**Anomaly Detection**

Structure

1 Definition of anomaly detection

2 Creation and interpretation of a BigML anomaly detector

3 Generating an anomaly-free dataset

4 Scoring instances with the trained anomaly detector

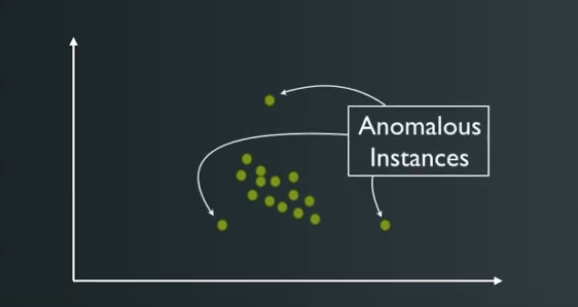
Supervised learning vs. unsupervised learning

|  |  |
| --- | --- |
| supervised learning | unsupervised learning |
| -one filed is the "objective field" (or "target variable", or "label") that is to be predicted  - The algorithm is trying to create a model that makes this prediction accurately | -Algorithm is trying to discover some structure in the data  - learned structure can often be applied to new data |

anomaly detection - unsupervised learning

anomaly detection: a way of detecting unusual instances in your dataset

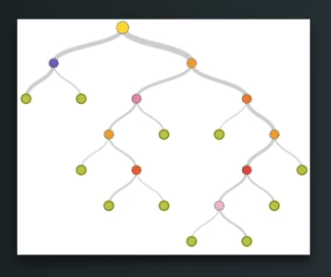
anomalous points – outliers



Application of anomaly detection (real-world use cases)

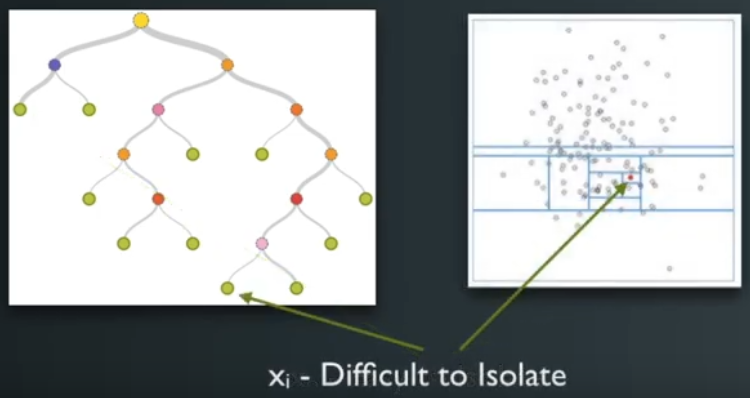
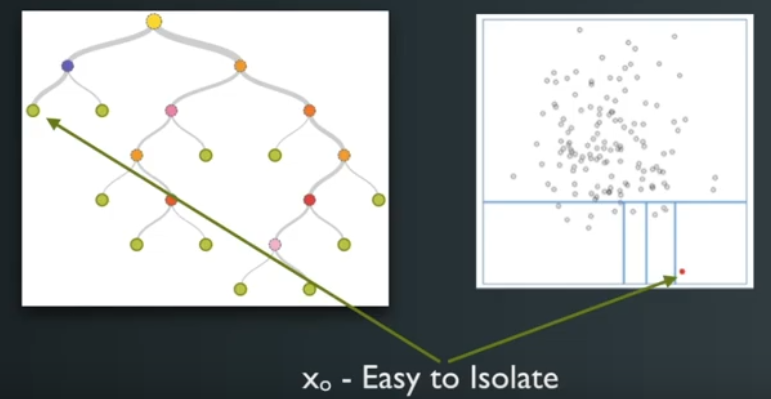
* Detecting rare, malicious behavior (fraud, intrusion) 🡪 credit card
* Alerting service technicians to possible failures 🡪 predictive maintenance
* Filtering of anomalies for “cleaner” supervised learning
* Assessing model competence

How does anomaly detection work?



Algorithm used to detect anomalies: isolation forest

🡪like a decision tree but with randomly chosen features and split-values



BigML interface

* BigML interface allows to easily view and interact with the detected anomalies in your dataset

|  |  |
| --- | --- |
| Possibility 1 | Possibility 2 |
| * Shows Top 10 anomalies in the dataset * Selected anomalies can then be removed from the dataset (and a new dataset without anomalies is created) * Anomaly detector – cannot predict a value for a field in the data | * scoring a new dataset with the trained anomaly detector * The anomaly score can be added to each line of the dataset – by creating a new dataset |