Extreme value theory for anomaly detection –the GPD classifier

Edoardo Vignotto and Sebastian Engelke

December 17, 2020

Published on Extremes(2020).

Presented by Liujun Chen.

Table of Contents

Normal VS Abnormal

- Modern classifiers are typically not able to discriminate between normal and abnormal classes and may give high confidence predictions for unrecognizable objects.
- We call a class normal if we have examples of it during the training phase, call a class abnormal if we have no examples of it during the training phase.

Normal VS Abnormal

- The ability to distinguish between these two cases is important if there is the possibility that new classes arise in the future or if there have not been any examples of some classes in the training set due to their rarity.
- The task of distinguishing between normal and abnormal test data points is called anomaly detection

Why extreme value thoery?

- In the last few years,extreme value theory has become an important tool in multivariate statistics and machine learning.
- This is due to the fact that the extreme features, rather than the average ones, are the most important for discriminating between different objects

Table of Contents

Extreme Value Theory

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables from the distribution function Fm and denote by $M_n = \max(X_1, \ldots, X_n)$ be the maximum of the first n samples.
- $$P\left(\frac{M_n - a_n}{b_n} \leq z\right) \to G(z)$$
- For a threshold $u \in \mathbb{R}$ that tends to the upper endpoint of distribution F of X, the distribution function of $X - u$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, y > 0.$$

Table of Contents

Setting

- Denote the training data by $x_i \in \mathbb{R}^p$, each of them labeled as class $y_i \in \{C_1, \dots, C_J\}, i = 1, 2, \dots, n$
- In total, then, we have J normal classes and we assume that each class is described by a continuous density function f_{C_j} defined on $\mathbb{R}^p$, where the probability that a point in class C_j falls in the set $A \subset \mathbb{R}^p$ is

$$\int_{x \in A} f_{C_j}(x) dx.$$

- The training data set can be described as a mixture of these density functions, with unconditional density

$$f(x) = \sum_{j=1}^{J} w_j f_{C_j}(x), x \in \mathbb{R}^p$$

for weights $w_j \in [0, 1]$ with $\sum_{j=1}^{J} w_j = 1$.

Normal VS Abnormal

- The value of the function f evaluated at some point is thus large if that point has high chance to be normal. In the following, we approximate directly f and then we do not have to estimate the weights w_j.
- suppose that we have a new unlabeled test point $x_0 \in \mathbb{R}$ that we would like to mark as normal or abnormal.
- The goal in anomaly detection is to decide if x_0 comes from the distribution with density f of the training set or not.

Normal VS Abnormal

Thus, we need to perform hypothesis test:

$$H_0\colon\ x_0 \text{ is normal}$$

$$H_1\colon\ x_0 \text{ is abnormal}$$

Table of Contents

The extreme value machine