

Training

June 10, 2024

```
[10]: import torch
import matplotlib.pyplot as plt
from model import GPT
from trainer import Trainer
from utils import set_seed
import json
from TransformerDataset import TransformerTrainDataset

DEBUG = False # Set to True to enable debugging

print('starting ....')

# Configuration
config = {
    'num_samples': 10000000,
    'input_dim': 20,
    'function_class': 'linear',
    'noise_std': 0,
    'model_type': 'standard',
    'batch_size': 64,
    'min_context_size': 1,
    'max_context_size': 40,
    'max_iters': 100000,
    'learning_rate': 10e-4,
    'num_workers': 0
}

# Create the TransformerTrainDataset
transformer_train_dataset = TransformerTrainDataset(
    num_samples=config['num_samples'],
    input_dim=config['input_dim'], max_context_size=config['max_context_size'],
    function_class=config['function_class'], noise_std=config['noise_std'])

# Debug TransformerTrainDataset
if DEBUG:
    print("TransformerTrainDataset samples:")
    for i in range(3):
```

```

        inputs, targets = transformer_train_dataset[i]
        print(f"Sample {i} - Inputs: {inputs.numpy()}, Targets: {targets.
        ↪numpy()}")

# Model configuration
set_seed(42)

model_config = GPT.get_default_config()
model_config.model_type = config['model_type']
model_config.vocab_size = config['input_dim'] + 1 # Adding 1 for the target,
        ↪dimension
model_config.block_size = (2 * config['max_context_size']) *
        ↪config['input_dim'] # Updated block_size based on max context size
model_config.input_dim = config['input_dim'] # Set the input dimension

# Create the model
model = GPT(model_config)
print("number of parameters: %.2fM" % (sum(p.numel() for p in model.
        ↪parameters()) / 1e6,))

# Training configuration
train_config = Trainer.get_default_config()
train_config.learning_rate = config['learning_rate']
train_config.max_iters = config['max_iters']
train_config.num_workers = config['num_workers']
train_config.batch_size = config['batch_size']

# Create the trainer
trainer = Trainer(train_config, model, transformer_train_dataset)
print("running on device", trainer.device)

# Initialize weights for tracking
initial_scalar_head_weight = model.scalar_head.weight.clone().detach()
initial_wte_weight = model.transformer.wte.weight.clone().detach()

# Lists to store weights and losses
trainer.scalar_head_weights = []
trainer.wte_weights = []
trainer.train_losses = []

# Define a callback for batch end to print training status and debug weights
def batch_end_callback(trainer):
    current_scalar_head_weight = model.scalar_head.weight.clone().detach()
    current_wte_weight = model.transformer.wte.weight.clone().detach()

    scalar_head_weight_changed = not torch.equal(initial_scalar_head_weight,
        ↪current_scalar_head_weight)

```

```

wte_weight_changed = not torch.equal(initial_wte_weight, current_wte_weight)

# Store weights for plotting
trainer.scalar_head_weights.append(current_scalar_head_weight.cpu().numpy())
trainer.wte_weights.append(current_wte_weight.cpu().numpy())
trainer.train_losses.append(trainer.loss.item())

if trainer.iter_num % 100 == 0:
    print(f"iter_dt {trainer.iter_dt * 1000:.2f}ms; iter {trainer.iter_num}:
    ↪ train loss {trainer.loss.item():.5f}; Scalar Head Weights Changed:
    ↪ {scalar_head_weight_changed}; WTE Weights Changed: {wte_weight_changed}")

# Update the initial weights
initial_scalar_head_weight.copy_(current_scalar_head_weight)
initial_wte_weight.copy_(current_wte_weight)

trainer.set_callback('on_batch_end', batch_end_callback)

# Run the training
trainer.run()

# Save the model and configurations
torch.save(model.state_dict(), 'trained_model.pth')
with open('config.json', 'w') as f:
    json.dump(config, f)
print("Model and configurations saved.")

# Plot the stored weight values
plt.figure(figsize=(12, 6))
for i in range(min(len(trainer.scalar_head_weights[0].flatten()), 5)): # Plot
    ↪ up to 5 scalar head weights
    plt.plot([w.flatten()[i] for w in trainer.scalar_head_weights],
    ↪ label=f'Scalar Head Weight {i}')
plt.xlabel('Iterations')
plt.ylabel('Weight Value')
plt.title('Scalar Head Weights Over Time')
plt.legend()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
for i in range(min(len(trainer.wte_weights[0].flatten()), 5)): # Plot up to 5
    ↪ wte weights
    plt.plot([w.flatten()[i] for w in trainer.wte_weights], label=f'WTE Weight
    ↪ {i}')
plt.xlabel('Iterations')
plt.ylabel('Weight Value')

```

```

plt.title('Word Token Embedding Weights Over Time')
plt.legend()
plt.grid(True)
plt.show()

# Plot the training loss
plt.figure(figsize=(12, 6))
plt.plot(trainer.train_losses, label='Training Loss')
plt.xlabel('Iterations')
plt.ylabel('Loss')
plt.title('Training Loss Over Time')
plt.legend()
plt.grid(True)
plt.show()

```

```

starting ...
number of parameters: 9.55M
number of parameters: 9.55M
running on device cuda
running on device cuda
iter_dt 0.00ms; iter 0: train loss 19.52942; Scalar Head Weights Changed: True;
WTE Weights Changed: True
iter_dt 144.69ms; iter 100: train loss 0.62856; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 144.89ms; iter 200: train loss 0.69854; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.16ms; iter 300: train loss 0.74673; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.33ms; iter 400: train loss 0.66965; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.62ms; iter 500: train loss 0.74123; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.45ms; iter 600: train loss 0.53343; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.17ms; iter 700: train loss 0.98147; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.62ms; iter 800: train loss 0.47394; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.39ms; iter 900: train loss 0.89346; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 144.68ms; iter 1000: train loss 0.48176; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.77ms; iter 1100: train loss 0.63574; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 144.95ms; iter 1200: train loss 0.51402; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 144.72ms; iter 1300: train loss 0.49737; Scalar Head Weights Changed:

```

True; WTE Weights Changed: True
iter_dt 145.11ms; iter 1400: train loss 0.53782; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.21ms; iter 1500: train loss 0.57458; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.66ms; iter 1600: train loss 0.52448; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.67ms; iter 1700: train loss 0.67910; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.63ms; iter 1800: train loss 0.63355; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 1900: train loss 0.97292; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.82ms; iter 2000: train loss 0.53982; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.47ms; iter 2100: train loss 0.70489; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.79ms; iter 2200: train loss 0.57215; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.74ms; iter 2300: train loss 0.33587; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.55ms; iter 2400: train loss 0.55425; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.39ms; iter 2500: train loss 0.46562; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.40ms; iter 2600: train loss 0.62389; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.17ms; iter 2700: train loss 0.57808; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.20ms; iter 2800: train loss 0.41227; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.71ms; iter 2900: train loss 0.59385; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.02ms; iter 3000: train loss 0.38178; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.19ms; iter 3100: train loss 0.53353; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.56ms; iter 3200: train loss 0.36764; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.00ms; iter 3300: train loss 0.37275; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.99ms; iter 3400: train loss 0.43256; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.36ms; iter 3500: train loss 0.58875; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.93ms; iter 3600: train loss 0.60638; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.46ms; iter 3700: train loss 0.62003; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 146.41ms; iter 3800: train loss 0.56051; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.28ms; iter 3900: train loss 0.38736; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.44ms; iter 4000: train loss 0.38032; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.40ms; iter 4100: train loss 0.39190; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.08ms; iter 4200: train loss 0.35043; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.41ms; iter 4300: train loss 0.55084; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.01ms; iter 4400: train loss 0.53248; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.06ms; iter 4500: train loss 0.55628; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.95ms; iter 4600: train loss 0.53651; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.94ms; iter 4700: train loss 0.46079; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.14ms; iter 4800: train loss 0.60207; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.00ms; iter 4900: train loss 0.46669; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.75ms; iter 5000: train loss 0.51970; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.77ms; iter 5100: train loss 0.39480; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.05ms; iter 5200: train loss 0.42524; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.79ms; iter 5300: train loss 0.61803; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.41ms; iter 5400: train loss 0.47270; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.53ms; iter 5500: train loss 0.59064; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.25ms; iter 5600: train loss 0.55647; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.79ms; iter 5700: train loss 0.57546; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.17ms; iter 5800: train loss 0.53681; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.40ms; iter 5900: train loss 0.63289; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.12ms; iter 6000: train loss 0.42247; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.50ms; iter 6100: train loss 0.54294; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 145.96ms; iter 6200: train loss 0.56602; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.23ms; iter 6300: train loss 0.58031; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.03ms; iter 6400: train loss 0.53278; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.34ms; iter 6500: train loss 0.44871; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.33ms; iter 6600: train loss 0.44718; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.15ms; iter 6700: train loss 0.43484; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.39ms; iter 6800: train loss 0.68152; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.85ms; iter 6900: train loss 0.55592; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.71ms; iter 7000: train loss 0.44976; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.90ms; iter 7100: train loss 0.57041; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.61ms; iter 7200: train loss 0.33931; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.77ms; iter 7300: train loss 0.35603; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.23ms; iter 7400: train loss 0.46964; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.45ms; iter 7500: train loss 0.49807; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.37ms; iter 7600: train loss 0.43095; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.85ms; iter 7700: train loss 0.50491; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.54ms; iter 7800: train loss 0.49115; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.07ms; iter 7900: train loss 0.51424; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.56ms; iter 8000: train loss 0.53375; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.53ms; iter 8100: train loss 0.51111; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.28ms; iter 8200: train loss 0.38158; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.57ms; iter 8300: train loss 0.43662; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.67ms; iter 8400: train loss 0.43506; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.20ms; iter 8500: train loss 0.49694; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 146.78ms; iter 8600: train loss 0.49603; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.79ms; iter 8700: train loss 0.48201; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.02ms; iter 8800: train loss 0.43923; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.20ms; iter 8900: train loss 0.37950; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.26ms; iter 9000: train loss 0.38280; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.75ms; iter 9100: train loss 0.68093; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.22ms; iter 9200: train loss 0.73557; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.35ms; iter 9300: train loss 0.43572; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.91ms; iter 9400: train loss 0.36230; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.12ms; iter 9500: train loss 0.52931; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.57ms; iter 9600: train loss 0.44004; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.72ms; iter 9700: train loss 0.47354; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.72ms; iter 9800: train loss 0.58448; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.68ms; iter 9900: train loss 0.52540; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.03ms; iter 10000: train loss 0.55499; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.15ms; iter 10100: train loss 0.52057; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.77ms; iter 10200: train loss 0.47404; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.16ms; iter 10300: train loss 0.59810; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.64ms; iter 10400: train loss 0.59210; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.41ms; iter 10500: train loss 0.50415; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.79ms; iter 10600: train loss 0.43100; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.37ms; iter 10700: train loss 0.65211; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.52ms; iter 10800: train loss 0.53678; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.00ms; iter 10900: train loss 0.44417; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 146.64ms; iter 11000: train loss 0.48874; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.25ms; iter 11100: train loss 0.37909; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.29ms; iter 11200: train loss 0.50516; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.37ms; iter 11300: train loss 0.48692; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.93ms; iter 11400: train loss 0.42601; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.56ms; iter 11500: train loss 0.41799; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.79ms; iter 11600: train loss 0.63287; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.03ms; iter 11700: train loss 0.55345; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.55ms; iter 11800: train loss 0.46742; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.54ms; iter 11900: train loss 0.44052; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.76ms; iter 12000: train loss 0.50567; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.66ms; iter 12100: train loss 0.54433; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.82ms; iter 12200: train loss 0.57733; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.26ms; iter 12300: train loss 0.52041; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.53ms; iter 12400: train loss 0.46209; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.11ms; iter 12500: train loss 0.42549; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.19ms; iter 12600: train loss 0.46260; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.46ms; iter 12700: train loss 0.50839; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.50ms; iter 12800: train loss 0.58660; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.18ms; iter 12900: train loss 0.62930; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.65ms; iter 13000: train loss 0.46134; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.89ms; iter 13100: train loss 0.43318; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.58ms; iter 13200: train loss 0.33272; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.55ms; iter 13300: train loss 0.43130; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.05ms; iter 13400: train loss 0.47379; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.75ms; iter 13500: train loss 0.52826; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.52ms; iter 13600: train loss 0.51625; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.18ms; iter 13700: train loss 0.59831; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.49ms; iter 13800: train loss 0.54090; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.24ms; iter 13900: train loss 0.53087; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.68ms; iter 14000: train loss 0.71684; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.62ms; iter 14100: train loss 0.56051; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.73ms; iter 14200: train loss 0.74299; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.06ms; iter 14300: train loss 0.56572; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.57ms; iter 14400: train loss 0.60779; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.14ms; iter 14500: train loss 0.45345; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.62ms; iter 14600: train loss 0.47487; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.95ms; iter 14700: train loss 0.59369; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.90ms; iter 14800: train loss 0.56569; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.02ms; iter 14900: train loss 0.47415; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.12ms; iter 15000: train loss 0.63294; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.83ms; iter 15100: train loss 0.50564; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.77ms; iter 15200: train loss 0.34854; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.96ms; iter 15300: train loss 0.54524; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.53ms; iter 15400: train loss 0.56384; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 157.11ms; iter 15500: train loss 0.47811; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.13ms; iter 15600: train loss 0.43503; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 145.96ms; iter 15700: train loss 0.52174; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 146.75ms; iter 15800: train loss 0.58439; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.81ms; iter 15900: train loss 0.38056; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.68ms; iter 16000: train loss 0.68189; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.01ms; iter 16100: train loss 0.56074; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.95ms; iter 16200: train loss 0.45823; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.68ms; iter 16300: train loss 0.53677; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.00ms; iter 16400: train loss 0.46328; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.09ms; iter 16500: train loss 0.52275; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.03ms; iter 16600: train loss 0.54705; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.59ms; iter 16700: train loss 0.55521; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.10ms; iter 16800: train loss 0.42525; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.48ms; iter 16900: train loss 0.44492; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.29ms; iter 17000: train loss 0.56726; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.76ms; iter 17100: train loss 0.41379; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.93ms; iter 17200: train loss 0.51332; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 158.16ms; iter 17300: train loss 0.61113; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.76ms; iter 17400: train loss 0.53722; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.96ms; iter 17500: train loss 0.60295; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.65ms; iter 17600: train loss 0.39602; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.68ms; iter 17700: train loss 0.53644; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.57ms; iter 17800: train loss 0.65315; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.00ms; iter 17900: train loss 0.47331; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.93ms; iter 18000: train loss 0.57458; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.54ms; iter 18100: train loss 0.71181; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 146.42ms; iter 18200: train loss 0.52434; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.65ms; iter 18300: train loss 0.52712; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.02ms; iter 18400: train loss 0.52704; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.73ms; iter 18500: train loss 0.38541; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.93ms; iter 18600: train loss 0.62611; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.33ms; iter 18700: train loss 0.47373; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 244.11ms; iter 18800: train loss 0.51647; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.45ms; iter 18900: train loss 0.53199; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.93ms; iter 19000: train loss 0.49867; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.53ms; iter 19100: train loss 0.46795; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.40ms; iter 19200: train loss 0.62225; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.03ms; iter 19300: train loss 0.52505; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.79ms; iter 19400: train loss 0.56411; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.22ms; iter 19500: train loss 0.59101; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.80ms; iter 19600: train loss 0.51708; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.58ms; iter 19700: train loss 0.60833; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.62ms; iter 19800: train loss 0.63614; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.40ms; iter 19900: train loss 0.47509; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.99ms; iter 20000: train loss 0.45207; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.14ms; iter 20100: train loss 0.50975; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.59ms; iter 20200: train loss 0.47424; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.09ms; iter 20300: train loss 0.44928; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.71ms; iter 20400: train loss 0.53301; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.73ms; iter 20500: train loss 0.39152; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.46ms; iter 20600: train loss 0.53159; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.91ms; iter 20700: train loss 0.57029; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.15ms; iter 20800: train loss 0.41330; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.80ms; iter 20900: train loss 0.47160; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.91ms; iter 21000: train loss 0.54067; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.97ms; iter 21100: train loss 0.50624; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.24ms; iter 21200: train loss 0.70877; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.68ms; iter 21300: train loss 0.62020; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.59ms; iter 21400: train loss 0.51279; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.35ms; iter 21500: train loss 0.54854; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.76ms; iter 21600: train loss 0.45866; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.47ms; iter 21700: train loss 0.48131; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.56ms; iter 21800: train loss 0.57958; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.93ms; iter 21900: train loss 0.59612; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.22ms; iter 22000: train loss 0.54223; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.67ms; iter 22100: train loss 0.54538; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.61ms; iter 22200: train loss 0.49735; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.32ms; iter 22300: train loss 0.38645; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.65ms; iter 22400: train loss 0.82830; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.60ms; iter 22500: train loss 0.33705; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.33ms; iter 22600: train loss 0.41276; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.98ms; iter 22700: train loss 0.51997; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.54ms; iter 22800: train loss 0.50562; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.55ms; iter 22900: train loss 0.48830; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.26ms; iter 23000: train loss 0.43096; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.68ms; iter 23100: train loss 0.34409; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.35ms; iter 23200: train loss 0.43720; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.47ms; iter 23300: train loss 0.59813; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.65ms; iter 23400: train loss 0.51463; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.98ms; iter 23500: train loss 0.34379; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 151.99ms; iter 23600: train loss 0.50737; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.35ms; iter 23700: train loss 0.55815; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.32ms; iter 23800: train loss 0.44088; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.29ms; iter 23900: train loss 0.44710; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.49ms; iter 24000: train loss 0.58429; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.89ms; iter 24100: train loss 0.48986; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.59ms; iter 24200: train loss 0.61183; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.33ms; iter 24300: train loss 0.61147; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.07ms; iter 24400: train loss 0.44933; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.17ms; iter 24500: train loss 0.48753; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.23ms; iter 24600: train loss 0.39008; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.72ms; iter 24700: train loss 0.47194; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.98ms; iter 24800: train loss 0.46595; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.68ms; iter 24900: train loss 0.54799; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.62ms; iter 25000: train loss 0.56965; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.46ms; iter 25100: train loss 0.43826; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.83ms; iter 25200: train loss 0.49249; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.11ms; iter 25300: train loss 0.46783; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.18ms; iter 25400: train loss 0.60475; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.97ms; iter 25500: train loss 0.53164; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.13ms; iter 25600: train loss 0.37976; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.95ms; iter 25700: train loss 0.52048; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.53ms; iter 25800: train loss 0.35676; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.54ms; iter 25900: train loss 0.42862; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.40ms; iter 26000: train loss 0.66270; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.61ms; iter 26100: train loss 0.39836; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 161.64ms; iter 26200: train loss 0.61173; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.07ms; iter 26300: train loss 0.40524; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.36ms; iter 26400: train loss 0.58951; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.34ms; iter 26500: train loss 0.59312; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.81ms; iter 26600: train loss 0.54574; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.27ms; iter 26700: train loss 0.39348; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.49ms; iter 26800: train loss 0.35174; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.66ms; iter 26900: train loss 0.48193; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.35ms; iter 27000: train loss 0.42175; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.63ms; iter 27100: train loss 0.70479; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.16ms; iter 27200: train loss 0.38942; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.54ms; iter 27300: train loss 0.47302; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.39ms; iter 27400: train loss 0.51502; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.75ms; iter 27500: train loss 0.61954; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 27600: train loss 0.36003; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.01ms; iter 27700: train loss 0.71835; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.06ms; iter 27800: train loss 0.53590; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.89ms; iter 27900: train loss 0.46157; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.22ms; iter 28000: train loss 0.55709; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.26ms; iter 28100: train loss 0.48947; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.33ms; iter 28200: train loss 0.52502; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.35ms; iter 28300: train loss 0.42972; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.18ms; iter 28400: train loss 0.42100; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.04ms; iter 28500: train loss 0.56297; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.35ms; iter 28600: train loss 0.44927; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.01ms; iter 28700: train loss 0.37663; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.85ms; iter 28800: train loss 0.39080; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.03ms; iter 28900: train loss 0.54649; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.79ms; iter 29000: train loss 0.51820; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.98ms; iter 29100: train loss 0.35039; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.38ms; iter 29200: train loss 0.43255; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.60ms; iter 29300: train loss 0.65895; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.08ms; iter 29400: train loss 0.47345; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.86ms; iter 29500: train loss 0.48606; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.65ms; iter 29600: train loss 0.41618; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.62ms; iter 29700: train loss 0.42806; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 146.80ms; iter 29800: train loss 0.57599; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.18ms; iter 29900: train loss 0.55792; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.67ms; iter 30000: train loss 0.59178; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.78ms; iter 30100: train loss 0.54502; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.44ms; iter 30200: train loss 0.61523; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 153.94ms; iter 30300: train loss 0.44703; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.77ms; iter 30400: train loss 0.48078; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.68ms; iter 30500: train loss 0.43281; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.01ms; iter 30600: train loss 0.42523; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.99ms; iter 30700: train loss 0.35911; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.66ms; iter 30800: train loss 0.50913; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.81ms; iter 30900: train loss 0.50656; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.62ms; iter 31000: train loss 0.37653; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.68ms; iter 31100: train loss 0.41658; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.58ms; iter 31200: train loss 0.63172; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.95ms; iter 31300: train loss 0.51059; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 31400: train loss 0.67105; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.67ms; iter 31500: train loss 0.39611; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.54ms; iter 31600: train loss 0.38770; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.26ms; iter 31700: train loss 0.45604; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.78ms; iter 31800: train loss 0.49818; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.56ms; iter 31900: train loss 0.50968; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.58ms; iter 32000: train loss 0.36213; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.13ms; iter 32100: train loss 0.57279; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.83ms; iter 32200: train loss 0.43500; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.22ms; iter 32300: train loss 0.55462; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.81ms; iter 32400: train loss 0.47479; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.42ms; iter 32500: train loss 0.41635; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.91ms; iter 32600: train loss 0.51185; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.80ms; iter 32700: train loss 0.55305; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.87ms; iter 32800: train loss 0.55632; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.02ms; iter 32900: train loss 0.42481; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 152.06ms; iter 33000: train loss 0.46423; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.70ms; iter 33100: train loss 0.64359; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.85ms; iter 33200: train loss 0.63284; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.15ms; iter 33300: train loss 0.38596; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.42ms; iter 33400: train loss 0.46048; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.98ms; iter 33500: train loss 0.36515; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.86ms; iter 33600: train loss 0.43926; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.71ms; iter 33700: train loss 0.44806; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.61ms; iter 33800: train loss 0.56925; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.23ms; iter 33900: train loss 0.59257; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.88ms; iter 34000: train loss 0.40141; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.24ms; iter 34100: train loss 0.36822; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.94ms; iter 34200: train loss 0.36599; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.53ms; iter 34300: train loss 0.66439; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.12ms; iter 34400: train loss 0.52810; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.18ms; iter 34500: train loss 0.34788; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.63ms; iter 34600: train loss 0.30023; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.56ms; iter 34700: train loss 0.57691; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.41ms; iter 34800: train loss 0.41627; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.75ms; iter 34900: train loss 0.51344; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.54ms; iter 35000: train loss 0.36907; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.83ms; iter 35100: train loss 0.52811; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 35200: train loss 0.58462; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.68ms; iter 35300: train loss 0.49299; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.65ms; iter 35400: train loss 0.41860; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.62ms; iter 35500: train loss 0.47599; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.56ms; iter 35600: train loss 0.43320; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.11ms; iter 35700: train loss 0.49346; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.64ms; iter 35800: train loss 0.50392; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.74ms; iter 35900: train loss 0.32192; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.42ms; iter 36000: train loss 0.60932; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.95ms; iter 36100: train loss 0.36499; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.04ms; iter 36200: train loss 0.52634; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 36300: train loss 0.34598; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.94ms; iter 36400: train loss 0.56389; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 36500: train loss 0.43583; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.80ms; iter 36600: train loss 0.62469; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.52ms; iter 36700: train loss 0.69948; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.03ms; iter 36800: train loss 0.45463; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.32ms; iter 36900: train loss 0.41849; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.11ms; iter 37000: train loss 0.42365; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.33ms; iter 37100: train loss 0.54803; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.59ms; iter 37200: train loss 0.43608; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.00ms; iter 37300: train loss 0.46428; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.92ms; iter 37400: train loss 0.36384; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.55ms; iter 37500: train loss 0.48128; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.37ms; iter 37600: train loss 0.47004; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.42ms; iter 37700: train loss 0.54939; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.34ms; iter 37800: train loss 0.38734; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.63ms; iter 37900: train loss 0.44278; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.27ms; iter 38000: train loss 0.35103; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.82ms; iter 38100: train loss 0.34842; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.99ms; iter 38200: train loss 0.44005; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 152.17ms; iter 38300: train loss 0.52717; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.79ms; iter 38400: train loss 0.51172; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.92ms; iter 38500: train loss 0.58026; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.70ms; iter 38600: train loss 0.56425; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.64ms; iter 38700: train loss 0.32486; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.61ms; iter 38800: train loss 0.50677; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.42ms; iter 38900: train loss 0.41568; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.33ms; iter 39000: train loss 0.44064; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.92ms; iter 39100: train loss 0.64083; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.01ms; iter 39200: train loss 0.61265; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.13ms; iter 39300: train loss 0.46056; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.71ms; iter 39400: train loss 0.58201; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.66ms; iter 39500: train loss 0.43280; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.89ms; iter 39600: train loss 0.29154; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.98ms; iter 39700: train loss 0.53998; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 147.99ms; iter 39800: train loss 0.55596; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.27ms; iter 39900: train loss 0.60456; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.19ms; iter 40000: train loss 0.35558; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.95ms; iter 40100: train loss 0.32074; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.06ms; iter 40200: train loss 0.68110; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.21ms; iter 40300: train loss 0.56357; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.21ms; iter 40400: train loss 0.53903; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.87ms; iter 40500: train loss 0.63371; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.61ms; iter 40600: train loss 0.41439; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.06ms; iter 40700: train loss 0.49408; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 40800: train loss 0.53648; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.99ms; iter 40900: train loss 0.49309; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 41000: train loss 0.53356; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.69ms; iter 41100: train loss 0.46919; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 41200: train loss 0.57356; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.49ms; iter 41300: train loss 0.49406; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.13ms; iter 41400: train loss 0.51662; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 41500: train loss 0.36048; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.07ms; iter 41600: train loss 0.45980; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.38ms; iter 41700: train loss 0.46885; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.30ms; iter 41800: train loss 0.41049; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.40ms; iter 41900: train loss 0.51344; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.90ms; iter 42000: train loss 0.54439; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.78ms; iter 42100: train loss 0.48675; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.27ms; iter 42200: train loss 0.43390; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.84ms; iter 42300: train loss 0.37841; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.03ms; iter 42400: train loss 0.38996; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 42500: train loss 0.40080; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 42600: train loss 0.52465; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.12ms; iter 42700: train loss 0.39666; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.08ms; iter 42800: train loss 0.42406; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 42900: train loss 0.43559; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.04ms; iter 43000: train loss 0.58246; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.08ms; iter 43100: train loss 0.51908; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.06ms; iter 43200: train loss 0.45205; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 43300: train loss 0.48519; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.95ms; iter 43400: train loss 0.55927; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 43500: train loss 0.40361; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 43600: train loss 0.47259; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.61ms; iter 43700: train loss 0.55064; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.66ms; iter 43800: train loss 0.55282; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 43900: train loss 0.52011; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.92ms; iter 44000: train loss 0.68800; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.91ms; iter 44100: train loss 0.61782; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.17ms; iter 44200: train loss 0.40096; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.78ms; iter 44300: train loss 0.54641; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 44400: train loss 0.58583; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 44500: train loss 0.52788; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.22ms; iter 44600: train loss 0.56999; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.91ms; iter 44700: train loss 0.63061; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.03ms; iter 44800: train loss 0.49471; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 44900: train loss 0.60386; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.57ms; iter 45000: train loss 0.40636; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.21ms; iter 45100: train loss 0.53070; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 45200: train loss 0.33746; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 45300: train loss 0.44590; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.57ms; iter 45400: train loss 0.55158; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.70ms; iter 45500: train loss 0.50897; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.82ms; iter 45600: train loss 0.45632; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.79ms; iter 45700: train loss 0.44151; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.82ms; iter 45800: train loss 0.40832; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.85ms; iter 45900: train loss 0.77285; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.70ms; iter 46000: train loss 0.65746; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.88ms; iter 46100: train loss 0.44152; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.78ms; iter 46200: train loss 0.57316; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.52ms; iter 46300: train loss 0.45629; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.17ms; iter 46400: train loss 0.56514; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.42ms; iter 46500: train loss 0.47382; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.62ms; iter 46600: train loss 0.48541; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.77ms; iter 46700: train loss 0.40703; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.79ms; iter 46800: train loss 0.60111; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.17ms; iter 46900: train loss 0.48130; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.47ms; iter 47000: train loss 0.40085; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.24ms; iter 47100: train loss 0.57231; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.60ms; iter 47200: train loss 0.65153; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.11ms; iter 47300: train loss 0.41018; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 47400: train loss 0.62240; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.03ms; iter 47500: train loss 0.56924; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 47600: train loss 0.65343; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.12ms; iter 47700: train loss 0.37070; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.90ms; iter 47800: train loss 0.46094; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.77ms; iter 47900: train loss 0.49270; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.61ms; iter 48000: train loss 0.52683; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.54ms; iter 48100: train loss 0.53304; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.72ms; iter 48200: train loss 0.55932; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 48300: train loss 0.43950; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.58ms; iter 48400: train loss 0.54174; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.95ms; iter 48500: train loss 0.41392; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.36ms; iter 48600: train loss 0.42087; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 48700: train loss 0.38575; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 48800: train loss 0.23426; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 48900: train loss 0.42086; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.30ms; iter 49000: train loss 0.51701; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 49100: train loss 0.44444; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.42ms; iter 49200: train loss 0.52614; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.08ms; iter 49300: train loss 0.45664; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.57ms; iter 49400: train loss 0.60150; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 49500: train loss 0.57587; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.99ms; iter 49600: train loss 0.50673; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.06ms; iter 49700: train loss 0.42255; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 49800: train loss 0.48547; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 151.83ms; iter 49900: train loss 0.42413; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.36ms; iter 50000: train loss 0.48415; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.68ms; iter 50100: train loss 0.48927; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.70ms; iter 50200: train loss 0.64529; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.91ms; iter 50300: train loss 0.55115; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.19ms; iter 50400: train loss 0.49316; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 151.16ms; iter 50500: train loss 0.70375; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.19ms; iter 50600: train loss 0.47762; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.90ms; iter 50700: train loss 0.45869; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 50800: train loss 0.38971; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.55ms; iter 50900: train loss 0.61643; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 51000: train loss 0.65662; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.30ms; iter 51100: train loss 0.57968; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.19ms; iter 51200: train loss 0.41090; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.98ms; iter 51300: train loss 0.37478; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.49ms; iter 51400: train loss 0.45117; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.53ms; iter 51500: train loss 0.46911; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.78ms; iter 51600: train loss 0.58669; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.93ms; iter 51700: train loss 0.36598; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.75ms; iter 51800: train loss 0.48692; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 51900: train loss 0.36815; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.36ms; iter 52000: train loss 0.56316; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.62ms; iter 52100: train loss 0.32876; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 52200: train loss 0.56824; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 52300: train loss 0.52460; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 52400: train loss 0.61214; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.66ms; iter 52500: train loss 0.38842; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.44ms; iter 52600: train loss 0.46370; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.20ms; iter 52700: train loss 0.39703; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 152.61ms; iter 52800: train loss 0.63321; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.95ms; iter 52900: train loss 0.44346; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.57ms; iter 53000: train loss 0.79102; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.35ms; iter 53100: train loss 0.61556; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.73ms; iter 53200: train loss 0.46950; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 53300: train loss 0.40998; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.92ms; iter 53400: train loss 0.44302; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.54ms; iter 53500: train loss 0.46639; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.74ms; iter 53600: train loss 0.33065; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.77ms; iter 53700: train loss 0.41478; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.40ms; iter 53800: train loss 0.36891; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.33ms; iter 53900: train loss 0.52401; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.24ms; iter 54000: train loss 0.43774; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 54100: train loss 0.40136; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.43ms; iter 54200: train loss 0.63041; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.33ms; iter 54300: train loss 0.50416; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.42ms; iter 54400: train loss 0.44469; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 54500: train loss 0.53276; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 54600: train loss 0.51786; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.21ms; iter 54700: train loss 0.57496; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.76ms; iter 54800: train loss 0.52555; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.94ms; iter 54900: train loss 0.60139; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.81ms; iter 55000: train loss 0.60403; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.84ms; iter 55100: train loss 0.45145; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.69ms; iter 55200: train loss 0.38282; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.78ms; iter 55300: train loss 0.44720; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 55400: train loss 0.49705; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.56ms; iter 55500: train loss 0.50883; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.09ms; iter 55600: train loss 0.64545; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.58ms; iter 55700: train loss 0.65138; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.88ms; iter 55800: train loss 0.64672; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.27ms; iter 55900: train loss 0.51890; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 56000: train loss 0.58201; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.69ms; iter 56100: train loss 0.45013; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 56200: train loss 0.31364; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.24ms; iter 56300: train loss 0.53310; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.85ms; iter 56400: train loss 0.51335; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 56500: train loss 0.45184; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.20ms; iter 56600: train loss 0.48479; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 56700: train loss 0.64899; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.84ms; iter 56800: train loss 0.51075; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 56900: train loss 0.59805; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 57000: train loss 0.45332; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.14ms; iter 57100: train loss 0.32264; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.66ms; iter 57200: train loss 0.45746; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.87ms; iter 57300: train loss 0.34342; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.37ms; iter 57400: train loss 0.47154; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.53ms; iter 57500: train loss 0.59661; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.37ms; iter 57600: train loss 0.65366; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 57700: train loss 0.60609; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.11ms; iter 57800: train loss 0.42578; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.56ms; iter 57900: train loss 0.69630; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 58000: train loss 0.54135; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 58100: train loss 0.45115; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.62ms; iter 58200: train loss 0.34850; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.97ms; iter 58300: train loss 0.63730; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 58400: train loss 0.34770; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.19ms; iter 58500: train loss 0.67676; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 58600: train loss 0.50880; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.25ms; iter 58700: train loss 0.31389; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 58800: train loss 0.53379; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.30ms; iter 58900: train loss 0.52512; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.31ms; iter 59000: train loss 0.37002; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.90ms; iter 59100: train loss 0.55566; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.58ms; iter 59200: train loss 0.70733; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.91ms; iter 59300: train loss 0.54053; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.73ms; iter 59400: train loss 0.42633; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 59500: train loss 0.45950; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.45ms; iter 59600: train loss 0.45391; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.30ms; iter 59700: train loss 0.59483; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.44ms; iter 59800: train loss 0.54588; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 59900: train loss 0.56907; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.14ms; iter 60000: train loss 0.44067; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 60100: train loss 0.62466; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.05ms; iter 60200: train loss 0.42418; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.80ms; iter 60300: train loss 0.40214; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.67ms; iter 60400: train loss 0.40408; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.77ms; iter 60500: train loss 0.55692; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 60600: train loss 0.54190; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.60ms; iter 60700: train loss 0.40834; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 60800: train loss 0.41176; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.02ms; iter 60900: train loss 0.36867; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.93ms; iter 61000: train loss 0.38397; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 61100: train loss 0.56751; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.45ms; iter 61200: train loss 0.54415; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.09ms; iter 61300: train loss 0.36314; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.27ms; iter 61400: train loss 0.56237; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.12ms; iter 61500: train loss 0.53382; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.06ms; iter 61600: train loss 0.47631; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.12ms; iter 61700: train loss 0.47018; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.97ms; iter 61800: train loss 0.48251; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 61900: train loss 0.36267; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.03ms; iter 62000: train loss 0.54516; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.31ms; iter 62100: train loss 0.45117; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.22ms; iter 62200: train loss 0.47876; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 153.27ms; iter 62300: train loss 0.75029; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.04ms; iter 62400: train loss 0.57896; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 62500: train loss 0.46039; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.13ms; iter 62600: train loss 0.55726; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 62700: train loss 0.46540; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.35ms; iter 62800: train loss 0.40586; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.92ms; iter 62900: train loss 0.50400; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 63000: train loss 0.55417; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.18ms; iter 63100: train loss 0.65757; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.82ms; iter 63200: train loss 0.41744; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.27ms; iter 63300: train loss 0.64955; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.35ms; iter 63400: train loss 0.47428; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 63500: train loss 0.66626; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.36ms; iter 63600: train loss 0.44296; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 63700: train loss 0.58406; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.41ms; iter 63800: train loss 0.45595; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 152.51ms; iter 63900: train loss 0.45529; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.21ms; iter 64000: train loss 0.42235; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.65ms; iter 64100: train loss 0.44469; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.99ms; iter 64200: train loss 0.48875; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.89ms; iter 64300: train loss 0.45890; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.99ms; iter 64400: train loss 0.52420; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.32ms; iter 64500: train loss 0.51009; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.57ms; iter 64600: train loss 0.54131; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 64700: train loss 0.48654; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.57ms; iter 64800: train loss 0.45285; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 64900: train loss 0.29194; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.67ms; iter 65000: train loss 0.51215; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.84ms; iter 65100: train loss 0.64356; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.82ms; iter 65200: train loss 0.47960; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.31ms; iter 65300: train loss 0.49811; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.96ms; iter 65400: train loss 0.51557; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.78ms; iter 65500: train loss 0.54949; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 153.01ms; iter 65600: train loss 0.32534; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 65700: train loss 0.44581; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.70ms; iter 65800: train loss 0.35586; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.56ms; iter 65900: train loss 0.68384; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 66000: train loss 0.43979; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.17ms; iter 66100: train loss 0.48710; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.38ms; iter 66200: train loss 0.37920; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 66300: train loss 0.57530; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.66ms; iter 66400: train loss 0.45651; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.70ms; iter 66500: train loss 0.36347; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.14ms; iter 66600: train loss 0.44594; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.35ms; iter 66700: train loss 0.43262; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.31ms; iter 66800: train loss 0.41019; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 66900: train loss 0.46235; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 67000: train loss 0.48932; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.65ms; iter 67100: train loss 0.30680; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 67200: train loss 0.62116; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.79ms; iter 67300: train loss 0.66550; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.51ms; iter 67400: train loss 0.41280; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.62ms; iter 67500: train loss 0.45105; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 67600: train loss 0.62239; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 67700: train loss 0.46419; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.33ms; iter 67800: train loss 0.47751; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.70ms; iter 67900: train loss 0.73504; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.68ms; iter 68000: train loss 0.50743; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.90ms; iter 68100: train loss 0.39677; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.88ms; iter 68200: train loss 0.55191; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.21ms; iter 68300: train loss 0.45912; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.83ms; iter 68400: train loss 0.47188; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.73ms; iter 68500: train loss 0.36961; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.28ms; iter 68600: train loss 0.44127; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.81ms; iter 68700: train loss 0.38613; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.80ms; iter 68800: train loss 0.52020; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 155.31ms; iter 68900: train loss 0.37727; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 69000: train loss 0.64066; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.57ms; iter 69100: train loss 0.58630; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.32ms; iter 69200: train loss 0.48146; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.98ms; iter 69300: train loss 0.52028; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.69ms; iter 69400: train loss 0.41447; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.27ms; iter 69500: train loss 0.39496; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.24ms; iter 69600: train loss 0.43877; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 69700: train loss 0.37427; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.05ms; iter 69800: train loss 0.44166; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 159.47ms; iter 69900: train loss 0.48109; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.55ms; iter 70000: train loss 0.46092; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.62ms; iter 70100: train loss 0.44659; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.45ms; iter 70200: train loss 0.52980; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.85ms; iter 70300: train loss 0.49812; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.65ms; iter 70400: train loss 0.53603; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.44ms; iter 70500: train loss 0.43183; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.51ms; iter 70600: train loss 0.49798; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.45ms; iter 70700: train loss 0.67490; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 70800: train loss 0.46206; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.31ms; iter 70900: train loss 0.59691; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.04ms; iter 71000: train loss 0.40905; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.13ms; iter 71100: train loss 0.36173; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 71200: train loss 0.47204; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.67ms; iter 71300: train loss 0.69139; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 71400: train loss 0.49518; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.61ms; iter 71500: train loss 0.43985; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.97ms; iter 71600: train loss 0.38024; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.61ms; iter 71700: train loss 0.38245; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.57ms; iter 71800: train loss 0.46083; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.13ms; iter 71900: train loss 0.51361; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.17ms; iter 72000: train loss 0.40415; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.02ms; iter 72100: train loss 0.49583; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.95ms; iter 72200: train loss 0.46609; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.15ms; iter 72300: train loss 0.49320; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 72400: train loss 0.40499; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.79ms; iter 72500: train loss 0.40608; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.49ms; iter 72600: train loss 0.58856; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.64ms; iter 72700: train loss 0.42370; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.32ms; iter 72800: train loss 0.48333; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.37ms; iter 72900: train loss 0.55218; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 73000: train loss 0.49799; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 73100: train loss 0.49173; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 73200: train loss 0.58049; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.15ms; iter 73300: train loss 0.40708; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.40ms; iter 73400: train loss 0.42961; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.42ms; iter 73500: train loss 0.51296; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.82ms; iter 73600: train loss 0.45294; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 73700: train loss 0.36199; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.11ms; iter 73800: train loss 0.48299; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.27ms; iter 73900: train loss 0.55879; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.70ms; iter 74000: train loss 0.47099; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 74100: train loss 0.38786; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.78ms; iter 74200: train loss 0.43310; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.14ms; iter 74300: train loss 0.49104; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.91ms; iter 74400: train loss 0.42735; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 74500: train loss 0.48566; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.58ms; iter 74600: train loss 0.48184; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.35ms; iter 74700: train loss 0.38843; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.58ms; iter 74800: train loss 0.48694; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.61ms; iter 74900: train loss 0.52554; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.55ms; iter 75000: train loss 0.59092; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.79ms; iter 75100: train loss 0.65007; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.17ms; iter 75200: train loss 0.62987; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.65ms; iter 75300: train loss 0.36331; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 75400: train loss 0.57212; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 75500: train loss 0.33220; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.61ms; iter 75600: train loss 0.47725; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 75700: train loss 0.65107; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.43ms; iter 75800: train loss 0.48643; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.52ms; iter 75900: train loss 0.53070; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 76000: train loss 0.46168; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.28ms; iter 76100: train loss 0.51315; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.60ms; iter 76200: train loss 0.47143; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 76300: train loss 0.62620; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.51ms; iter 76400: train loss 0.56989; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.80ms; iter 76500: train loss 0.47482; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.70ms; iter 76600: train loss 0.62957; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.49ms; iter 76700: train loss 0.59962; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 76800: train loss 0.52682; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.73ms; iter 76900: train loss 0.61584; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 77000: train loss 0.34515; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.72ms; iter 77100: train loss 0.62000; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.12ms; iter 77200: train loss 0.49525; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 77300: train loss 0.57703; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.85ms; iter 77400: train loss 0.39534; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 77500: train loss 0.45053; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 77600: train loss 0.63707; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.01ms; iter 77700: train loss 0.52764; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.32ms; iter 77800: train loss 0.56470; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.04ms; iter 77900: train loss 0.34698; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.93ms; iter 78000: train loss 0.48043; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.01ms; iter 78100: train loss 0.31623; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.42ms; iter 78200: train loss 0.59416; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.92ms; iter 78300: train loss 0.43678; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 78400: train loss 0.46715; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 78500: train loss 0.51720; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.66ms; iter 78600: train loss 0.56911; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 78700: train loss 0.46715; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.54ms; iter 78800: train loss 0.50754; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.09ms; iter 78900: train loss 0.59692; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 79000: train loss 0.50382; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.73ms; iter 79100: train loss 0.41227; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.54ms; iter 79200: train loss 0.50897; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.04ms; iter 79300: train loss 0.35166; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.56ms; iter 79400: train loss 0.44105; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 79500: train loss 0.47203; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.33ms; iter 79600: train loss 0.59542; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 79700: train loss 0.76791; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.60ms; iter 79800: train loss 0.46543; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.19ms; iter 79900: train loss 0.32035; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.35ms; iter 80000: train loss 0.57288; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 80100: train loss 0.42274; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.90ms; iter 80200: train loss 0.54686; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 80300: train loss 0.43211; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.11ms; iter 80400: train loss 0.62104; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.68ms; iter 80500: train loss 0.52070; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.57ms; iter 80600: train loss 0.47416; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.86ms; iter 80700: train loss 0.42917; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 80800: train loss 0.40497; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.35ms; iter 80900: train loss 0.65675; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 81000: train loss 0.41228; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 81100: train loss 0.61769; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 81200: train loss 0.55321; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.60ms; iter 81300: train loss 0.39717; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 81400: train loss 0.57197; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 81500: train loss 0.45548; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 81600: train loss 0.68465; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.86ms; iter 81700: train loss 0.66011; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.54ms; iter 81800: train loss 0.55569; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.57ms; iter 81900: train loss 0.49504; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.92ms; iter 82000: train loss 0.43306; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.76ms; iter 82100: train loss 0.43802; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.76ms; iter 82200: train loss 0.42795; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.83ms; iter 82300: train loss 0.49116; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.71ms; iter 82400: train loss 0.65064; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.71ms; iter 82500: train loss 0.60685; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 82600: train loss 0.42263; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.57ms; iter 82700: train loss 0.36096; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.12ms; iter 82800: train loss 0.38444; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.74ms; iter 82900: train loss 0.63693; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.35ms; iter 83000: train loss 0.42336; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.18ms; iter 83100: train loss 0.46072; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.90ms; iter 83200: train loss 0.49703; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.95ms; iter 83300: train loss 0.42911; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.44ms; iter 83400: train loss 0.38233; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 83500: train loss 0.36242; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 83600: train loss 0.42400; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.64ms; iter 83700: train loss 0.43989; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.97ms; iter 83800: train loss 0.44002; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.92ms; iter 83900: train loss 0.34843; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.93ms; iter 84000: train loss 0.44552; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.52ms; iter 84100: train loss 0.56598; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.36ms; iter 84200: train loss 0.49026; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.35ms; iter 84300: train loss 0.49404; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.88ms; iter 84400: train loss 0.58676; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.44ms; iter 84500: train loss 0.43138; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.44ms; iter 84600: train loss 0.46209; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 84700: train loss 0.40178; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.44ms; iter 84800: train loss 0.60415; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 84900: train loss 0.49535; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.10ms; iter 85000: train loss 0.51751; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 85100: train loss 0.54969; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 85200: train loss 0.43784; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 85300: train loss 0.40197; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.22ms; iter 85400: train loss 0.63568; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.54ms; iter 85500: train loss 0.60615; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.56ms; iter 85600: train loss 0.52542; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.37ms; iter 85700: train loss 0.53152; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 85800: train loss 0.57706; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.75ms; iter 85900: train loss 0.40858; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 86000: train loss 0.49338; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.71ms; iter 86100: train loss 0.54322; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.87ms; iter 86200: train loss 0.63948; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.84ms; iter 86300: train loss 0.58739; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.79ms; iter 86400: train loss 0.66609; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 86500: train loss 0.45218; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.04ms; iter 86600: train loss 0.48417; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 86700: train loss 0.57376; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 86800: train loss 0.63062; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.64ms; iter 86900: train loss 0.43291; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.93ms; iter 87000: train loss 0.45584; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.12ms; iter 87100: train loss 0.44938; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.84ms; iter 87200: train loss 0.52027; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.42ms; iter 87300: train loss 0.61757; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.27ms; iter 87400: train loss 0.37396; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.62ms; iter 87500: train loss 0.50048; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.55ms; iter 87600: train loss 0.40016; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.27ms; iter 87700: train loss 0.59443; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.86ms; iter 87800: train loss 0.40972; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 87900: train loss 0.62009; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.17ms; iter 88000: train loss 0.37281; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.67ms; iter 88100: train loss 0.41763; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.71ms; iter 88200: train loss 0.42187; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.84ms; iter 88300: train loss 0.41405; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.57ms; iter 88400: train loss 0.45039; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.64ms; iter 88500: train loss 0.47757; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.38ms; iter 88600: train loss 0.38589; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 147.96ms; iter 88700: train loss 0.36060; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 88800: train loss 0.47744; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.15ms; iter 88900: train loss 0.47254; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.07ms; iter 89000: train loss 0.44773; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.78ms; iter 89100: train loss 0.68159; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.48ms; iter 89200: train loss 0.46157; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.33ms; iter 89300: train loss 0.60319; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.93ms; iter 89400: train loss 0.55714; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.24ms; iter 89500: train loss 0.44464; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 89600: train loss 0.55199; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.99ms; iter 89700: train loss 0.40101; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.14ms; iter 89800: train loss 0.38005; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 155.98ms; iter 89900: train loss 0.40263; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.78ms; iter 90000: train loss 0.64324; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.24ms; iter 90100: train loss 0.43415; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.59ms; iter 90200: train loss 0.58633; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.23ms; iter 90300: train loss 0.54817; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.65ms; iter 90400: train loss 0.65709; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.04ms; iter 90500: train loss 0.34003; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.32ms; iter 90600: train loss 0.50063; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.05ms; iter 90700: train loss 0.40547; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.88ms; iter 90800: train loss 0.49603; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 90900: train loss 0.39680; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.50ms; iter 91000: train loss 0.69112; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.06ms; iter 91100: train loss 0.55817; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.16ms; iter 91200: train loss 0.67155; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.64ms; iter 91300: train loss 0.64031; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 150.66ms; iter 91400: train loss 0.56061; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.77ms; iter 91500: train loss 0.64446; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.31ms; iter 91600: train loss 0.55554; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.55ms; iter 91700: train loss 0.72179; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.41ms; iter 91800: train loss 0.38836; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 91900: train loss 0.59607; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.42ms; iter 92000: train loss 0.52547; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.62ms; iter 92100: train loss 0.41869; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.52ms; iter 92200: train loss 0.35627; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.04ms; iter 92300: train loss 0.61984; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.25ms; iter 92400: train loss 0.36968; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.64ms; iter 92500: train loss 0.57799; Scalar Head Weights Changed:

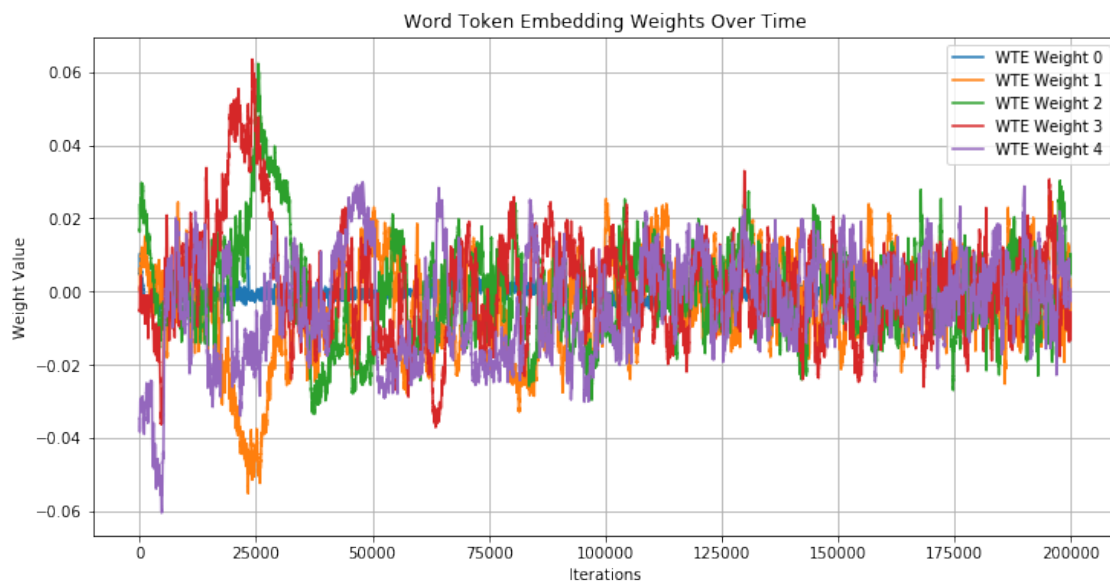
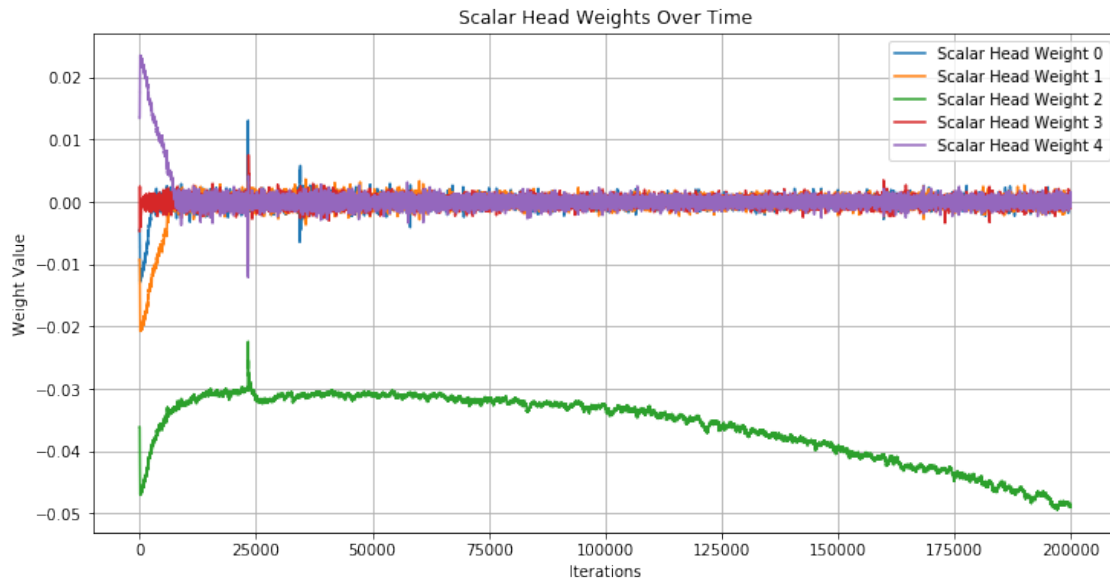
True; WTE Weights Changed: True
iter_dt 148.49ms; iter 92600: train loss 0.54874; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.80ms; iter 92700: train loss 0.69165; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.78ms; iter 92800: train loss 0.42107; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 92900: train loss 0.62525; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.33ms; iter 93000: train loss 0.32712; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.67ms; iter 93100: train loss 0.53642; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 93200: train loss 0.47450; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.18ms; iter 93300: train loss 0.46898; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.05ms; iter 93400: train loss 0.45168; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.72ms; iter 93500: train loss 0.59239; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.98ms; iter 93600: train loss 0.47619; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.83ms; iter 93700: train loss 0.43648; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.68ms; iter 93800: train loss 0.35145; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.40ms; iter 93900: train loss 0.40681; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 94000: train loss 0.47619; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.74ms; iter 94100: train loss 0.44362; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.71ms; iter 94200: train loss 0.53425; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.73ms; iter 94300: train loss 0.44538; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.69ms; iter 94400: train loss 0.51202; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.65ms; iter 94500: train loss 0.47702; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.61ms; iter 94600: train loss 0.53359; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.29ms; iter 94700: train loss 0.48481; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.02ms; iter 94800: train loss 0.48738; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.87ms; iter 94900: train loss 0.39439; Scalar Head Weights Changed:

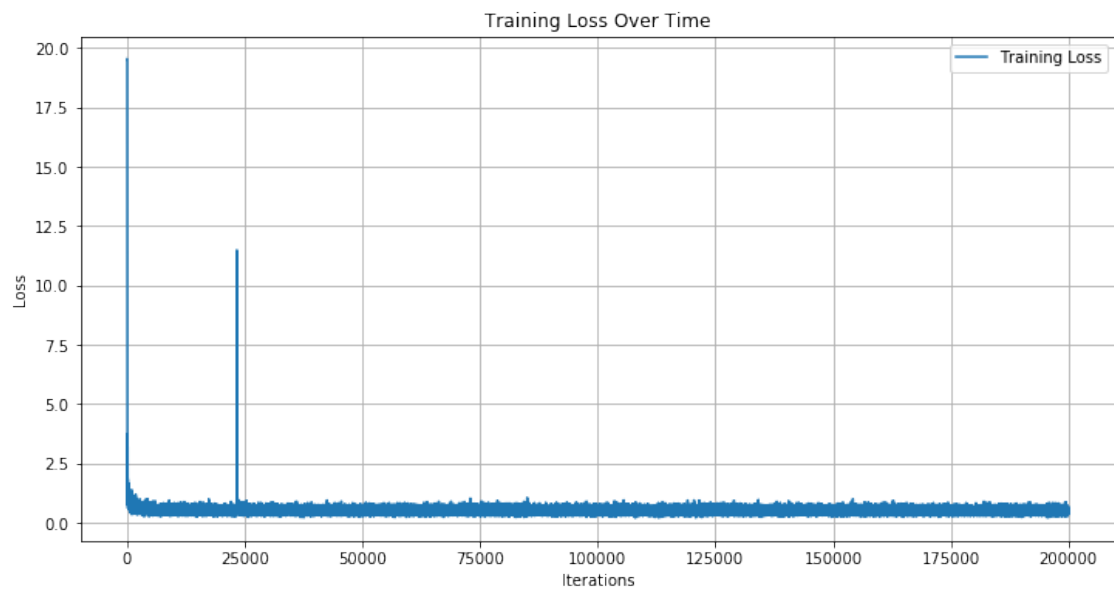
True; WTE Weights Changed: True
iter_dt 148.73ms; iter 95000: train loss 0.61579; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.84ms; iter 95100: train loss 0.40437; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.64ms; iter 95200: train loss 0.56244; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.48ms; iter 95300: train loss 0.54238; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.59ms; iter 95400: train loss 0.39859; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.41ms; iter 95500: train loss 0.55147; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 95600: train loss 0.37121; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 95700: train loss 0.34519; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.45ms; iter 95800: train loss 0.37281; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.55ms; iter 95900: train loss 0.39655; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.98ms; iter 96000: train loss 0.46637; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.01ms; iter 96100: train loss 0.43971; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.58ms; iter 96200: train loss 0.54240; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.82ms; iter 96300: train loss 0.41753; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.64ms; iter 96400: train loss 0.71320; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.69ms; iter 96500: train loss 0.46912; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 96600: train loss 0.41934; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 96700: train loss 0.52082; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.87ms; iter 96800: train loss 0.52871; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.26ms; iter 96900: train loss 0.39203; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.85ms; iter 97000: train loss 0.44085; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 97100: train loss 0.50463; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.39ms; iter 97200: train loss 0.44293; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.31ms; iter 97300: train loss 0.40255; Scalar Head Weights Changed:

True; WTE Weights Changed: True
iter_dt 148.79ms; iter 97400: train loss 0.55271; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 152.18ms; iter 97500: train loss 0.48486; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.61ms; iter 97600: train loss 0.33561; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.19ms; iter 97700: train loss 0.53912; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.54ms; iter 97800: train loss 0.35606; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.31ms; iter 97900: train loss 0.38773; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.46ms; iter 98000: train loss 0.46793; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.07ms; iter 98100: train loss 0.52294; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.63ms; iter 98200: train loss 0.37782; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.36ms; iter 98300: train loss 0.54627; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.55ms; iter 98400: train loss 0.48002; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.38ms; iter 98500: train loss 0.41080; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.54ms; iter 98600: train loss 0.55352; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.01ms; iter 98700: train loss 0.56464; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.20ms; iter 98800: train loss 0.47956; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.43ms; iter 98900: train loss 0.39955; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.47ms; iter 99000: train loss 0.63692; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.94ms; iter 99100: train loss 0.48004; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.01ms; iter 99200: train loss 0.35181; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.90ms; iter 99300: train loss 0.54919; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.73ms; iter 99400: train loss 0.43294; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 149.09ms; iter 99500: train loss 0.59665; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.34ms; iter 99600: train loss 0.45821; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.19ms; iter 99700: train loss 0.43233; Scalar Head Weights Changed:

```
True; WTE Weights Changed: True
iter_dt 148.38ms; iter 99800: train loss 0.47065; Scalar Head Weights Changed:
True; WTE Weights Changed: True
iter_dt 148.92ms; iter 99900: train loss 0.50988; Scalar Head Weights Changed:
True; WTE Weights Changed: True
Model and configurations saved.
```

```
/usr/lib/python3/dist-packages/IPython/core/pylabtools.py:132: UserWarning:
Creating legend with loc="best" can be slow with large amounts of data.
fig.canvas.print_figure(bytes_io, **kw)
```





[]: