对抗训练

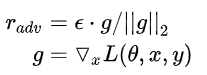
基本思想：

对抗训练的min-max公式

Max的目的为找到样本空间中使模型loss最大的一组扰动，外层min使得模型在面对这种数据分布时候，模型受到的最大干扰的期望最小

由此得到的模型能对噪声有一定的防御能力。

Fast Gradient Method （FGM）



先计算原样本的梯度 ，将梯度方向上的小扰动 叠加到原样本上得到对抗样本，将对抗样本的经过模型作用后的新梯度叠加到原来梯度上，再将两次梯度累加后的合方向进行梯度更新。