CME 213, ME 339-Spring 2021

Eric Darve, ICME, Stanford



"Physics is the universe's operating system." (Steven R Garman)

Final project

Goal

Implementing a neural network in order to recognize hand-written digits

Logistics

Turn in	Date
Preliminary report + code	Friday May 28th
Final report (4 pages) + code	Sunday June 6th

Grade 20% 80%

Preliminary report

Focus is on correctness

Final report

Profiling and analysis, performance, quality of report

What are the performance bottlenecks in your code?

How can they be addressed?

Correctness

Discuss your strategy to test your code

Test outputs for valid inputs

Make sure you distinguish roundoff errors from genuine bugs

000000000000000000000000000000000000000	\00\000000000000000000000000000000000
000000000000000000000000000000000000000	000000000000000000000000000000000000000
<i>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~</i>	300=0000000000000000000000000000000000
000000000000000000000000000000000000000	\0000000000000000000000000000000000000
00000000000000000000000000000000000000	
000000000000000000000000000000000000000)
000000000000000000000000000000000000000	000000000000000000000000000000000000000
00000000000000000000000000000000000000	\$ C S C C C C C C C C C C C C C C C C C
0 000000000000000000000000000000000000	>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
000000000000000000000000000000000000000	000000000000000000000000000000000000000
001000000000000000000000000000000000000	
000000000000000000000000000000000000000)00 <i>-0</i> 000000000000000000000000000000000
000000000000000000000000000000000000000	000000000000000000000000000000000000000
000000000000000000000000000000000000000	
~ ~ ~ ~ ~ ~ ~ ~ ~ ~	}000000-000000000000000000000000000000
000000000000000000000000000000000000000	<i> 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</i>
	000000000000000000000000000000000000000
<u> </u>	0 <i>0</i> 0 0 0 − 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
00000000000000000000000000000000000000	>0000000000000000000000000000000000000
000000000000000000000000000000000000000	106000000000000000000000000000000000000
00080666008800000000000000000000000000	
000000000000000000000000000000000000000	; <i>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~</i>
000000000000000000000000000000000000000	000000000000000000000000000000000000000
00 < 0 < 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	70030000000000000000000000000000000000
000000000000000000000000000000000000000	300 <i>0</i> 0 <i>0000000000000000000000000000000</i>
000000000000000000000000000000000000000	
96609966666666666666666666666666666666	
000000000000000000000000000000000000000)@{``C`C`O`C``C``C`C`C`C`C`C`C`C`C`C`C`C`C
000000000000000000000000000000000000000	00000000000000000000000000000000000000
000000000000000000000000000000000000000	
00000000000000000000000000000000000000)
000000000000000000000000000000000000000	<i></i>
	000000000000000000000000000000000000000
000000000000000000000000000000000000000	00000000000000000000000000000000000000
000000000000000000000000000000000000000	//////////////////////////////////////
	000000000000000000000000000000000000000
00000000000000000000000000000000000000	
C0000000000000000000000000000000000000	>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
000000000000000000000000000000000000000	000000000000000000000000000000000000000
000000000000000000000000000000000000000	\$0000000000000000000000000000000000000
000000000000000000000000000000000000000	x~ <i>\0~10~0</i> 00000000000000000000000000000000
000000000000000000000000000000000000000	000000000000000000000000000000000000000
000000000000000000000000000000000000000	
001101000000000000000000000000000000000	> 0 0 ~ ~ C ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~
000000000000000000000000000000000000000	
00 6 0000000000000000000000000000000000	
000000000000000000000000000000000000000	<i>1000000000000000000000000000000000000</i>
000000000000000000000000000000000000000	
	>00-0 -0 -000000000000000000000000000000
000000000000000000000000000000000000000	000/00000000000000000000000000000000000
~~~0 <b>0</b> 00000000000000000000000000000000	10C0U00C0000000000000000000000000000000
000000000000000000000000000000000000000	000000000000000000000000000000000000000
000000000000000000000000000000000000000	000000000000000000000000000000000000000
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	0, 10, <i>0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0</i>
000000000000000000000000000000000000000	<b>200</b> 0000000000000000000000000000000000
000000000000000000000000000000000000000	000000000000000000000000000000000000000
000000000000000000000000000000000000000	
000000000000000000000000000000000000000	<b>↓ ○ ④ ④ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ </b>
000000000000000000000000000000000000000	1408608000000000000000000000000000000000
<b>e</b> coolooga a coola a coo	<u>  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~</u>
000000000000000000000000000000000000000	000000000000000000000000000000000000000
000000000000000000000000000000000000000	20000000000000000000000000000000000000
000000000000000000000000000000000000000	000000000000000000000000000000000000000
0 <b>000000000000000000000000000000000000</b>	<u> </u>
000000000000000000000000000000000000000	000000000000000000000000000000000000000
<b>BODDOODDOODDOODOODOODOODOO</b>	
00000000000000000000000000000000000000	<u> </u>
000000000000000000000000000000000000000	00000000000000000000000000000000000000
100000000000000000000000000000000000000	000000000000000000000000000000000000000
C000CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>



99999999999999999999999999999999999999	<b>A 9 5 9 8 1</b> 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
29429999999999999999999999999999999999	99994494999999999999999999999999999999
1 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	9 9 9 9 9 G G G G G G G G F 9 9 9 9 9 9
7 7 / 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	
<i>y u y t y y y y y y y y y y</i>	7434744494749474994744994744994744994999944999994
9999999999 <b>9</b> 9 <b>9</b> 9 <b>9</b> 99999999999999999	<i>\$ 7 9 9 9 4 4 4 9 4 9 4 9 4 9 4 9 4 9 9 9 9 9 9 9 9 4 4 4 9 9 4 7 4 4 4 9 9 4 7 4 4 4 9 9 9 4 9 4</i>
3 <i>9999999999</i> 99999999999999999999999999	} <b>79</b> 999999999 <b>9999999999999999999999999</b>
99999999999999999999999999999999999999	)
99 <b>99999999999999999999999999999999999</b>	7 9 <b>9 4 9 4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 </b>
8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	999044944949999999949949994999499449944
900999009090990990990990999999999999999	9 4 4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
a 6 6 6 9 9 0 9 9 0 9 9 9 9 8 8 8 8 8 9 9 9 8 9 9 9 9	0 8 9 8 8 0 6 9 9 8 9 9 0 6 8 9 0 9 9 0 9 9 9 9 8 9 8 9 8 9 8 9 8 9
	4 { 4 } 7 7 4 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
7 7 7 7 9 9 7 7 7 7 7 7 7 7 7 7 7 7 7 7	
9999944 <b>49</b> 999 <b>9</b> 749944999949994494 <b>9</b> 944 <b>9</b> 4 <b>49</b> 44	C
7 U 4 9 U 9 U 9 U 9 U 9 I 9 I 9 I 9 I 9 I 9 I	1 4 <b>9</b> 4 9 4 9 4 4 4 9 <b>4</b> 4 9 4 4 4 4 9 <b>4</b> 9 <b>4</b> 9 9 9 <b>9 9 9 9 9 9</b> 4 <b>9</b> 4 9 4 9 4 9 9 9 9 9 9 7
99 <b>999999999999999999</b> 9999999999999999	1999919191999 <b>999999999999999999999999</b>
99999999999999999999999999999999999999	<b>                                     </b>
99 <b>9</b> 999999999999999999999999999999999	1799 <b>91</b> 40995 <b>9</b> 99 <b>09</b> 997999 <b>959</b> 99 <b>959</b> 90 <b>9</b> 999999999999
79999999999999999999999999999999999999	199949999999999999999999999999999999999
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	999999999999999999999999999999999999999
999090999999999999999999999999999999999	
9 4 0 9 M 9 9 4 9 4 9 4 0 9 9 1 9 9 4 9 5 9 9 9 4 0 6 9 9 9 9 9 9 9	29/01/29 00 9 00 00 00 00 00 00 00 00 00 00 00
999999999999999999999999999999999999999	
7 9 4 7 4 4 9 <b>9</b> 4 9 9 4 9 9 9 <b>9 9</b> 4 9 4 9 4 9 4 9 4 9 4 9 4 9 4 9 4 9 4	9 4 4 9 5 5 5 7 4 4 9 9 4 9 4 9 9 9 9 9 9 9 9 4 4 9 9 9 4 9 9 9 4 4 9 9 9 4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
4 4 9 4 7 9 9 9 4 9 9 9 9 4 4 9 9 9 9 4 4 9 9 7 9 9 4 4 9 9 9 9	<b>44</b> 99974 <b>2</b> 49 <b>4</b> 4944779979 <b>4</b> 99 <b>4</b> 999994999444494499
YY Y Y Y Y Y Y Y 9 9 9 9 9 9 9 9 9 9 9	1 4 4 9 9 9 9 9 4 4 9 9 4 9 9 4 9 9 4 9 3 4 9 4 9
999 <b>9999999999999999999999999999999999</b>	9 4 4 9 4 4 9 9 4 8 9 4 9 9 4 4 9 9 9 4 9 9 9 9
99999999999999999999999999999999999999	119915911199 <b>99999999999</b> 999999999999999
9 <b>99999999999</b> 99999999 <b>999999999</b>	1 <b>991</b> 9 <b>999999999999999999</b> 999999999999
99999999999999999999999999999999999999	<i>J 9 9 F <b>9</b> 7 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9</i>
99999999999999999999999999999999999999	)
99 × 999999999999999999999999999999999	99999999999999999999999999999999999999
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	99999999999999999999999999999999999999
G 4 9 4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	
4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	99999999999999999999999999999999999999
4 7 7 4 4 9 7 7 J 9 4 1 7 4 4 7 7 7 7 4 7 7 4 4 7 7 4 7 7 4 7 1 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 4 7 7 7 4 7 7 7 4 7 7 7 4 7 7 7 4 7 7 7 4 7 7 7 4 7 7 7 4 7 7 7 4 7 7 7 4 7 7 7 7 4 7 7 7 7 4 7 7 7 7 4 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	
	777477777777777777777777777777777777777
9 4 9 5 9 9 9 9 9 9 4 1 4 1 4 1 9 4 4 <b>9</b> 4 7 4 9 7 4 9 0 9 9 4 4 7 9 9 0 4	797957077779977799797947944949494949499999999
99949999999999999999999999999999999999	4 4 4 4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 4 4 4 3 9 4 9 9 9 9
9 <b>5</b> 7 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	79917999999999999999999999999999999999
<b>9919999999999999999999999999999999999</b>	793 <b>9</b> <i>9999999999999999999999999999999999</i>
99 <b>999</b> 9999 <b>99999999999999999999999999</b>	j <b>9999999999999999</b> 99999999999999999999
<b>999</b> 999999999999999999999999999999999	3 <b>9</b> 9 9 9 9 1 9 2 9 9 9 9 9 9 9 9 9 9 9 9 9
999919999999999999999999999999999999999	19999999999999999999999999999999999999
93799999999999999999999999999999999999	<b>7999999999999999</b> 999999999999999999999
99999999999999999999999999999999999999	799999399599599999999999999999999999999
999999999999999999999999999999999999999	99999999999999999999999999999999999999
99999999999999999999999999999999999999	q 7 q 9 q 9 q 9 q 9 x 7 7 7 4 9 7 9 8 4 7 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
4 9 4 4 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	0 9 9 7 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
9 4 4 9 9 A 9 9 9 9 9 9 4 9 9 9 9 9 9 9	
999999999999999999999999999999999999999	1 1 1 2 2 3 4 1 3 5 9 5 9 9 4 9 9 9 4 9 9 6 9 9 9 9 9 9 9 9 9 9
9 4 9 6 9 9 9 9 9 6 6 9 9 9 9 9 9 9 9 9	999909999999999999999999999999999999999
9996 4 9 9 4 2 9 0 9 0 9 9 9 9 9 9 9 0 0 0 0 0 0 0 9 7 0 4 9 1 9 9 0	a a a a a a a a a a a a a a a a a a a
09999999999999999999999909000000000000	399903900990490494949494149004904904904904
	17777771777777777777777777777777777777
	<b>17 5 5 5 7 7 7 7 7 7 7 7 7 7</b>
4 <b>5</b> <i>4 5 4</i> <b>7 7 7 7 7 7 7 7 7 7</b>	777777777777777777777777777777777777777
<b><i>y y y y y y y y y y</i> </b>	1 3 9 4 4 9 4 9 9 9 9 9 4 9 4 9 4 9 4 9 9 9 9 9 9 4 <b>4</b> 4 9 9 9 9 9 4 4 4 9 4 9 4 9 4 9 4 9 4
99199997994444959999994994949494949999	1 1 1 1 4 9 4 1 9 4 4 9 4 4 9 7 1 9 9 9 9 1 9 1 9 9 9 9 9 9 9 9 9 9
999 <b>9999999999999999999999999999999999</b>	<b>1999</b> 999999999999999999999999999999999
1 <b>9</b> 9 <b>999999999999999999999999999999999</b>	<b>19999999999999999999999999</b> 99999999999
99999999499999999999999999999999999999	<b>99999993999999999999</b> 999999999999 <b>99999999</b>
999999 <b>999999999999</b> 999999999999999999	<b>\999{{999}}999\$999\9999\$999\$999\$\$999\$\$999\$\$999\$\$999\$\$999\$</b>
<i>99999</i> 999 <b>99</b> 99999999999999999999999999	<b>19</b> 9999 <b>9999999999</b> 999999999999994 <b>9</b> 99 <b>9</b> 9 <b>9</b>
89989999999999999999999999999999999999	₹ <b>9</b> 99999 <b>994491</b> 999 <i>999</i> 99999999999999999999999999999
19999999999999999999999999999999999999	87199945999999999999999999999999999999999
<i>9 / <b>9</b> 9 9 9 <b>9 9 9 9 9 9 9 9 9 9 9 9 9 9 </b></i>	\
<i>9 9</i> <b>9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 </b>	\
<i>₣</i> <b>₮९२</b> <i>₮९<b>१</b><i>९११७९९१९१९१९१११११११११११११</i></i>	\
<i>\$</i> ? <b>\</b> ? <b>9 \</b> 4 <b>\</b> 9 <b>\</b> 7 <b>?</b> 9 <b>\</b> 7 <b>? \</b> 9 <b>\</b> 9 <b>\</b> 7 <b>\</b> 7 <b>\</b> 9 <b>\</b> 7 <b>\</b>	\
<i>\$</i> 7 <b>9 9 9 9 9 9 7 7 9 9 9 9 9 9 9 9 9 9</b>	\ 9 9 9 9 9 1 9 9 9 9 9 9 9 9 9 9 9 9 9
\$7\$\$\$\$9\$\$9\$\$9\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$	\4999499 9999954999999999999999999999999





Input layer: image

Hidden layer: – n num; variable size

Output layer: softmax vector with 10 digits

### Softmax

$$ext{softmax}(z)_j = rac{\exp(z_j)}{\sum_{i=0}^9 \exp(z_i)}$$

Interpreted as a probability

Each layer is a matrix multiplication and a non-linear function

$$z = Wx + b$$
 $a = \sigma(z)$ 



We will use sigmoid

### How do you train a network?

Many methods but most are based on gradient descent

**Error function** 

$$J(p) = rac{1}{N}\sum_{i=1}^N \mathrm{error}^{(i)}(y_i, \hat{y}_i)$$

p: weights and biases of network = all parameters

### Gradient update

 $p \leftarrow p - lpha 
abla_p J$ 

Gradient is computed by repeated application of the chain rule

Backpropagation



# Stochastic gradient descent

$$J_r(p) = rac{1}{N} \sum_{i \in ext{random subset}} ext{error}^{(i)}(y_i, \hat{y}_i)$$

$$p \leftarrow p - lpha 
abla_p J_r$$

If we use a small subset, this allows more updates to the DNN coefficients

 $\Rightarrow$  more accurate

Randomness of subset selection allows avoiding local minima and escaping saddle points

⇒ better convergence



### Sequence of operations

Forward pass = left to right; DNN prediction; compare with label

Backward propagation = right to left; chain rule; compute gradient and update DNN

Iterate until convergence





# Core building blocks to implement

- Matrix-matrix products
- Non-linear activation functions

https://playground.tensorflow.org











# Regularization

$$J(p) = rac{1}{N} \sum_{i=1}^{N} \mathrm{error}^{(i)}(y_i, \hat{y}_i) + rac{1}{2} \, \lambda \, ||p||_2^2$$

*p*: weights and biases of the network

$$egin{aligned} z = Wx + b \ a = \sigma(z) \end{aligned}$$

Regularization makes the DNN more linear

How can we figure out how much regularization is needed?

Training set: used to minimize loss; involved in defining the gradient Validation set: used to evaluate model; how accurate is it? Avoids overfitting



#### Example: small 2-layer DNN with width 8





#### With noise added to data





#### Fix 1: reduce the size of the DNN; for example with width 1



![](_page_35_Figure_2.jpeg)

Fix 2: add regularization, e.g.,  $\lambda = 10^{-3}$ 

![](_page_36_Figure_1.jpeg)

![](_page_36_Figure_2.jpeg)

![](_page_37_Picture_0.jpeg)

### Diagnostic

### Overfitting; † $\lambda$

# Too much regularization; $\downarrow \lambda$

### Regularization is good

- Training set: optimize DNN parameters
- Validation set: optimize regularization

Two main tasks in the project

Implement a matrix-matrix product (GEMM) algorithm
 Implement the MPI algorithm for distributed memory

Naive implementation; shared memory is not used

![](_page_40_Figure_1.jpeg)

# GEMM performance

The key is to increase the arithmetic intensity.

This requires reducing the memory traffic.

![](_page_42_Figure_0.jpeg)

$$c_{ij} = \sum_k a_{ik} b_{kj}$$
 $c_{ij} \leftarrow c_{ij} + a_{ik} b_{kj}$ 

# $c_{ij} \leftarrow c_{ij} + a_{ik} b_{kj}$

Block size: b

Memory traffic: 2b

Flops:  $b^2$ 

![](_page_44_Picture_4.jpeg)

High arithmetic intensity

![](_page_45_Figure_0.jpeg)

# MPI, distributed memory algorithm

Topic of upcoming lectures

High-level discussion of approaches

![](_page_47_Figure_0.jpeg)

![](_page_48_Figure_0.jpeg)

![](_page_48_Picture_1.jpeg)

Communication

$$J(p) = rac{1}{N}\sum_{i=1}^N \mathrm{error}^{(i)}(y_i, \hat{y}_i)$$

Sum is required over all input images to compute gradient

Parallel reduction to get  $abla_p J$ 

= Reduction for all DNN coefficients across all nodes

![](_page_51_Figure_0.jpeg)

Time for MPI communication is fairly significant.

A better implementation exists!

### Model parallelism

Much more complicated to understand but implementation is not more difficult than previous approach

![](_page_54_Figure_0.jpeg)

Reduction is required at the end to get the output labels  $m{y}$ 

### Backpropagation

You have not seen the details yet. So, it will be hard to follow.

The take-home message is that no MPI communication is required between nodes.

![](_page_57_Picture_0.jpeg)

 $[W^{(2)}]^T y \qquad W^{(1)} \leftarrow W^{(1)} - \alpha [(W^{(2)})^T y] x^T$ 

![](_page_57_Picture_3.jpeg)

#### Warning!

Equations in previous slide were simplified for clarity

See Part1 write-up for details

No communication is required during the backpropagation

This implementation is much more efficient

![](_page_60_Picture_0.jpeg)