

异常检测 vs 监督学习

Anomaly detection	vs.	Supervised learning
<ul style="list-style-type: none">→ Very small number of positive examples ($y = 1$). (0-20 is common).→ Large number of negative ($y = 0$) examples. $p(x)$→ Many different "types" of anomalies. Hard for any algorithm to learn from positive examples what the anomalies look like;→ future anomalies may look nothing like any of the anomalous examples we've seen so far.		<ul style="list-style-type: none">Large number of positive and negative examples.Enough positive examples for algorithm to get a sense of what positive examples are like, future positive examples likely to be similar to ones in training set.Spam

当只有很少数量的正样本和大量负样本时，我们通常使用异常检测算法，因为监督学习很难从很少的正样本中学习出什么样的数据是属于没有故障的，并且会出现很多之前并未出现过的故障，通常这种情况下，使用大量的负样本进行建模 $p(x)$ ，再把正样本用在交叉验证集和训练集进行验证。

当有大量的正样本和负样本时，使用监督算法。

Anomaly detection	vs.	Supervised learning
<ul style="list-style-type: none">→ • Fraud detection $y=1$→ • Manufacturing (e.g. aircraft engines)→ • Monitoring machines in a data center•		<ul style="list-style-type: none">• Email spam classification• Weather prediction (sunny/rainy/etc).• Cancer classification•

在负样本数量也不大时，运用异常检测，但数量很大时，使用监督学习。

