Vertical Federated Learning (VFL):
- Features of the samples are partitioned across clients and the labels are owned by server.

Challenge:
- Aggregation: averaging local embeddings fails to capture the unique properties of each client.
- Communication: communicating gradients for each training step incurs high costs.

Our contributions:
- Framework: an effective framework with multiple heads (VIM).
- Algorithm: an ADMM-based method (VIMADMM) reducing communication costs by allowing multiple local updates at each step.
- Empirical study: 1) VIMADMM converges faster and achieves higher accuracy; 2) client-level explanation under VIM based on the linear heads.

VIMADMM Workflow
- Key steps of VIMADMM at each round $t$:
  1. Communication from client to server: a batch of local embeddings $(\tilde{y}_i^{(t)})_{i=1}^N$.
  2. Sever updates auxiliary variables:
     \[
     x_i^{(t)} = \arg\min_{x_i} \sum_{j=0}^{n_i} \beta_{ij} x_i \quad \text{s.t.} \quad \sum_{i=0}^n f(z_i^{(t)}; \theta_j) W_{ik} = z_j, \quad (i,j) \in E
     \]
  3. Sever updates dual variables:
     \[
     y_{j}^{(t+1)} = \sum_{i=0}^N \beta_{ij} x_i^{(t)} + \sum_{i=0}^N \alpha_{ij} W_{ik}^{(t)}, \quad (j,k) \in E
     \]
  4. Sever updates linear heads:
     \[
     x_{i}^{(t+1)} = \sum_{j=0}^{n_i} \beta_{ij} x_i + \sum_{i=0}^N \alpha_{ij} W_{ik}^{(t)}, \quad (i,j) \in E
     \]
  5. Communication from server to each client: residual variable, dual variables, and one corresponding linear head.
  6. Client updates local model parameters:
     \[
     \theta_{i}^{(t)} = \arg\min_{\theta_i} \sum_{j=0}^{n_i} \beta_{ij} x_i + \sum_{i=0}^N \alpha_{ij} W_{ik}^{(t)}, \quad (i,j) \in E
     \]

- Reduce frequency by allowing multiple local updating steps at each round.
- Reduce dimensionality of information by exchanging ADMM-related variables.

VIMADMM vs SOTA [1,2]
- VFL classification on four datasets: MNIST, CIFAR, NUS-WIDE and ModelNet40.

Experiments
- Faster convergence; improved communication efficiency; higher accuracy.

Client-level Explainability of VIM
- The weights of linear heads reflect the importance of local clients; Perturbing the client with high weights has higher impact on test accuracy; VIM enables client denoising by lowering their weights.

More details and results are in our paper:
- Details for algorithm VIMADMM-J for VFL without model splitting setting.
- Results on communication costs comparison, effect of penalty factor and local steps, client summarization and the visualization of the local embedding.

References: