# 隐私保护机器学习(PPML)背景

数据很重要：

模型训练需要大量数据

模型服务商需要接收用户输入作为推理输入

并且数据含有敏感性息

技术路线：主要介绍MPC，其他路线还有TEE、FL

多方计算同一个函数，但是不会泄露信息

同样有隐私训练、隐私推理两步。

# NN密态训练和推理

## 简单案例

以逻辑回归为案例，需要这四步：

1. 数据来源

数据提供方加载数据

1. 如何加密保护

把明文数据发送到SPU（计算能力提供方）上

计算方拿到了密文

1. 如何定义模型

使用jax手动实现前向和反向传播

1. 如何执行密态计算

计算过程被编译为一个密态的计算图，按照MPC协议执行

## 复杂案例

### 手动建模

使用flax/stax库

参考：spu/examples/python/ml/{flax\_mlp, stax\_nn}

### 复用模型

模型的定义和计算和明文一样，但是run\_on\_spu需要把input\_ids和params需要加密后提供到spu上去

### 支持更多模型

需要实现密态算子，目前XLA的语义没有全部被SPU支持

SPU的实现不一定是性能最好的，有优化空间

# 实践

见两个作业文件