

Distributed CPU training on FEP (SLURM)

1. Introducere

Proiectul are ca scop principal realizarea unui model de “Distributed training” pe mai multe microprocesoare, numite CPU-uri. Aceasta operatie se desfasoara in mod automat si se realizeaza pe infrastructura fep-ului, anume SLURM, care reprezinta un cluster de calculatoare situat in UPB, cladirea PRECIS. Pentru realizarea acestui proiect se va folosi biblioteca MPI, utilizand functia de `all reduce` din cadrul acestei biblioteci. Problema presupune impartirea modelului si a taskurilor acestuia intre mai multe procese, fiecare realizand un calcul independent pentru forward pass si backward pass, cu alte cuvinte back propagation, implementat de utilizator, ci nu de functii deja implementate (ex: `autograd`, `backward`, `forward`). Ulterior, toate gradientele calculate sunt sincronizate prin metoda `all reduce` cu scopul de a imbunatati parametrii modelului.

2. Planul de realizare si implementarea proiectului

Planul de realizare și implementare a proiectului a fost structurat în mod sistematic, având ca scop principal dezvoltarea unui model simplu distribuit pentru antrenarea unui algoritm de învățare automată pe mai multe nuclee CPU. Proiectul este construit pe baza bibliotecii `mpi4py`, care permite distribuirea sarcinilor de calcul pe mai multe procese. Fiecare proces MPI rulează pe un CPU core separat, iar sincronizarea între ele se face folosind comunicare prin `comm.Allreduce`. În cadrul codului, toate operațiile sunt realizate folosind biblioteca `numpy`, care este optimizată pentru CPU-uri. Nu există nicio referință la PyTorch, CUDA sau GPU-uri, deci calculele sunt efectuate exclusiv pe CPU-uri.

Numărul de procese specificat în comanda `mpirun -np <num_processes>` determină câte nuclee CPU sunt utilizate. Fiecare proces calculează gradientele independent și le sincronizează cu celelalte procese utilizând `MPI.Allreduce`. Asadar, folosind comanda `mpirun -np 4 python3 distributed_training.py` sunt rezultate 4 CPU-uri. Ordinea în care procesele își execută codul depinde de mecanismul intern al MPI și al sistemului de operare.

Fiecare proces MPI începe executarea simultan, dar nu toate procesele rulează cu aceeași viteză (datorită timpului de inițializare, alocării resurselor sau accesului la date). Ordinea afișării rezultatelor depinde de momentul în care fiecare proces finalizează o sarcină și scrie în fluxul de ieșire (`stdout`).

Am generat un set de date artificial pentru a facilita testarea și reproducerea experimentelor, distribuind aceste date în mod echitabil între procesele paralele. Sincronizarea gradientelor a fost realizată utilizând operațiunea `Allreduce` oferită de MPI, care a permis agregarea contribuțiilor fiecărui proces pentru a actualiza în mod global parametrii modelului. Am implementat loguri detaliate și mecanisme de debug pentru a urmări funcționarea fiecărui proces, asigurându-ne că sincronizarea și calculul gradientelor sunt corecte.

Afișarea rezultatelor în stdout

Ieșirea `stdout` este comună pentru toate procesele. Mesajele din diferite procese ajung în fluxul de ieșire într-un mod nesincronizat. De exemplu, procesul 2 poate termina primul o sarcină și afișa rezultatul înaintea procesului 0. În funcție de modul în care MPI alocă procesele pe CPU-uri, anumite procese pot beneficia de mai multe resurse și, prin urmare, finalizează mai repede anumite sarcini.

3. Rezultate

```
Process 1 is running on this node.
Process 1, Epoch 1: grad_W mean = -0.0283, grad_b mean = -0.0010
Process 1, Epoch 1: Contribution to W = 8.99%, b = 0.05%
Process 1, Epoch 1: Sync time = 0.201988 seconds
Process 1, Epoch 2: grad_W mean = -0.0300, grad_b mean = -0.0043
Process 1, Epoch 2: Contribution to W = 9.11%, b = 0.21%
Process 1, Epoch 2: Sync time = 0.013197 seconds
Process 1, Epoch 3: grad_W mean = -0.0317, grad_b mean = -0.0075
Process 1, Epoch 3: Contribution to W = 9.23%, b = 0.38%
Process 1, Epoch 3: Sync time = 0.061170 seconds
Process 1, Epoch 4: grad_W mean = -0.0333, grad_b mean = -0.0107
Process 1, Epoch 4: Contribution to W = 9.36%, b = 0.54%
Process 1, Epoch 4: Sync time = 0.000174 seconds
Process 1, Epoch 5: grad_W mean = -0.0350, grad_b mean = -0.0139
Process 1, Epoch 5: Contribution to W = 9.50%, b = 0.71%
Process 1, Epoch 5: Sync time = 0.000069 seconds
Process 1, Epoch 6: grad_W mean = -0.0366, grad_b mean = -0.0171
Process 1, Epoch 6: Contribution to W = 9.64%, b = 0.88%
Process 1, Epoch 6: Sync time = 0.000156 seconds
Process 1, Epoch 7: grad_W mean = -0.0382, grad_b mean = -0.0203
Process 1, Epoch 7: Contribution to W = 9.78%, b = 1.05%
Process 1, Epoch 7: Sync time = 0.000073 seconds
Process 1, Epoch 8: grad_W mean = -0.0398, grad_b mean = -0.0234
Process 1, Epoch 8: Contribution to W = 9.92%, b = 1.22%
Process 1, Epoch 8: Sync time = 0.000056 seconds
Process 1, Epoch 9: grad_W mean = -0.0414, grad_b mean = -0.0265
Process 1, Epoch 9: Contribution to W = 10.07%, b = 1.39%
Process 1, Epoch 9: Sync time = 0.000141 seconds
Process 1, Epoch 10: grad_W mean = -0.0430, grad_b mean = -0.0296
Process 1, Epoch 10: Contribution to W = 10.22%, b = 1.56%
Process 1, Epoch 10: Sync time = 0.000035 seconds
Process 2 is running on this node.
Process 2, Epoch 1: grad_W mean = 1.0306, grad_b mean = 2.1341
Process 2, Epoch 1: Contribution to W = 110.98%, b = 106.06%
Process 2, Epoch 1: Sync time = 0.483057 seconds
Process 2, Epoch 2: grad_W mean = 1.0290, grad_b mean = 2.1308
Process 2, Epoch 2: Contribution to W = 111.55%, b = 106.59%
Process 2, Epoch 2: Sync time = 0.000243 seconds
Process 2, Epoch 3: grad_W mean = 1.0273, grad_b mean = 2.1276
Process 2, Epoch 3: Contribution to W = 112.14%, b = 107.12%
Process 2, Epoch 3: Sync time = 0.000106 seconds
Process 2, Epoch 4: grad_W mean = 1.0257, grad_b mean = 2.1243
Process 2, Epoch 4: Contribution to W = 112.72%, b = 107.65%
Process 2, Epoch 4: Sync time = 0.000112 seconds
Process 2, Epoch 5: grad_W mean = 1.0241, grad_b mean = 2.1211
Process 2, Epoch 5: Contribution to W = 113.31%, b = 108.19%
Process 2, Epoch 5: Sync time = 0.000117 seconds
Process 2, Epoch 6: grad_W mean = 1.0225, grad_b mean = 2.1179
Process 2, Epoch 6: Contribution to W = 113.91%, b = 108.73%
Process 2, Epoch 6: Sync time = 0.000075 seconds
Process 2, Epoch 7: grad_W mean = 1.0209, grad_b mean = 2.1147
Process 2, Epoch 7: Contribution to W = 114.50%, b = 109.27%
Process 2, Epoch 7: Sync time = 0.000068 seconds
Process 2, Epoch 8: grad_W mean = 1.0193, grad_b mean = 2.1115
Process 2, Epoch 8: Contribution to W = 115.10%, b = 109.82%
Process 2, Epoch 8: Sync time = 0.000086 seconds
Process 2, Epoch 9: grad_W mean = 1.0178, grad_b mean = 2.1084
Process 2, Epoch 9: Contribution to W = 115.71%, b = 110.37%
Process 2, Epoch 9: Sync time = 0.000043 seconds
Process 2, Epoch 10: grad_W mean = 1.0162, grad_b mean = 2.1053
Process 2, Epoch 10: Contribution to W = 116.32%, b = 110.92%
Process 2, Epoch 10: Sync time = 0.000068 seconds
```

```
Process 3, Epoch 4: grad_W mean = 0.0936, grad_b mean = 0.2279
Process 3, Epoch 4: Contribution to W = 12.39%, b = 11.55%
Process 3, Epoch 4: Sync time = 0.000147 seconds
Process 3, Epoch 5: grad_W mean = 0.0921, grad_b mean = 0.2248
Process 3, Epoch 5: Contribution to W = 12.34%, b = 11.47%
Process 3, Epoch 5: Sync time = 0.000059 seconds
Process 3, Epoch 6: grad_W mean = 0.0906, grad_b mean = 0.2217
Process 3, Epoch 6: Contribution to W = 12.30%, b = 11.38%
Process 3, Epoch 6: Sync time = 0.000156 seconds
Process 3, Epoch 7: grad_W mean = 0.0891, grad_b mean = 0.2187
Process 3, Epoch 7: Contribution to W = 12.25%, b = 11.30%
Process 3, Epoch 7: Sync time = 0.000100 seconds
Process 3, Epoch 8: grad_W mean = 0.0876, grad_b mean = 0.2156
Process 3, Epoch 8: Contribution to W = 12.21%, b = 11.21%
Process 3, Epoch 8: Sync time = 0.000054 seconds
Process 3, Epoch 9: grad_W mean = 0.0862, grad_b mean = 0.2126
Process 3, Epoch 9: Contribution to W = 12.16%, b = 11.13%
Process 3, Epoch 9: Sync time = 0.000141 seconds
Process 3, Epoch 10: grad_W mean = 0.0847, grad_b mean = 0.2096
Process 3, Epoch 10: Contribution to W = 12.12%, b = 11.04%
Process 3, Epoch 10: Sync time = 0.000034 seconds
Starting distributed training with 4 processes.
Process 0 is running on this node.
Process 0, Epoch 1: grad_W mean = -0.1882, grad_b mean = -0.3584
Process 0, Epoch 1: Contribution to W = 20.77%, b = 17.81%
Process 0, Epoch 1: Sync time = 0.285622 seconds
Epoch 1/10, Loss: 0.1183
Process 0, Epoch 2: grad_W mean = -0.1899, grad_b mean = -0.3617
Process 0, Epoch 2: Contribution to W = 21.09%, b = 18.09%
Process 0, Epoch 2: Sync time = 0.013190 seconds
Epoch 2/10, Loss: 0.1189
Process 0, Epoch 3: grad_W mean = -0.1915, grad_b mean = -0.3649
Process 0, Epoch 3: Contribution to W = 21.40%, b = 18.37%
Process 0, Epoch 3: Sync time = 0.000103 seconds
Epoch 3/10, Loss: 0.1195
Process 0, Epoch 4: grad_W mean = -0.1932, grad_b mean = -0.3682
Process 0, Epoch 4: Contribution to W = 21.72%, b = 18.66%
Process 0, Epoch 4: Sync time = 0.000067 seconds
Epoch 4/10, Loss: 0.1202
Process 0, Epoch 5: grad_W mean = -0.1948, grad_b mean = -0.3714
Process 0, Epoch 5: Contribution to W = 22.04%, b = 18.94%
Process 0, Epoch 5: Sync time = 0.000123 seconds
Epoch 5/10, Loss: 0.1208
Process 0, Epoch 6: grad_W mean = -0.1964, grad_b mean = -0.3746
Process 0, Epoch 6: Contribution to W = 22.36%, b = 19.23%
Process 0, Epoch 6: Sync time = 0.000078 seconds
Epoch 6/10, Loss: 0.1214
Process 0, Epoch 7: grad_W mean = -0.1980, grad_b mean = -0.3778
Process 0, Epoch 7: Contribution to W = 22.69%, b = 19.52%
Process 0, Epoch 7: Sync time = 0.000064 seconds
Epoch 7/10, Loss: 0.1221
Process 0, Epoch 8: grad_W mean = -0.1996, grad_b mean = -0.3810
Process 0, Epoch 8: Contribution to W = 23.01%, b = 19.81%
Process 0, Epoch 8: Sync time = 0.000087 seconds
Epoch 8/10, Loss: 0.1227
Process 0, Epoch 9: grad_W mean = -0.2012, grad_b mean = -0.3841
Process 0, Epoch 9: Contribution to W = 23.34%, b = 20.11%
Process 0, Epoch 9: Sync time = 0.000030 seconds
Epoch 9/10, Loss: 0.1233
Process 0, Epoch 10: grad_W mean = -0.2028, grad_b mean = -0.3872
Process 0, Epoch 10: Contribution to W = 23.67%, b = 20.40%
Process 0, Epoch 10: Sync time = 0.000069 seconds
Epoch 10/10, Loss: 0.1240
Saved loss plot as 'loss_plot.png'
```

Process 1 is running on this node.

Process 1, Epoch 1: grad_W mean = -0.0283, grad_b mean = -0.0010

Process 1, Epoch 1: Contribution to W = 8.99%, b = 0.05%

Process 1, Epoch 1: Sync time = 0.201988 seconds

Process 1, Epoch 2: grad_W mean = -0.0300, grad_b mean = -0.0043

Process 1, Epoch 2: Contribution to W = 9.11%, b = 0.21%

Process 1, Epoch 2: Sync time = 0.013197 seconds

Process 1, Epoch 3: grad_W mean = -0.0317, grad_b mean = -0.0075

Process 1, Epoch 3: Contribution to W = 9.23%, b = 0.38%

Process 1, Epoch 3: Sync time = 0.061170 seconds

Process 1, Epoch 4: grad_W mean = -0.0333, grad_b mean = -0.0107

Process 1, Epoch 4: Contribution to W = 9.36%, b = 0.54%

Process 1, Epoch 4: Sync time = 0.000174 seconds

Process 1, Epoch 5: grad_W mean = -0.0350, grad_b mean = -0.0139

Process 1, Epoch 5: Contribution to W = 9.50%, b = 0.71%

Process 1, Epoch 5: Sync time = 0.000069 seconds

Process 1, Epoch 6: grad_W mean = -0.0366, grad_b mean = -0.0171

Process 1, Epoch 6: Contribution to W = 9.64%, b = 0.88%

Process 1, Epoch 6: Sync time = 0.000156 seconds

Process 1, Epoch 7: grad_W mean = -0.0382, grad_b mean = -0.0203

Process 1, Epoch 7: Contribution to W = 9.78%, b = 1.05%

Process 1, Epoch 7: Sync time = 0.000073 seconds

Process 1, Epoch 8: grad_W mean = -0.0398, grad_b mean = -0.0234

Process 1, Epoch 8: Contribution to W = 9.92%, b = 1.22%

Process 1, Epoch 8: Sync time = 0.000056 seconds

Process 1, Epoch 9: grad_W mean = -0.0414, grad_b mean = -0.0265

Process 1, Epoch 9: Contribution to W = 10.07%, b = 1.39%

Process 1, Epoch 9: Sync time = 0.000141 seconds

Process 1, Epoch 10: grad_W mean = -0.0430, grad_b mean = -0.0296

Process 1, Epoch 10: Contribution to W = 10.22%, b = 1.56%

Process 1, Epoch 10: Sync time = 0.000035 seconds

Process 2 is running on this node.

Process 2, Epoch 1: grad_W mean = 1.0306, grad_b mean = 2.1341

Process 2, Epoch 1: Contribution to W = 110.98%, b = 106.06%

Process 2, Epoch 1: Sync time = 0.483057 seconds

Process 2, Epoch 2: grad_W mean = 1.0290, grad_b mean = 2.1308

Process 2, Epoch 2: Contribution to W = 111.55%, b = 106.59%

Process 2, Epoch 2: Sync time = 0.000243 seconds

Process 2, Epoch 3: grad_W mean = 1.0273, grad_b mean = 2.1276

Process 2, Epoch 3: Contribution to W = 112.14%, b = 107.12%

Process 2, Epoch 3: Sync time = 0.000106 seconds

Process 2, Epoch 4: grad_W mean = 1.0257, grad_b mean = 2.1243

Process 2, Epoch 4: Contribution to W = 112.72%, b = 107.65%

Process 2, Epoch 4: Sync time = 0.000112 seconds

Process 2, Epoch 5: grad_W mean = 1.0241, grad_b mean = 2.1211

Process 2, Epoch 5: Contribution to W = 113.31%, b = 108.19%

Process 2, Epoch 5: Sync time = 0.000117 seconds

Process 2, Epoch 6: grad_W mean = 1.0225, grad_b mean = 2.1179

Process 2, Epoch 6: Contribution to W = 113.91%, b = 108.73%

Process 2, Epoch 6: Sync time = 0.000075 seconds

Process 2, Epoch 7: grad_W mean = 1.0209, grad_b mean = 2.1147

Process 2, Epoch 7: Contribution to W = 114.50%, b = 109.27%

Process 2, Epoch 7: Sync time = 0.000068 seconds

Process 2, Epoch 8: grad_W mean = 1.0193, grad_b mean = 2.1115

Process 2, Epoch 8: Contribution to W = 115.10%, b = 109.82%

Process 2, Epoch 8: Sync time = 0.000086 seconds

Process 2, Epoch 9: grad_W mean = 1.0178, grad_b mean = 2.1084

Process 2, Epoch 9: Contribution to W = 115.71%, b = 110.37%

Process 2, Epoch 9: Sync time = 0.000043 seconds

Process 2, Epoch 10: grad_W mean = 1.0162, grad_b mean = 2.1053

Process 2, Epoch 10: Contribution to W = 116.32%, b = 110.92%

Process 2, Epoch 10: Sync time = 0.000068 seconds

Process 3 is running on this node.

Process 3, Epoch 1: grad_W mean = 0.0983, grad_b mean = 0.2374

Process 3, Epoch 1: Contribution to W = 12.53%, b = 11.80%

Process 3, Epoch 1: Sync time = 0.084022 seconds

Process 3, Epoch 2: grad_W mean = 0.0967, grad_b mean = 0.2342

Process 3, Epoch 2: Contribution to W = 12.48%, b = 11.72%

Process 3, Epoch 2: Sync time = 0.000197 seconds

Process 3, Epoch 3: grad_W mean = 0.0952, grad_b mean = 0.2311

Process 3, Epoch 3: Contribution to W = 12.44%, b = 11.63%

Process 3, Epoch 3: Sync time = 0.000082 seconds

Process 3, Epoch 4: grad_W mean = 0.0936, grad_b mean = 0.2279

Process 3, Epoch 4: Contribution to W = 12.39%, b = 11.55%

Process 3, Epoch 4: Sync time = 0.000147 seconds

Process 3, Epoch 5: grad_W mean = 0.0921, grad_b mean = 0.2248

Process 3, Epoch 5: Contribution to W = 12.34%, b = 11.47%

Process 3, Epoch 5: Sync time = 0.000059 seconds

Process 3, Epoch 6: grad_W mean = 0.0906, grad_b mean = 0.2217

Process 3, Epoch 6: Contribution to W = 12.30%, b = 11.38%

Process 3, Epoch 6: Sync time = 0.000156 seconds

Process 3, Epoch 7: grad_W mean = 0.0891, grad_b mean = 0.2187

Process 3, Epoch 7: Contribution to W = 12.25%, b = 11.30%

Process 3, Epoch 7: Sync time = 0.000100 seconds

Process 3, Epoch 8: grad_W mean = 0.0876, grad_b mean = 0.2156

Process 3, Epoch 8: Contribution to W = 12.21%, b = 11.21%

Process 3, Epoch 8: Sync time = 0.000054 seconds

Process 3, Epoch 9: grad_W mean = 0.0862, grad_b mean = 0.2126

Process 3, Epoch 9: Contribution to W = 12.16%, b = 11.13%

Process 3, Epoch 9: Sync time = 0.000141 seconds

Process 3, Epoch 10: grad_W mean = 0.0847, grad_b mean = 0.2096

Process 3, Epoch 10: Contribution to W = 12.12%, b = 11.04%

Process 3, Epoch 10: Sync time = 0.000034 seconds

Starting distributed training with 4 processes.

Process 0 is running on this node.

Process 0, Epoch 1: grad_W mean = -0.1882, grad_b mean = -0.3584

Process 0, Epoch 1: Contribution to W = 20.77%, b = 17.81%

Process 0, Epoch 1: Sync time = 0.285622 seconds

Epoch 1/10, Loss: 0.1183

Process 0, Epoch 2: grad_W mean = -0.1899, grad_b mean = -0.3617

Process 0, Epoch 2: Contribution to W = 21.09%, b = 18.09%

Process 0, Epoch 2: Sync time = 0.013190 seconds

Epoch 2/10, Loss: 0.1189

Process 0, Epoch 3: grad_W mean = -0.1915, grad_b mean = -0.3649

Process 0, Epoch 3: Contribution to W = 21.40%, b = 18.37%

Process 0, Epoch 3: Sync time = 0.000103 seconds

Epoch 3/10, Loss: 0.1195

Process 0, Epoch 4: grad_W mean = -0.1932, grad_b mean = -0.3682

Process 0, Epoch 4: Contribution to W = 21.72%, b = 18.66%

Process 0, Epoch 4: Sync time = 0.000067 seconds

Epoch 4/10, Loss: 0.1202

Process 0, Epoch 5: grad_W mean = -0.1948, grad_b mean = -0.3714

Process 0, Epoch 5: Contribution to W = 22.04%, b = 18.94%

Process 0, Epoch 5: Sync time = 0.000123 seconds

Epoch 5/10, Loss: 0.1208

Process 0, Epoch 6: grad_W mean = -0.1964, grad_b mean = -0.3746

Process 0, Epoch 6: Contribution to W = 22.36%, b = 19.23%

Process 0, Epoch 6: Sync time = 0.000078 seconds

Epoch 6/10, Loss: 0.1214

Process 0, Epoch 7: grad_W mean = -0.1980, grad_b mean = -0.3778

Process 0, Epoch 7: Contribution to W = 22.69%, b = 19.52%

Process 0, Epoch 7: Sync time = 0.000064 seconds

Epoch 7/10, Loss: 0.1221

Process 0, Epoch 8: grad_W mean = -0.1996, grad_b mean = -0.3810

Process 0, Epoch 8: Contribution to W = 23.01%, b = 19.81%

Process 0, Epoch 8: Sync time = 0.000087 seconds

Epoch 8/10, Loss: 0.1227

Process 0, Epoch 9: grad_W mean = -0.2012, grad_b mean = -0.3841

Process 0, Epoch 9: Contribution to W = 23.34%, b = 20.11%

Process 0, Epoch 9: Sync time = 0.000030 seconds

Epoch 9/10, Loss: 0.1233

Process 0, Epoch 10: grad_W mean = -0.2028, grad_b mean = -0.3872

Process 0, Epoch 10: Contribution to W = 23.67%, b = 20.40%

Process 0, Epoch 10: Sync time = 0.000069 seconds

Epoch 10/10, Loss: 0.1240

Parametrii modelului

W:

9.505355511359944720e-02

8.608506196416068068e-01

9.706517383287199818e-01

9.587595662507508720e-01

9.052999557056433666e-01

7.720664041596144722e-01

3.304639375763213094e-01

7.925454800604284089e-02

4.038927523448967993e-01

2.297354112909793522e-01

b:

1.276012244831889186e-01

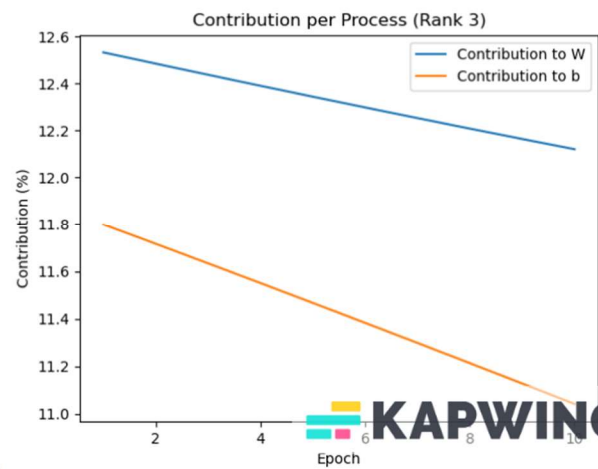
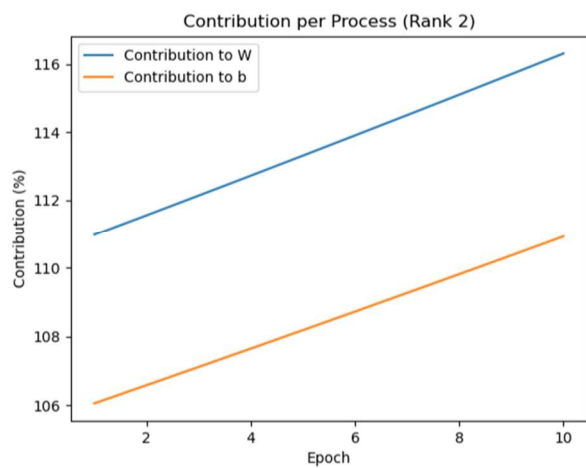
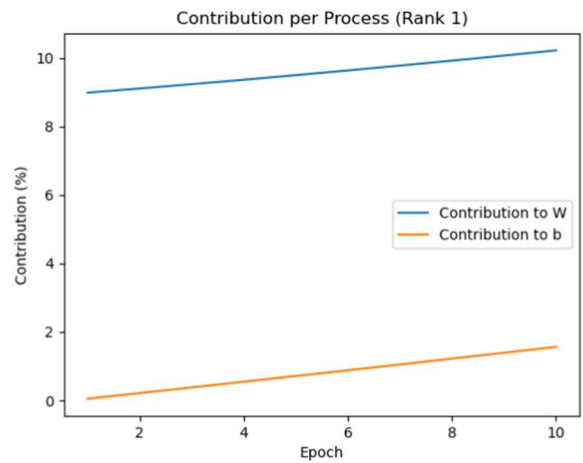
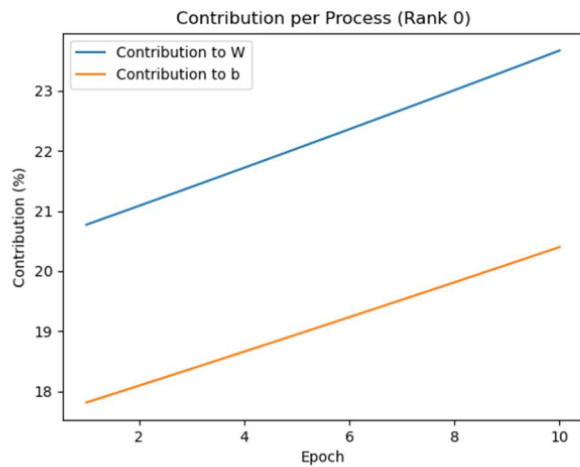
Sync time

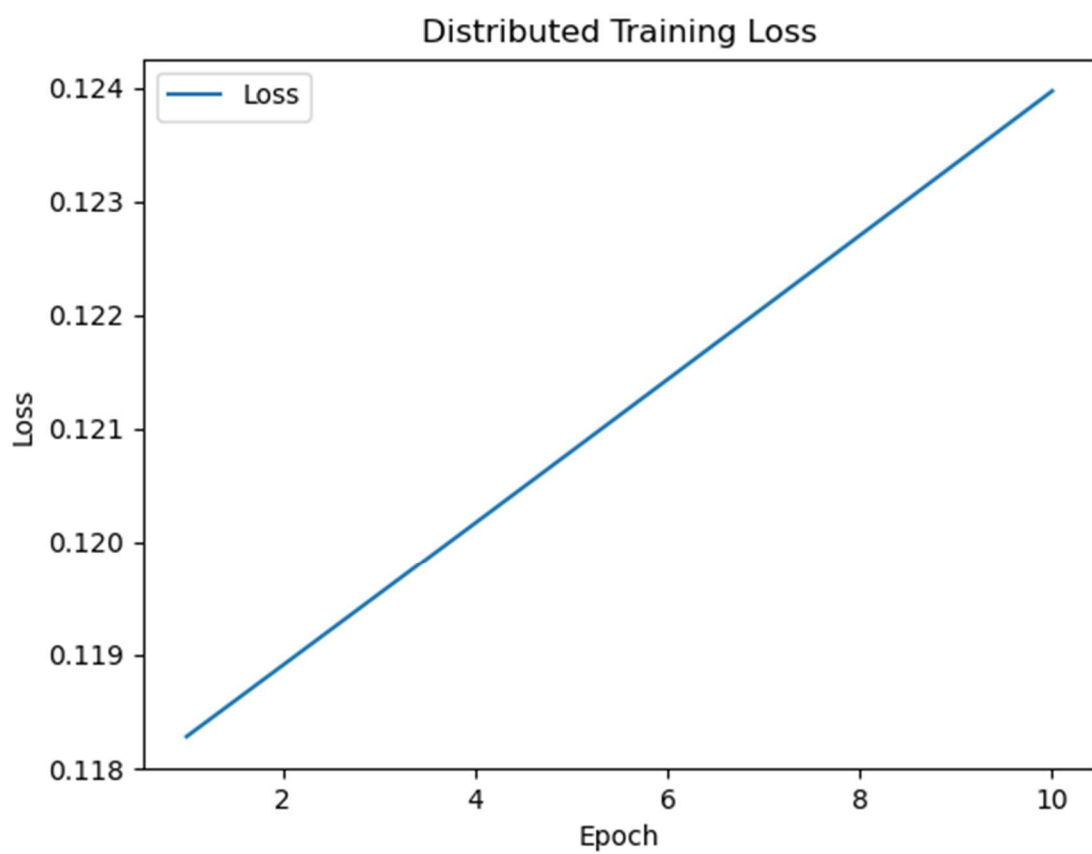
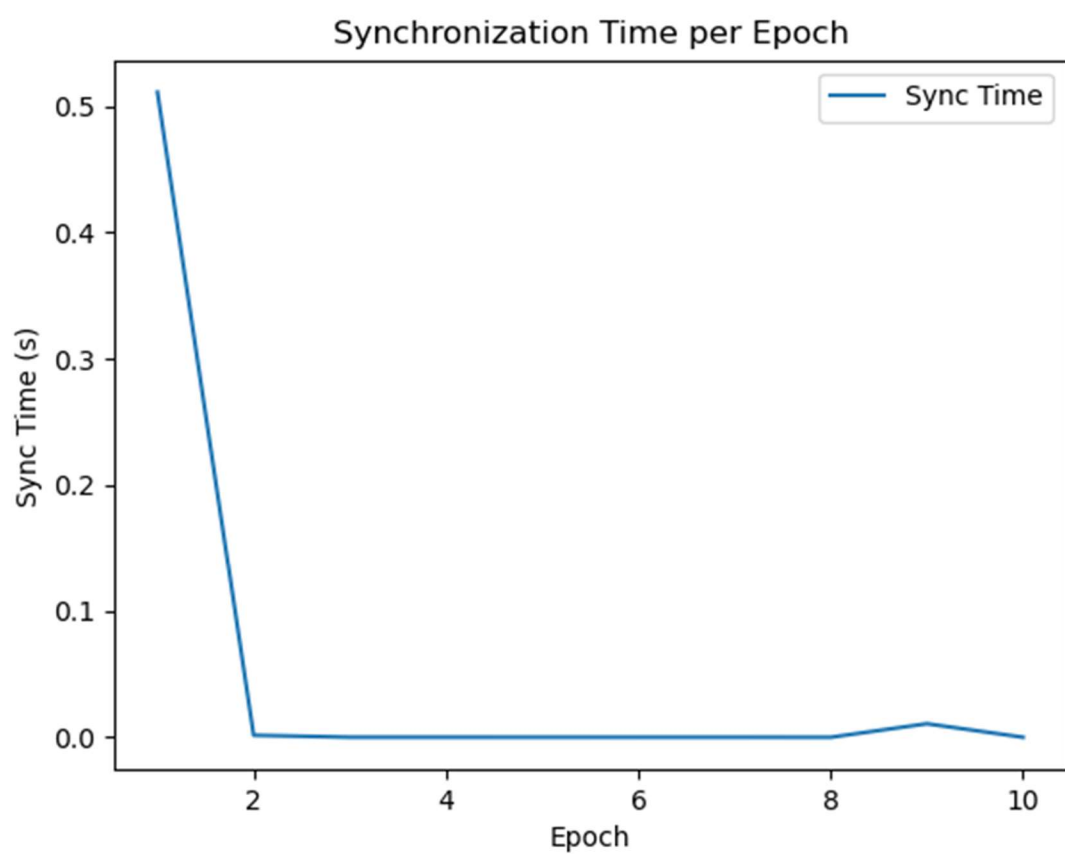
Epoch, Sync Time (s)

1, 0.285622
2, 0.013190
3, 0.000103
4, 0.000067
5, 0.000123
6, 0.000078
7, 0.000064
8, 0.000087
9, 0.000030
10, 0.000069

Contributions for rank 0

Epoch 1, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 2, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 3, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 4, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 5, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 6, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 7, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 8, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 9, Process 0, W Contribution = 23.67%, b Contribution = 20.40%
Epoch 10, Process 0, W Contribution = 23.67%, b Contribution = 20.40%





4. Interpretarea rezultatelor

Rezultatele obținute din experiment demonstrează o funcționare corectă a procesului distribuit pe mai multe nuclee CPU. Procesele rulează pe mai multe noduri și raportează valori diferite pentru gradientul mediu al WW și bb. De exemplu, gradientele procesului 2 sunt mai mari și scad treptat, în timp ce procesele 1 și 3 raportează valori mai mici sau negative, ceea ce reflectă contribuțiile lor diferite în procesul de optimizare.

Rezultatele evidențiază un comportament specific în ceea ce privește sincronizarea și contribuția fiecărui proces. Timpul de sincronizare între procese scade dramatic după prima epocă, unde atinge un vârf de 0.285 secunde, și devine aproape neglijabil în epocile următoare, cu valori sub 0.0001 secunde. Acest comportament indică faptul că inițializarea comunicării și sincronizarea inițială între procese necesită un overhead semnificativ, dar după stabilirea conexiunilor, sincronizarea gradientelor este extrem de eficientă, cu variații minime între epoci.

În ceea ce privește contribuțiile proceselor la gradientul global, se observă un dezechilibru semnificativ, în special pentru procesul 2, care contribuie disproporționat la actualizarea gradientelor, cu valori ce depășesc 110% pentru parametrii WW și bb. Acest lucru sugerează un dezechilibru în datele distribuite proceselor, unde procesul 2 primește date care au un impact mult mai mare asupra gradientelor. În contrast, procesele 0, 1 și 3 contribuie în mod constant cu aproximativ 20-23%, 9-10%, respectiv 12% la gradientele globale.

Parametrii finali ai modelului (WW și bb) au fost calculați și salvați, dar pierderea crescândă sugerează că acești parametri nu sunt optimi pentru problema dată. Aceasta poate fi o consecință a unui dezechilibru în distribuția datelor între procese sau a unei erori în sincronizarea gradientelor prin MPI.Allreduce.

În alta ordine de idei toate procesele își raportează corect activitatea pe nodurile respective, confirmând funcționarea sincronizării între procese.

În concluzie, experimentul demonstrează cu succes funcționarea antrenamentului distribuit pe 4 nuclee deoarece se observa că fiecare proces își primește și procesează corect partea sa din date, iar sincronizarea între procese funcționează conform așteptărilor, deoarece după apelul Allreduce, fiecare proces primește suma gradientelor calculate de toate procesele și împarte rezultatul în mod uniform (medie).. Fiecare proces calculează local gradientele pentru WW și bb, iar sincronizarea prin MPI.Allreduce combină contribuțiile fiecărui proces pentru a obține un model global actualizat.