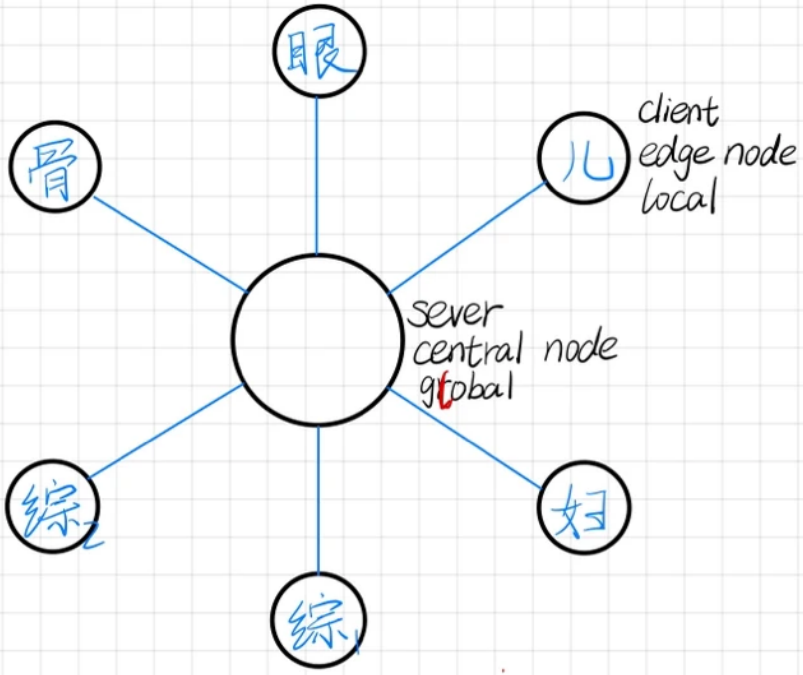
① FL

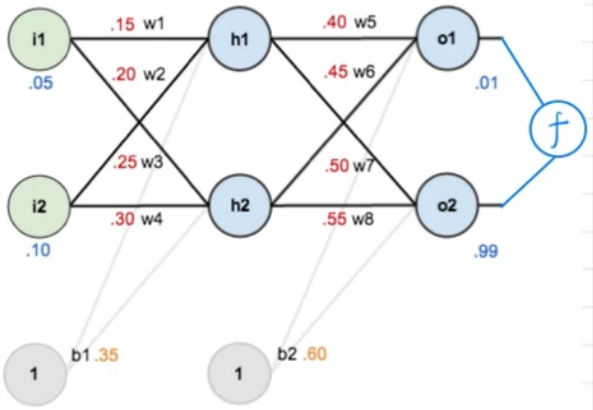


独立同分布

② IID & Non-IID

数据	眼CT	肺CT	腿CT
眼	30	0	0
儿	0	18	2
妇	0	15	1
骨	0	20	20
综1	15	15	15
综2	5	5	5

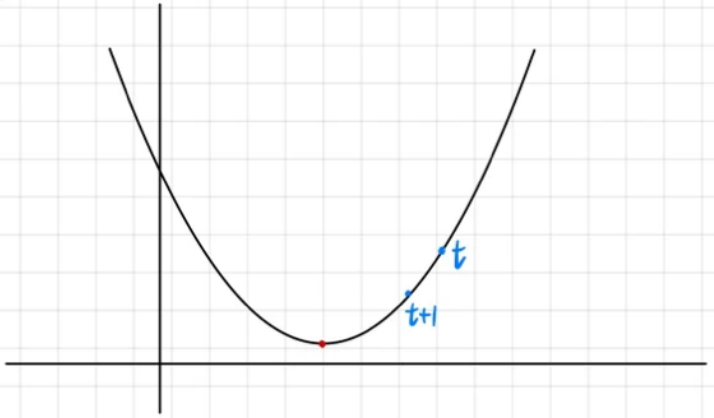
③ model W



r_{real}

$E_{误差} = |f - r|$

④ SGD



S: Stochastic 随机(选一个数据)

GD: gradient desent 梯度下降

$$W_{t+1} = W_t - \eta * \nabla F(W_t)$$

A. 前向传播：对于h1来说

$$IN = i1 \cdot W1 + i2 \cdot W2$$
$$= 0.05 \cdot 0.15 + 0.10 \cdot 025$$
$$= 0.0325 \text{ (net)}$$
$$OUT = ReLu(IN) = 0.0325 \text{ (out)}$$

B. 后向传播

以权重参数w5为例，如果我们想知道w5对整体误差产生了多少影响，可以用整体误差对w5求偏导求出：（链式法则）

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial w_5}$$

最后我们来更新w5的值：

$$w_5^+ = w_5 - \eta * \frac{\partial E_{total}}{\partial w_5}$$

⑤FedAvg

Algorithm 1 FederatedAveraging

Server executes:

initialize w_0

for each round $t = 1, 2, \ldots$ **do**

$S_t = (\text{random set of } \max(C \cdot K, 1) \text{ clients})$

for each client $k \in S_t$ **in parallel do**

$w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$

$w_{t+1} \leftarrow \sum_{t=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w): *// Executed on client k*

for each local epoch i from 1 to E **do**

batches $\leftarrow$ (data $\mathcal{P}_k$ split into batches of size B) #有点MBGD的意思

for batch b in batches **do**

$w \leftarrow w - \eta \nabla \ell(w; b)$

return w to server

C：在每轮上执行计算的客户端的比例；
E：每个客户端在每轮上对其本地数据集执行的训练通过数；
B：客户端更新所使用的小批量大小