Probabilistic deep learning model for contextual anomaly detection

Lovre Mrčela, mag. ing.

(lovre.mrcela@fer.hr)

mentor: Prof. Zvonko Kostanjčar, PhD

University of Zagreb Faculty of Electrical Engineering and Computing

1. Introduction

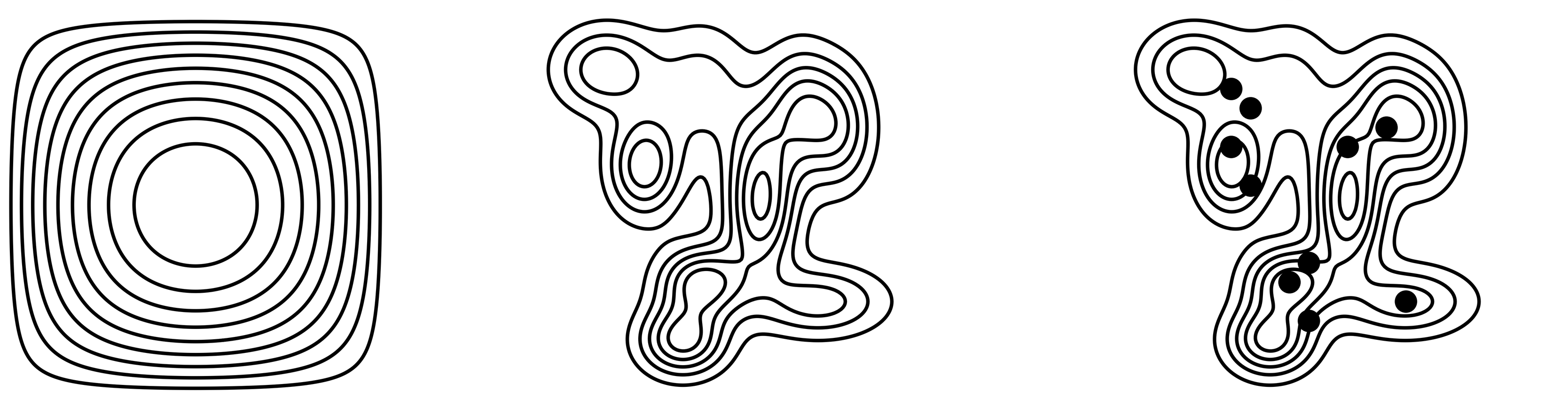
Anomaly detection is a **machine learning** discipline which is concerned with finding data points that **do not fit into the established pattern.**

The **probabilistic approach** to anomaly detection aims to find a **distribution** which **best fits the given data set.** The underlying assumption of this approach is that **anomalies are the data points whose probabilities are extremely small.**

A **latent variable model (LVM)** is a probabilistic model that represents the relation between the **latent variables z** and **data x** as

$$p(x) = \int_{\mathcal{Z}} p(x \mid z)\, p(z)\, dz$$

The **latent distribution p(z)** is decoded into **the data distribution p(x)** where we can look for anomalies.



2. Context

Context refers to the known situation in which a data point occurs. There are two factors which determine the context:

1. **co-occurring data points**
2. **additional contextual information**

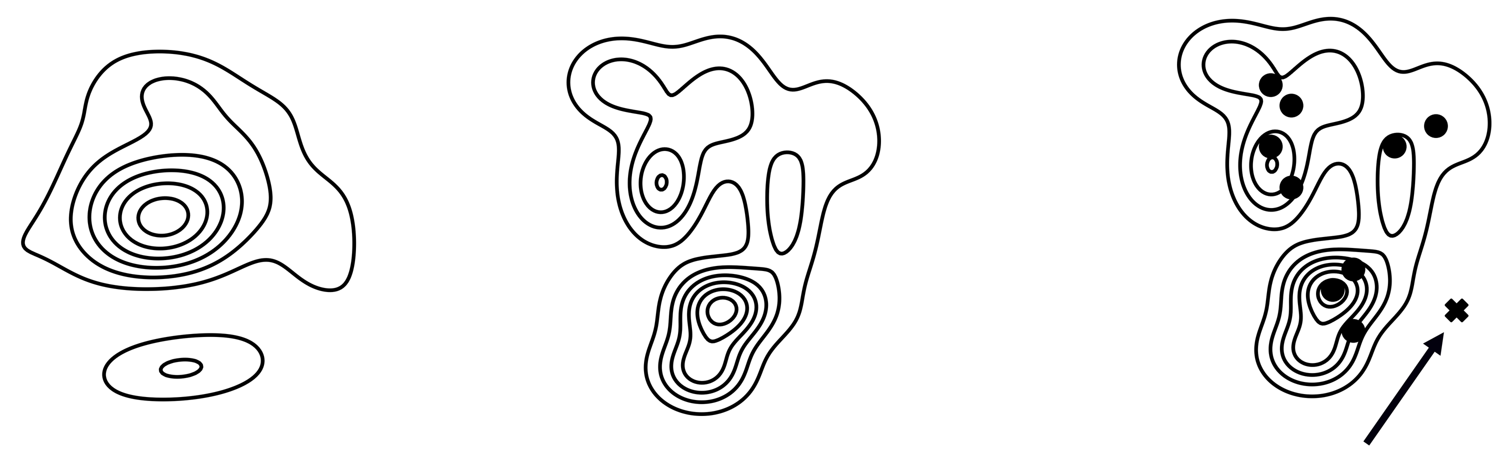
Contextual anomaly is a special kind of an anomaly where irregularity arises from the interaction between a data point and its context. Such data points may be normal in some contexts, but anomalies in others.

3. Model

The **proposed model** for contextual anomaly detection introduces the **contextual variable c** and extends the **LVM** so that the **contextual data distribution** can be recovered by

$$p(x \mid c) = \int_{\mathcal{Z}} p(x \mid z)\, p(z \mid c)\, dz$$

By allowing the latent distribution to adapt to a given context, model can detect contextual anomalies.



Whether something is an **anomaly** depends on the **context!**

4. Multivariate Polynomial PDF

The choice of family of **probability density functions (PDFs)** for the contextual latent distribution is crucial for the optimization, as allowing **too much freedom** can easily lead to **overfitting.**

To describe contextual latent distributions, we construct a family of PDFs based on **multivariate polynomial sum of squares (MPSOS).** The hyperparameters of this family are:

- the maximal degree of the polynomial square, 2D
- the number of variables, N
  (which corresponds to the dimension of the latent space)

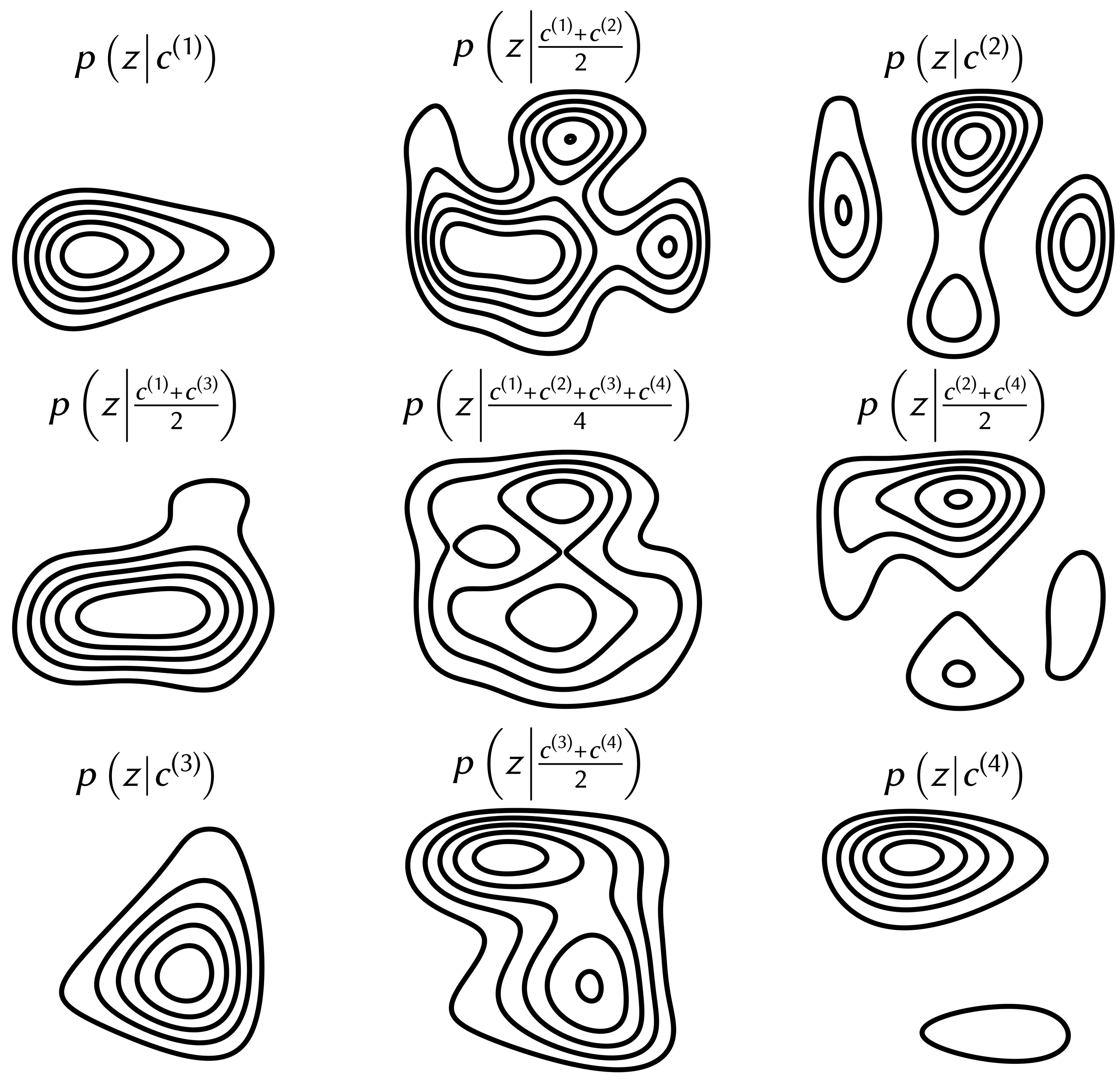
One such PDF is calculated as

$$p(z \mid c) := \left(q_1^2(z) + q_2^2(z) + \ldots + q_K^2(z)\right) / S$$

where the choice of polynomials $q_i(z)$ of N variables and degree D determines a context. Sum of squares is **always positive**, and the normalizing constant S ensures that the integral over the support is **equal to 1.**

MPSOS PDFs have a number of useful properties:

1. Hyperparameter D controls the complexity of the distribution (i.e. **the maximal number of modes**)
2. PDF can have any number of modes between 1 and the maximum, and **locations of modes are not fixed**
3. Any **convex combination** of any number of MPSOS PDFs **is another MPSOS PDF** with the same number of variables N and the same maximal degree 2D
4. Current known result is that **at most $K \le 2^N$** squares are needed to **represent any MPSOS**



All contexts are **interpolations** of each other, which makes model **reuse learned contexts** and **improves generalization.**

5. Conclusion

Preliminary results on synthetic data sets show improvement when compared to other probabilistic approaches that do not take contextual anomalies into account. Our approach is novel and integrating useful properties of MPSOS into contextual anomaly detection seems promising but needs more research.

PhD Day Faculty of Electrical Engineering and Computing, 1 June 2023.