Outlier Detection
Mathematical Notes

Global, Contextual, and Collective Outliers; Statistical, Proximity, Reconstruction, Clustering, Classification, and High-Dimensional Methods

Central idea. An **outlier** is an object that deviates strongly from the pattern of normal data, possibly because it was generated by a different mechanism. Detection identifies unusual observations; domain investigation determines whether they are fraud, failure, a valid rare event, noise, or a new population.

I. Foundations of Outlier Detection

1. What Is an Outlier?

Let $D = \{x_1, x_2, \ldots, x_n\}$ be a data set and let $s(x)$ be an **outlier score**. A larger score means that x appears more unusual with respect to a chosen reference population. An alert is created only after choosing a threshold τ:

$$\boxed{\text{Flag } x \text{ as an outlier if } s(x) \geq \tau.}$$

Outlierness is therefore relative to four choices: the representation of x, the reference data considered normal, the context, and time.

2. Outlier versus Noise

	Outlier	Noise
Meaning	A meaningful object that strongly deviates from a pattern	Random error or unwanted variation
Example	A real high-value transaction from a compromised account	A sensor spike caused by electrical interference
Goal	Detect, rank, explain, and investigate	Remove, smooth, correct, or model robustly
Risk	Deleting it may remove the event of interest	Keeping it may distort the learned pattern

Important warning. A valid extreme value is not automatically an error, and a value inside the normal range is not automatically normal.

3. Three Types of Outlier

- **Global outlier:** one object is unusual compared with the complete data set. Example: a transaction amount of Tk. 50,000 when nearly all amounts are below Tk. 5,000.
- **Contextual outlier:** one object is unusual only under a specific context. Example: 30°C is normal in July but abnormal in January.
- **Collective outlier:** a group, window, sequence, or subgraph is unusual even though its individual members may look ordinary. Example: many login attempts in a very short time.

Global	Contextual	Collective
$s(x)$ uses D	$s(x)$ uses D_c	$s(W)$ uses a set or sequence

4. Learning Settings

Setting	Available training data	Main issue
Supervised	Labeled normal and outlier objects	Outliers are rare, diverse, and expensive to label
Semi-supervised	Mostly verified normal objects	Contamination in normal training data weakens the boundary
Unsupervised	No reliable labels	The method's assumptions define what is normal

II. Statistical Approaches

Statistical methods assume that normal data follow a probability model. An observation receiving a very small probability under the fitted model is considered unusual.

1. Negative Log-Likelihood Score

Suppose that normal data follow the probability distribution

$$x \sim p(x \mid \theta).$$

A common outlier score is

$$\boxed{s(x) = -\log p(x \mid \hat{\theta})}.$$

A low probability produces a large negative log-likelihood score.

Worked Example: Probability-Based Detection

A fitted model assigns the following probabilities to four observations.

Object	$p(x \mid \hat{\theta})$	$s(x) = -\log p(x \mid \hat{\theta})$
x_1	0.30	1.204
x_2	0.25	1.386
x_3	0.20	1.609
x_4	0.01	4.605

For x_4,

$$\begin{aligned} s(x_4) &= -\log(0.01) \\ &= -(-4.605) \\ &= 4.605. \end{aligned}$$

If the alert threshold is $\tau = 3$, then

$$s(x_4) = 4.605 > 3.$$

Therefore,

$$\boxed{x_4 \text{ is flagged as an outlier.}}$$

Interpretation. The fitted model assigns only a 1% probability to x_4. Therefore, x_4 receives the largest outlier score.

2. Z-Score Detection

For a univariate attribute with mean μ and standard deviation σ, the z-score is

$$\boxed{z = \frac{x - \mu}{\sigma}}.$$

The absolute value $|z|$ measures how many standard deviations the observation lies from the mean.

Worked Example: Response-Time Detection

The following response times are considered normal.

Request	**Response Time** x_i	$(x_i - \mu)^2$
x_1	94	36
x_2	97	9
x_3	99	1
x_4	100	0
x_5	102	4
x_6	108	64
Total	600	114

The population mean is

$$\mu = \frac{94 + 97 + 99 + 100 + 102 + 108}{6} = \frac{600}{6} = 100.$$

The population variance is

$$\sigma^2 = \frac{\sum_{i=1}^{6}(x_i - \mu)^2}{6} = \frac{114}{6} = 19.$$

Therefore,

$$\sigma = \sqrt{19} \approx 4.359.$$

Suppose a new request has response time $x = 116$. Its z-score is

$$\begin{aligned} z &= \frac{x-\mu}{\sigma} \\ &= \frac{116-100}{4.359} \\ &\approx 3.67. \end{aligned}$$

Using $|z| > 3$ as the decision rule,

$$\boxed{116 \text{ is an outlier because } |z| = 3.67 > 3.}$$

3. IQR-Based Detection

For sorted data, the interquartile range is

$$\text{IQR} = Q_3 - Q_1.$$

The lower and upper fences are

$$\boxed{L = Q_1 - 1.5\,\text{IQR}, \qquad U = Q_3 + 1.5\,\text{IQR}.}$$

An observation is flagged if it is below L or above U.

Worked Example: IQR Fences

Consider the sorted data

$$D = \{4, 5, 5, 6, 6, 7, 8, 20\}.$$

Position	1	2	3	4	5	6	7	8
Value	4	5	5	6	6	7	8	20

The lower half is

$$\{4, 5, 5, 6\},$$

so

$$Q_1 = \frac{5+5}{2} = 5.$$

The upper half is

$$\{6, 7, 8, 20\},$$

so

$$Q_3 = \frac{7+8}{2} = 7.5.$$

Therefore,

$$\text{IQR} = 7.5 - 5 = 2.5.$$

The lower fence is

$$L = 5 - 1.5(2.5) = 1.25.$$

The upper fence is

$$U = 7.5 + 1.5(2.5) = 11.25.$$

Since

$$20 > 11.25,$$

we obtain

$$\boxed{\text{20 is an outlier under the IQR rule.}}$$

4. Robust Median and MAD

The median and median absolute deviation are

$$\tilde{x} = \text{median}(x_i),$$

and

$$\text{MAD} = \text{median}\,|x_i - \tilde{x}|.$$

The robust z-score is

$$\boxed{z_i^{(r)} = 0.6745\frac{x_i - \tilde{x}}{\text{MAD}}}.$$

Worked Example: Robust Z-Score

Consider

$$D = \{4, 5, 5, 6, 6, 7, 8, 20\}.$$

The complete deviation table is

x_i	$x_i - \tilde{x}$	$\lvert x_i - \tilde{x}\rvert$
4	-2	2
5	-1	1
5	-1	1
6	0	0
6	0	0
7	1	1
8	2	2
20	14	14

The median is

$$\tilde{x} = \frac{6+6}{2} = 6.$$

The sorted absolute deviations are

$$\{0, 0, 1, 1, 1, 2, 2, 14\}.$$

Therefore,

$$\text{MAD} = \frac{1+1}{2} = 1.$$

For $x = 20$,

$$\begin{aligned} z^{(r)} &= 0.6745\frac{20-6}{1} \\ &= 0.6745(14) \\ &= 9.443. \end{aligned}$$

Using $|z^{(r)}| > 3.5$ as the threshold,

$$\boxed{\text{20 is a strong robust-statistical outlier.}}$$

5. Mahalanobis Distance

For multivariate data with mean vector μ and covariance matrix Σ,

$$\boxed{D_M^2(x) = (x-\mu)^\top \Sigma^{-1}(x-\mu)}.$$

Mahalanobis distance accounts for both scale and correlation.

Worked Example: Two-Dimensional Data

Consider the following reference data.

Object	x_1	x_2
A	-2	-1
B	-2	1
C	0	-1
D	0	1

The mean vector is

$$\mu = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

The sample covariance matrix is

$$\Sigma = \begin{bmatrix} \frac{4}{3} & 0 \\ 0 & \frac{4}{3} \end{bmatrix}.$$

Therefore,

$$\Sigma^{-1} = \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{4} \end{bmatrix}.$$

Suppose the test point is

$$x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Then

$$x - \mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} D_M^2(x) &= \begin{bmatrix} 3 & 3 \end{bmatrix} \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{4} \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} \\ &= \frac{3}{4}(3^2) + \frac{3}{4}(3^2) \\ &= 13.5. \end{aligned}$$

Thus,

$$D_M(x) = \sqrt{13.5} \approx 3.674.$$

For two dimensions,

$$\chi^2_{2,0.95} = 5.99.$$

Since

$$D_M^2(x) = 13.5 > 5.99,$$

we conclude that

$$\boxed{x = (2, 3) \text{ is a multivariate outlier.}}$$

6. Kernel Density Estimation

Kernel density estimation is

$$\boxed{\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right).}$$

A common anomaly score is

$$s(x) = -\log \hat{f}_h(x).$$

For a Gaussian kernel,

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

Worked Example: Gaussian KDE

Consider

$$D = \{1, 2, 2.5, 3\}, \qquad h = 1,$$

and calculate the density of $x = 6$.

x_i	$u_i = (6 - x_i)/1$	$K(u_i)$	**Contribution**
1	5.0	0.00000149	0.00000149
2	4.0	0.00013383	0.00013383
2.5	3.5	0.00087268	0.00087268
3	3.0	0.00443185	0.00443185

Therefore,

$$\begin{aligned} \hat{f}_1(6) &= \frac{1}{4}\,(0.00000149 + 0.00013383 + 0.00087268 + 0.00443185) \\ &= 0.001360. \end{aligned}$$

The anomaly score is

$$s(6) = -\log(0.001360) \approx 6.60.$$

$$\boxed{\hat{f}_1(6) \approx 0.001360, \qquad s(6) \approx 6.60}$$

The estimated density is extremely small, so $x = 6$ is a strong outlier candidate.

III. Proximity-Based Approaches

Proximity methods detect observations that are far from other objects or have lower local density than their neighbors.

1. Minkowski Distance

The Minkowski distance is

$$\boxed{d_p(x, y) = \left(\sum_{j=1}^{d} |x_j - y_j|^p \right)^{1/p}}.$$

For $p = 1$, it becomes Manhattan distance. For $p = 2$, it becomes Euclidean distance.

Worked Example: Manhattan and Euclidean Distance

Consider

$$x = (2, 3), \qquad y = (5, 7).$$

Feature	x_j	y_j	$\lvert x_j - y_j \rvert$
1	2	5	3
2	3	7	4

The Manhattan distance is

$$d_1(x, y) = |2 - 5| + |3 - 7| = 3 + 4 = 7.$$

The Euclidean distance is

$$\begin{aligned} d_2(x, y) &= \sqrt{(2 - 5)^2 + (3 - 7)^2} \\ &= \sqrt{9 + 16} \\ &= 5. \end{aligned}$$

Therefore,

$$\boxed{d_1(x, y) = 7, \qquad d_2(x, y) = 5.}$$

2. Distance-Based $DB(p, D)$ Outlier

An object o is a $DB(p, D)$ outlier if at least fraction p of all objects are farther than distance D from it:

$$\boxed{\frac{|\{x \in D : d(x, o) > D\}|}{|D|} \geq p}.$$

Worked Example: Radius-Based Detection

Let

$$D = \{1, 2, 2.5, 3, 10\}, \qquad o = 10, \qquad p = 0.60, \qquad D_{\text{radius}} = 3.$$

Object	**Distance from 10**	**Distance > 3?**
1	9.0	Yes
2	8.0	Yes
2.5	7.5	Yes
3	7.0	Yes
10	0.0	No

Four of the five objects are farther than three units from 10:

$$\frac{4}{5} = 0.80.$$

Since

$$0.80 \geq 0.60,$$

we conclude that

$$\boxed{10 \text{ is a } DB(0.60, 3) \text{ outlier.}}$$

3. k-Nearest-Neighbor Score

Two common k-NN outlier scores are

$$s_{\max}(x) = d\big(x, k\text{-th NN}(x)\big),$$

and

$$\boxed{s_{\text{avg}}(x) = \frac{1}{k} \sum_{o \in N_k(x)} d(x, o)}.$$

Worked Example: Average Two-NN Distance

Let

$$D = \{1, 2, 2.5, 3, 10\}, \qquad k = 2.$$

The complete pairwise distance table is

	1	2	2.5	3	10
1	0	1	1.5	2	9
2	1	0	0.5	1	8
2.5	1.5	0.5	0	0.5	7.5
3	2	1	0.5	0	7
10	9	8	7.5	7	0

The average two-neighbor scores are

Point	**Two Nearest Distances**	$s_{\text{avg}}(x)$
1	1, 1.5	1.25
2	0.5, 1	0.75
2.5	0.5, 0.5	0.50
3	0.5, 1	0.75
10	7, 7.5	7.25

For $x = 10$,

$$s_{\text{avg}}(10) = \frac{7 + 7.5}{2} = 7.25.$$

Therefore,

$$\boxed{10 \text{ has the largest } k\text{-NN outlier score.}}$$

4. Local Outlier Factor

The reachability distance is

$$\text{reach-dist}_k(p, o) = \max\{k\text{-distance}(o), d(p, o)\}.$$

The local reachability density is

$$\text{lrd}_k(p) = \left(\frac{\sum_{o \in N_k(p)} \text{reach-dist}_k(p, o)}{|N_k(p)|} \right)^{-1}.$$

The LOF score is

$$\boxed{\text{LOF}_k(p) = \frac{1}{|N_k(p)|} \sum_{o \in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)}}.$$

Worked Example: Local Density Comparison

Suppose

$$\text{lrd}_k(p) = 0.25,$$

and the local densities of its three nearest neighbors are

$$0.50, \qquad 0.40, \qquad 0.60.$$

Object	**Local Density**	**Ratio to Density of p**
p	0.25	1.0
o_1	0.50	$0.50/0.25 = 2.0$
o_2	0.40	$0.40/0.25 = 1.6$
o_3	0.60	$0.60/0.25 = 2.4$

Therefore,

$$\begin{aligned} \text{LOF}_3(p) &= \frac{1}{3} \left(\frac{0.50}{0.25} + \frac{0.40}{0.25} + \frac{0.60}{0.25} \right) \\ &= \frac{1}{3}(2.0 + 1.6 + 2.4) \\ &= 2.0. \end{aligned}$$

$$\boxed{\text{LOF}_3(p) = 2.0}$$

The neighbors are, on average, twice as dense as the region containing p. Therefore, p is a strong local-outlier candidate.

IV. Reconstruction Methods

Reconstruction methods learn a compact representation of normal data. Observations that produce large reconstruction errors are considered unusual.

1. PCA Reconstruction Error

If U contains orthonormal principal directions, then

$$\hat{x} = UU^\top x.$$

The squared reconstruction score is

$$\boxed{s(x) = \|x - \hat{x}\|_2^2}.$$

Worked Example: Distance from a PCA Subspace

Normal observations lie along the line $x_1 = x_2$.

Object	x_1	x_2	Type
A	1	1	Reference
B	2	2	Reference
C	3	3	Reference
D	4	4	Reference
x	4	0	Test point

The unit direction of the normal subspace is

$$u = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Therefore,

$$uu^\top = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

The projection of $x = (4, 0)^\top$ is

$$\begin{aligned} \hat{x} &= uu^\top x \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 2 \end{bmatrix}. \end{aligned}$$

The residual is

$$e = x - \hat{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}.$$

Thus,

$$s(x) = 2^2 + (-2)^2 = 8.$$

$$\boxed{s(x) = 8}$$

The test point has a large distance from the learned normal subspace.

2. Robust PCA

Robust PCA decomposes the data matrix as

$$\boxed{X = L + S},$$

where L is a low-rank normal matrix and S is a sparse anomaly matrix.
The optimization problem is

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \qquad \text{subject to } X = L + S.$$

Worked Example: Sparse Anomalous Entry

Suppose

$$X = \begin{bmatrix} 1 & 2 & 9 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

The learned low-rank structure is

$$L = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} S &= X - L \\ &= \begin{bmatrix} 1 & 2 & 9 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The complete entry comparison is

Position	X_{ij}	L_{ij}	S_{ij}
$(1,1)$	1	1	0
$(1,2)$	2	2	0
$(1,3)$	9	3	6
$(2,1)$	2	2	0
$(2,2)$	4	4	0
$(2,3)$	6	6	0
$(3,1)$	3	3	0
$(3,2)$	6	6	0
$(3,3)$	9	9	0

The only large sparse value is

$$\boxed{S_{13} = 6}.$$

Therefore, the observation at position $(1,3)$ is anomalous.

3. Autoencoder Reconstruction

An autoencoder calculates

$$z = f_\theta(x), \qquad \hat{x} = g_\phi(z).$$

The anomaly score is

$$\boxed{s(x) = \|x - \hat{x}\|_2^2}.$$

Worked Example: Autoencoder Residual

An autoencoder produces the following reconstructions.

Object	**Original** x_1	x_2	**Reconstructed** $\hat{x}_1$	$\hat{x}_2$
A	1.0	1.0	1.1	0.9
B	2.0	2.0	1.9	2.1
C	5.0	1.0	3.0	2.0

For object A,

$$s(A) = (1.0 - 1.1)^2 + (1.0 - 0.9)^2 = 0.02.$$

For object B,

$$s(B) = (2.0 - 1.9)^2 + (2.0 - 2.1)^2 = 0.02.$$

For object C,

$$s(C) = (5.0 - 3.0)^2 + (1.0 - 2.0)^2 = 4 + 1 = 5.$$

Object	Reconstruction Score	Decision
A	0.02	Normal
B	0.02	Normal
C	5.00	Outlier

Therefore,

$$\boxed{C \text{ is the strongest autoencoder-based outlier.}}$$

V. Clustering and Classification Approaches

1. Clustering-Based Scores

Normal objects are often assumed to belong to a cluster and lie near its center. For cluster centroid $\mu_{c(i)}$,

$$\boxed{s(x_i) = d(x_i, \mu_{c(i)}).}$$

A cluster-level score can combine small size and low density:

$$s(C_j) = \alpha \frac{1}{|C_j|} + (1-\alpha) \frac{1}{\text{density}(C_j)}.$$

Worked example. With $\mu = (2, 3)$ and $x_3 = (7, 8)$,

$$s(x_3) = \sqrt{(7-2)^2 + (8-3)^2} = \sqrt{50} = 7.07.$$

Small clusters are not automatically anomalous: they may be valid minority populations, new cohorts, or coordinated attacks.

2. Classification-Based Scores

Two-class classification learns from labeled normal and outlier cases:

$$s(x) = P(y = \text{outlier} \mid x).$$

One-class classification learns a boundary around normal data. A standard one-class SVM formulation is

$$\min_{w,\rho,\xi} \ \frac{1}{2} \|w\|^2 + \frac{1}{\nu n} \sum_{i=1}^{n} \xi_i - \rho$$

subject to

$$w^\top \phi(x_i) \geq \rho - \xi_i, \qquad \xi_i \geq 0,$$

with decision function

$$f(x) = \text{sign}\big(w^\top \phi(x) - \rho\big).$$

The parameter ν controls the tolerated violations. Scaling and kernel width are important, especially with an RBF kernel.

VI. Contextual and Collective Outliers

1. Contextual Detection

Split an object into contextual attributes c and behavioral attributes b:

$$x = (c, b), \qquad \boxed{s(x) = -\log p(b \mid c).}$$

Two common strategies are:

1. Build a separate reference distribution per context:

$$z_{\text{context}} = \frac{x - \mu_{\text{context}}}{\sigma_{\text{context}}}.$$

2. Predict behavior from context and score its residual:

$$\hat{b} = f(c), \qquad e = b - \hat{b}, \qquad s = |e|.$$

Worked example. For January, $\mu_{\text{Jan}} = 14°\text{C}$, $\sigma_{\text{Jan}} = 3°\text{C}$, and $x = 30°\text{C}$:

$$z_{\text{Jan}} = \frac{30 - 14}{3} = 5.33.$$

For July, $\mu_{\text{Jul}} = 31°\text{C}$ and $\sigma_{\text{Jul}} = 2°\text{C}$:

$$z_{\text{Jul}} = \frac{30 - 31}{2} = -0.5.$$

The same temperature is anomalous in January but normal in July.

2. Collective Detection

For a sequence window $W_t = (x_{t-w+1}, \ldots, x_t)$, a likelihood score is

$$\boxed{s(W_t) = -\sum_{j=t-w+1}^{t} \log p(x_j \mid x_{j-1}, \ldots).}$$

Other collective evidence includes unusual subsequence shapes, rate bursts, unexpected transitions, coordinated groups, and suspicious graph substructures.

Worked example: login burst. Suppose the normal number of logins per minute follows a Poisson distribution with $\lambda = 2$:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

For ten or more attempts in one minute,

$$P(X \geq 10) = 1 - \sum_{k=0}^{9} e^{-2} \frac{2^k}{k!} \approx 4.65 \times 10^{-5}.$$

Each attempt can look ordinary, but the burst is highly improbable and is therefore a collective outlier.

VII. High-Dimensional Outliers

In high-dimensional data, pairwise distances become similar. For independent standardized features, an approximate relative spread is

$$\boxed{\mathrm{CV}(d) \approx \frac{1}{\sqrt{2d}}.}$$

As d increases, nearest and farthest distances become less distinguishable. An anomaly may be visible in only a small informative subspace and hidden by many irrelevant features.

Useful strategies are: regularized or robust conventional detectors, subspace search, projections/factorization, learned representations, and feature bagging. For feature subsets S_m,

$$\boxed{s(x) = \mathrm{aggregate}_{m=1}^{M}\, s_m(x_{S_m}).}$$

Angle-based methods use

$$\theta_{axb} = \cos^{-1}\left(\frac{(a-x)^\top (b-x)}{\|a-x\|\,\|b-x\|}\right).$$

A far-away point sees many other points in a similar direction and tends to have low angle variance.

Reduction caution. Do not retain components only because they explain high variance. A low-variance direction may contain the exact signal that makes an object anomalous. Fit scaling and dimensionality reduction only on the training or reference data.

VIII. Thresholds, Evaluation, and Practice

1. Choosing a Threshold

Possible threshold policies include a clean-score quantile, a target false-positive rate, daily review capacity, or expected cost. Monitor score distributions, alerts per day, review outcomes, and feature/population drift after deployment.

2. Evaluation Metrics

With rare anomalies, accuracy is misleading. Let TP, FP, and FN be true positives, false positives, and false negatives:

$$\boxed{\mathrm{Precision} = \frac{TP}{TP+FP}, \qquad \mathrm{Recall} = \frac{TP}{TP+FN}.}$$

Precision-recall curves, Precision@k, recall, delay, and review burden are usually more informative than accuracy for rare-event detection.

If a missed anomaly costs C_{FN} and an unnecessary review costs C_{FP},

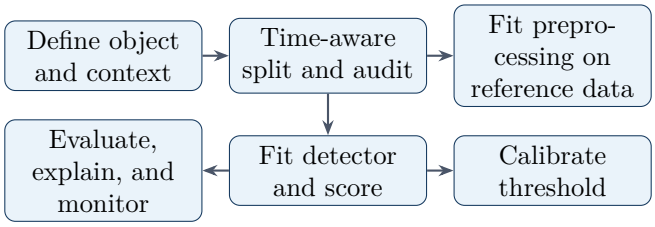
$$\boxed{\mathrm{ExpectedCost}(\tau) = C_{FN}FN(\tau) + C_{FP}FP(\tau).}$$

For $C_{FN} = 1000$ and $C_{FP} = 10$, compare:

$$\tau_1:\ 8(1000) + 20(10) = 8200, \qquad \tau_2:\ 2(1000) + 100(10) = 3000.$$

Even though τ_2 creates more false alerts, it is better under these costs.

3. Leakage-Safe Workflow



Common leakage sources are full-data scaling, future statistics, post-incident attributes, and the same entity appearing in both training and test data.

4. Method Selection Guide

Data condition	Strong starting method
One interpretable variable	IQR or robust z-score
Elliptical correlated continuous data	Robust Mahalanobis distance
Unequal local density	LOF
Low-rank numerical structure	PCA or robust factorization
Several meaningful groups	Clustering-based score
Reliable anomaly labels	Supervised classifier
Trusted normal reference	One-class model

5. Practice Problems with Answers

1. For $\{4, 5, 5, 6, 6, 7, 8, 20\}$, calculate IQR fences and identify an outlier.
2. For $\mu = (0, 0)$, $\Sigma = \mathrm{diag}(4, 1)$, and $x = (4, 2)$, calculate D_M^2.
3. For $\{1, 2, 3, 4, 12\}$ and $k = 2$, calculate the average k-NN score for every point.
4. A point has lrd $= 0.4$ and neighbor densities $0.8, 0.6, 1.0$. Calculate LOF.
5. A PCA model reconstructs $x = (5, 1)$ as $\hat{x} = (3, 2)$. Calculate squared reconstruction error.

Answers.

1. $Q_1 = 5$, $Q_3 = 7.5$, IQR $= 2.5$; fences are 1.25 and 11.25, so 20 is flagged.
2. $D_M^2 = 4^2/4 + 2^2/1 = 8$.
3. The scores are $1.5, 1, 1, 1.5, 8.5$, respectively.
4. LOF $= \frac{1}{3}(0.8/0.4 + 0.6/0.4 + 1.0/0.4) = 2.0$.
5. $\|x - \hat{x}\|_2^2 = (5 - 3)^2 + (1 - 2)^2 = 5$.

IX. Key Takeaways

1. Outlierness depends on representation, reference population, context, and time.
2. Each detector encodes an assumption: improbability, distance, local density, reconstruction error, cluster membership, or class boundary.
3. Global, contextual, and collective outliers require different units of analysis.
4. High-dimensional data need informative subspaces or structure-aware representations.

5. Scores rank unusualness; thresholds turn scores into operational decisions.
6. Evaluation must account for rarity, review capacity, false-alert cost, missed-event cost, and time-based deployment conditions.

Prepared By
Md. Atikuzzaman
Lecturer, Department of Computer Science and Engineering
Green University of Bangladesh
Email: atik@cse.green.edu.bd