OOD（out of distribution detection）是数据挖掘领域的一个重要分支，通俗来讲OOD检测的任务就是**识别出分布与模型训练数据不同的输入数据。**

最简单的例子：
一个模型接受一堆宠物狗的照片，并试图将他们按照不同品种分成不同的类别。在训练过程中，我们所提供给模型的自然是一堆狗的照片，但是在实际预测过程中，如果输入中混入了其他动物的照片（比如一只猫）那我们当然希望模型能够识别出，这个异常输入（猫）不属于任何一个品种，从而将它划分为分布外数据（即OOD数据）。

由于OOD检测的特殊性，它被广泛应用于增强模型的泛化能力，以及提高模型的鲁棒性，应对可能发生极端情况的场景。例如无人驾驶领域，实验室数据很难模拟实际路况可能发生的极端情景，当无人驾驶的操作系统面临现实中的异常输入时，我们希望模型能够正确识别输入模式并非训练数据中的任何一种模式，从而采取不同的应对方案，而不是继续按照实验训练一样处理，否则可能会造成严重后果。

在正式开始之前，我们重新回顾一下OOD检测中的两个基本概念：

- OOD：即OOD数据，全称为分布外（out of distribution）数据，是我们希望OOD检测模型能够正确识别并剔除出来的数据。
- ID：即ID数据，全程分部内数据（in distribution）顾名思义与OOD的概念相对，意味着模型接受数据分布与训练数据集相同。

下面我们来看几个传统OOD检测基本方向。

softmax-based 方法

可以先看看softmax的讲解入门级都能看懂的softmax详解-CSDN博客

OOD检测的softmax-based方法基于softmax概率，大致过程分为以下几步：

- 预训练模型：对网络进行预训练，使其能够在已知的类别上正确进行分类。
- 计算softmax概率分数：网络输出一个分数向量，我们通过softmax方法将其转换为概率分数：

$$softmax(x_i) = e^{x_i} / \sum_j e^{x_j}$$

对于一个ID样本，我们有理由相信模型对于它的分类结果总是比较自信的，所有类别对应的概率分数中的最高项应该明显高于一个阈值 γ （这个阈值比较empirical，阈值的选取标准也是OOD检测的一个研究方向），如果对于一个样本模型输出的最大概率分数小于这个阈值，我们就可以将他归类为OOD数据，反之为ID数据。

uncertainty 方法

当模型面对OOD数据时，模型输出的不确定性会显著增加，因此也可以将输出的不确定性作为区分ID和OOD样本的依据之一。不确定性检测方法通常与贝叶斯等数学方法挂钩，这里只简单介绍下贝叶斯神经网络（BNN）。BNNs通过在网络的权重上引入概率分布来捕捉不确定性。对于每个输入，网络不仅提供一个预测，还提供预测的不确定性度量。通过多次前向传播（每次使用不同的权重样本）并计算预测的方差，可以估计epistemic uncertainty。

PU方法

OOD检测的概念与传统半监督学习（positive and unlabeled learning）有许多相似之处，近年来将PU方法引入OOD检测也是一个新颖的研究方向。
PU方法意味着能提供给样本的数据只有确保为阳性的正样本（positive sample）与没有被打上标签的样本（unlabeled sample），这种情况更接近实际中OOD检测面临的数据分布，因此可以显著提升模型的OOD检测能力。

PU方法通常对已知正样本数据进行特征挖掘，并试图克服没有负样本带来的选择偏差（labeled bias），通常用无标记样本与正样本的特征组合来模拟得到负样本的分布，进而将问题转化为全监督学习。