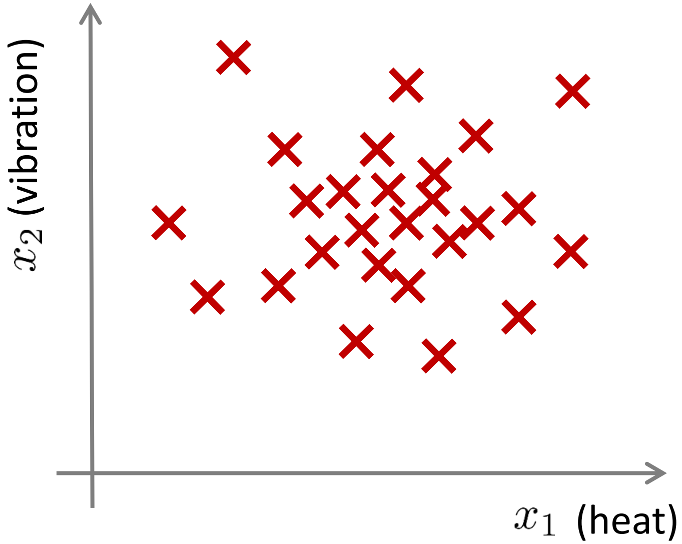
New engine: x_{test}



Density estimation

Dataset: $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

Is x_{test} anomalous?



Anomaly detection example

Fraud detection:

$x^{(i)}$ = features of user i 's activities

Model $p(x)$ from data.

Identify unusual users by checking which have $p(x) < \varepsilon$

Manufacturing:

Monitoring computers in a data center.

$x^{(i)}$ = features of machine i

x_1 = memory use, x_2 = number of disk accesses/sec,

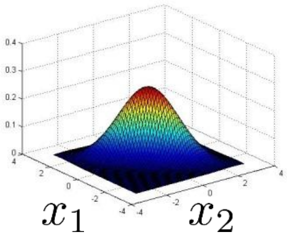
x_3 = CPU load, x_4 = CPU load/network traffic.

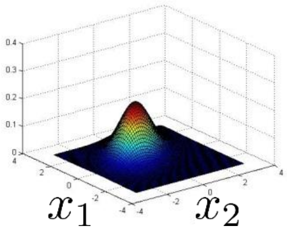
…

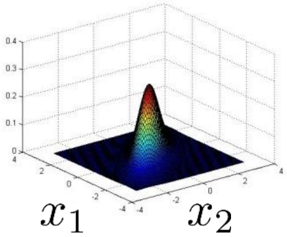
Example density estimation method: Multivariate Gaussian (Normal) distribution

Parameters μ, Σ

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$







Parameter fitting:

Given training set $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

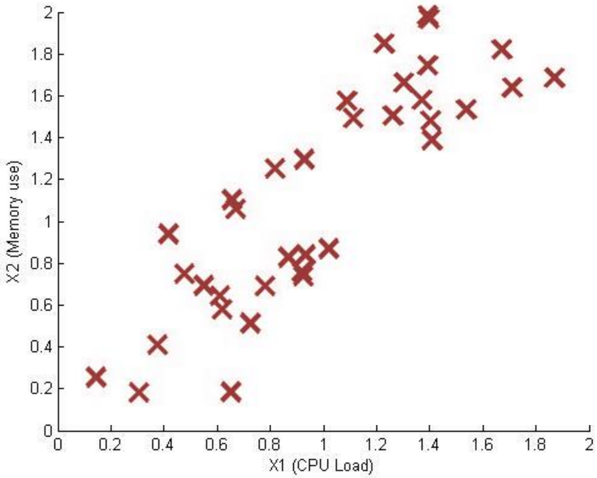
$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \qquad \Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu)(x^{(i)}-\mu)^T$$

Anomaly detection with the multivariate Gaussian

1. Fit model $p(x)$ by setting

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

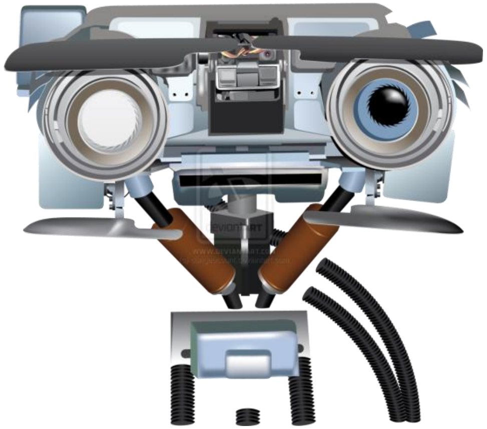
$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \mu)(x^{(i)} - \mu)^T$$



2. Given a new example x, compute

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Flag an anomaly if $p(x) < \varepsilon$



Evaluation

Anomaly Detection

Evaluating an anomaly detection model

When developing a learning algorithm (choosing features, etc.), making decisions is much easier if we have a way of evaluating our learning algorithm.

Assume we have some labeled data, of anomalous and non-anomalous examples. ($y = 0$ if normal, $y = 1$ if anomalous).

Training set: $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$ (assume normal examples/not anomalous)

Cross validation set: $(x_{cv}^{(1)}, y_{cv}^{(1)}), \ldots, (x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})})$
Test set: $(x_{test}^{(1)}, y_{test}^{(1)}), \ldots, (x_{test}^{(m_{test})}, y_{test}^{(m_{test})})$

Aircraft engines motivating example

10000	good (normal) engines
20	flawed engines (anomalous)

Training set: 6000 good engines
CV: 2000 good engines ($y = 0$), 10 anomalous ($y = 1$)
Test: 2000 good engines ($y = 0$), 10 anomalous ($y = 1$)

Alternative:
Training set: 6000 good engines
CV: 4000 good engines ($y = 0$), 10 anomalous ($y = 1$)
Test: 4000 good engines ($y = 0$), 10 anomalous ($y = 1$)

Algorithm evaluation

Fit model $p(x)$ on training set $\{x^{(1)}, \ldots, x^{(m)}\}$

On a cross validation/test example x, predict

$$y = \begin{cases} 1 & \text{if } p(x) < \varepsilon \text{ (anomaly)} \\ 0 & \text{if } p(x) \geq \varepsilon \text{ (normal)} \end{cases}$$

Possible evaluation metrics:

- Precision/Recall

$$precision = \frac{true\ positives}{predicted\ positives}$$

$$recall = \frac{true\ positives}{actual\ positives}$$

Can also use cross validation set to choose parameter ε

Anomaly detection *vs.* Supervised learning

- Very small number of positive examples (y=1)
- Large number of negative (y=0) examples

- Many different “types” of anomalies. Hard for any algorithm to learn from positive examples what the anomalies look like; future anomalies may look nothing like any of the anomalous examples we’ve seen so far.

- Large number of positive and negative examples.

- Enough positive examples for algorithm to get a sense of what positive examples are like, future positive examples likely to be similar to ones in training set.

Anomaly detection vs. Supervised learning

- Fraud detection
- Manufacturing (e.g. aircraft engines)
- Monitoring machines in a data center

⋮

- Email spam classification
- Weather prediction (sunny/rainy/etc).
- Cancer classification

⋮

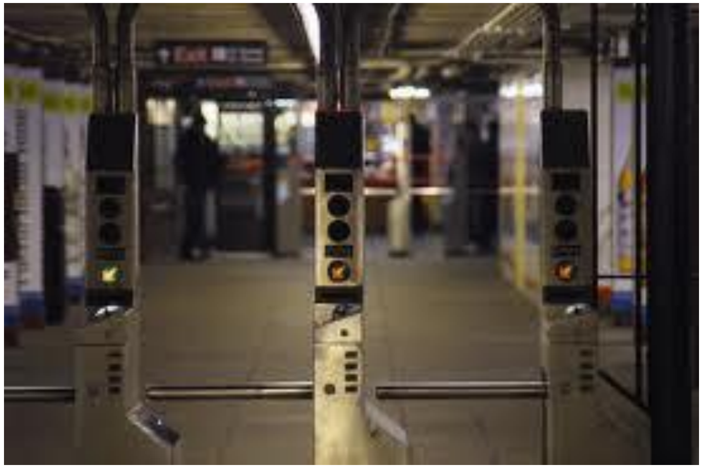
Online Detection of Unusual Events in Videos via Dynamic Sparse Coding

Bin Zhao, Li Fei-Fei, Eric Xing

Proceedings of the International Conference on Computer Vision and Pattern Recognition
(*CVPR 2011*), Colorado Springs, CO, USA, June 2011

Goal: Detect Unusual Events in Videos

- Example unusual event: entering subway via exit
- Videos are described as spatio-temporal features



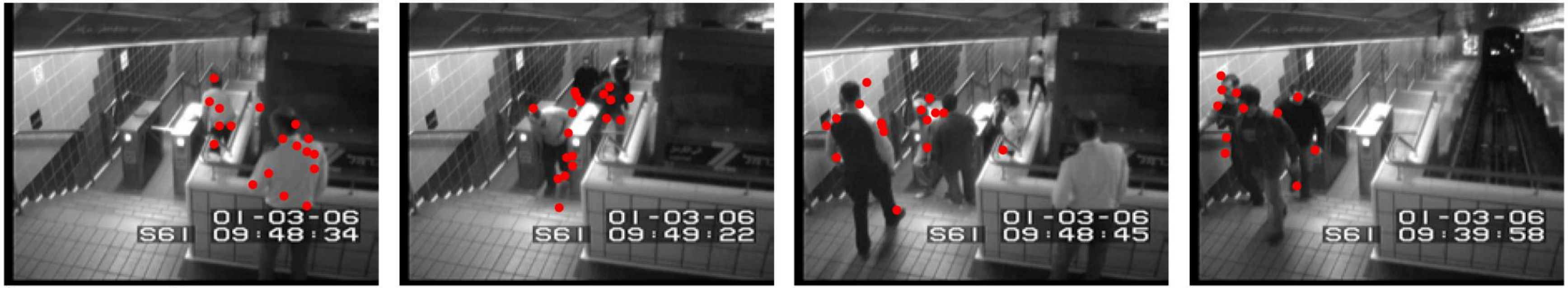


Figure 2. Example spatio-temporal interest points

Dictionary-based Anomaly Detection

- Learn a dictionary of bases corresponding to usual events:
  - a usual event should be reconstructible from a small number of such bases, and
  - the reconstruction weights should change smoothly over space/time across actions in such events.
  - an unusual event is either not reconstructible from the dictionary of usual events with small error, or,
  - Needs a large number of bases, in a temporal-spatially non-smooth fashion.
- Must: Learn a good dictionary of bases representing usual events
- Must: Update the dictionary online to adapt to changing content of the video

Algorithm: Look for High Reconstruction Error

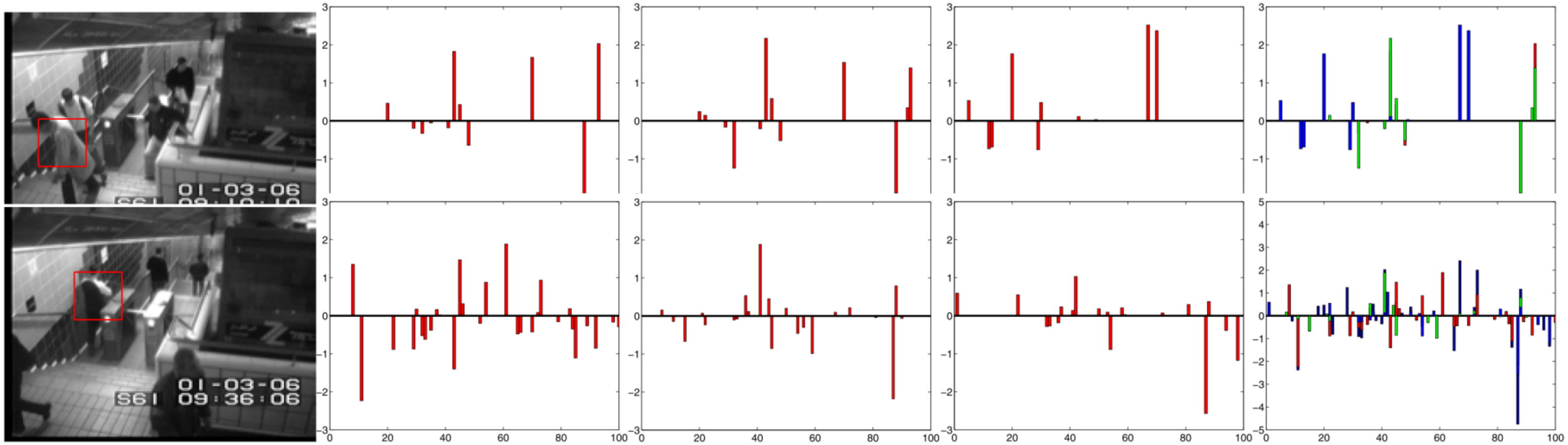
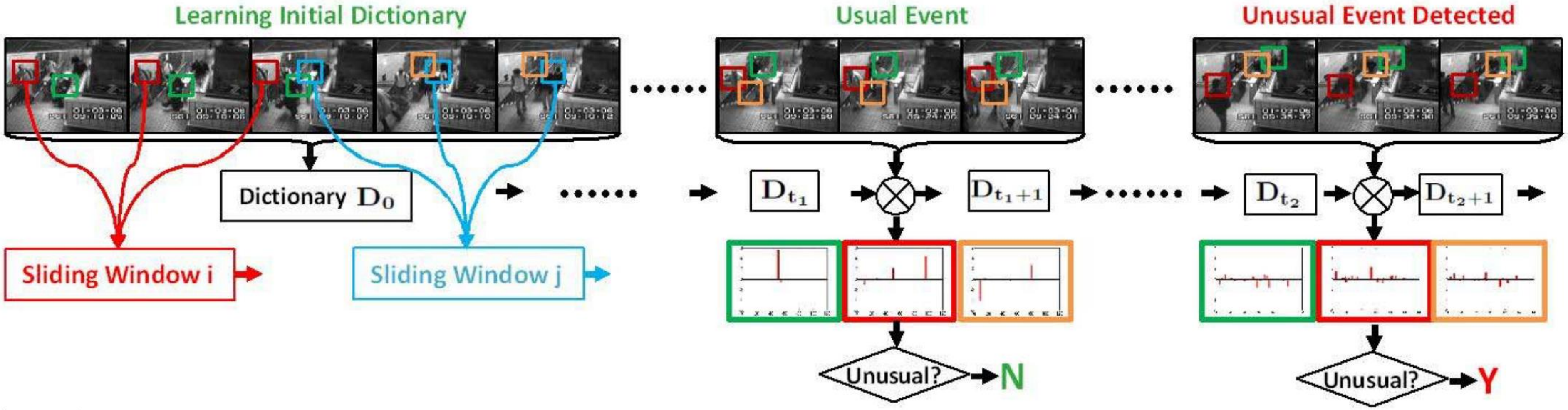


Figure 3. First row: usual event (leaving subway exit); second row: unusual event (entering subway exit). From left to right: example frame and sliding window, reconstruction vectors for 3 cuboids, plot all 3 reconstruction vectors on the same figure.

Algorithm


Learning Initial Dictionary
Usual Event
Unusual Event Detected
Dictionary D_0
Sliding Window i
Sliding Window j
D_{t_1}
D_{t_1+1}
D_{t_2}
D_{t_2+1}
Unusual?
N
Unusual?
Y