Popular Anomaly Detection Algorithms: Comparison

Algorithm	Pros	Cons
 K-NEAREST NEIGHBOR: K-NN	• VERY EASY TO UNDERSTAND • GOOD FOR CREATING MODELS THAT INCLUDE NONSTANDARD DATA TYPES SUCH AS TEXT	• LARGE STORAGE REQUIREMENTS • COMPUTATIONALLY-EXPENSIVE • SENSITIVE TO THE CHOICE OF THE SIMILARITY FUNCTION FOR COMPARING INSTANCES
 LOCAL OUTLIER FACTOR (LOF)	• WELL-KNOWN AND GOOD ALGORITHM FOR LOCAL ANOMALY DETECTION	• ONLY RELIES ON ITS DIRECT NEIGHBORHOOD • PERFORM POORLY ON DATASETS WITH GLOBAL ANOMALIES
 K-MEANS	• LOW COMPLEXITY • VERY EASY TO IMPLEMENT	• EACH CLUSTER HAS PRETTY EQUAL NUMBERS OF OBSERVATIONS • NECESSITY OF SPECIFYING K • ONLY WORK WITH NUMERICAL DATA
 SUPPORT VECTOR MACHINE (SVM)	• FIND THE BEST SEPARATION HYPERPLANE • DEAL WITH VERY HIGH DIMENSIONAL DATA • CAN LEARN VERY ELABORATE CONCEPTS • WORK VERY WELL	• REQUIRE BOTH POSITIVE & NEGATIVE EXAMPLES • REQUIRE LOTS OF MEMORY • SOME NUMERICAL STABILITY PROBLEMS • NEED TO SELECT A GOOD KERNEL FUNCTION
 NEURAL NETWORKS BASED ANOMALY DETECTION	• FULFIL TASKS THAT A LINEAR PROGRAM CANNOT • LEARNS AND DOES NOT NEED TO BE REPROGRAMMED • CAN BE IMPLEMENTED IN ANY APPLICATION	• NEEDS TRAINING TO OPERATE • REQUIRES HIGH PROCESSING TIME FOR LARGE NEURAL NETWORKS • THE ARCHITECTURE NEEDS TO BE EMULATED