ML4AAD - Final Project

Winter Semeseter 18/19

Neeratyoy Mallik

4774378

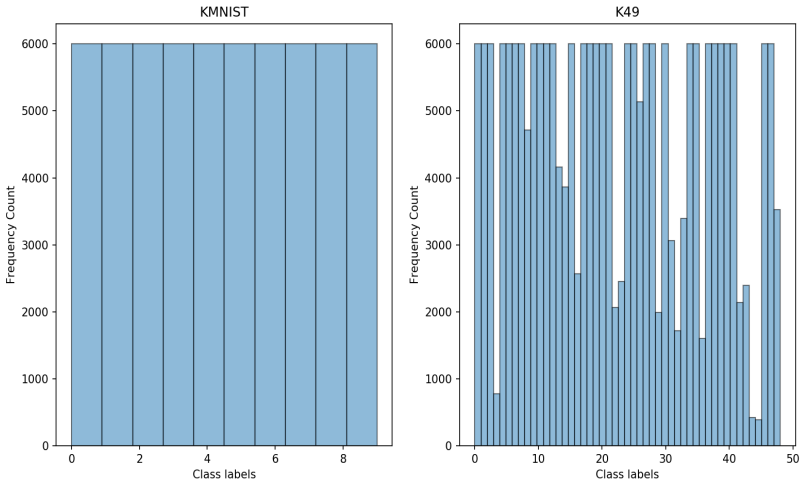
February 24, 2019

Motivation (The Why)

Motivation (The Why)

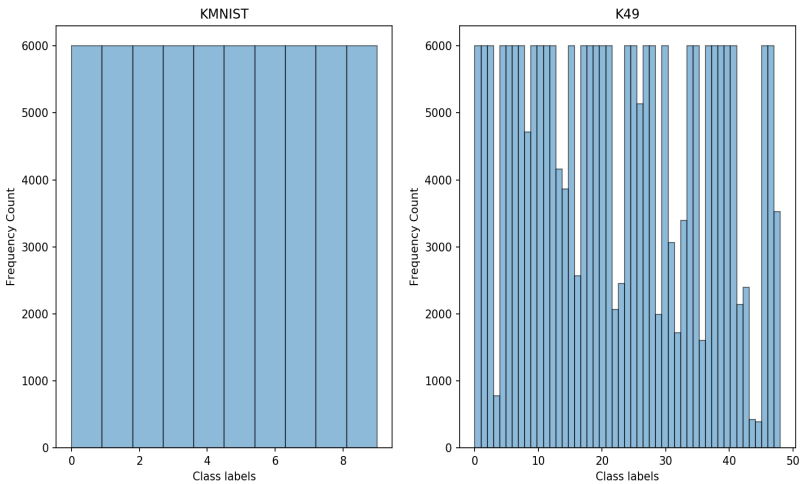
Motivation (The Why)

2 (similar) datasets with differing size, # of classes, class balance

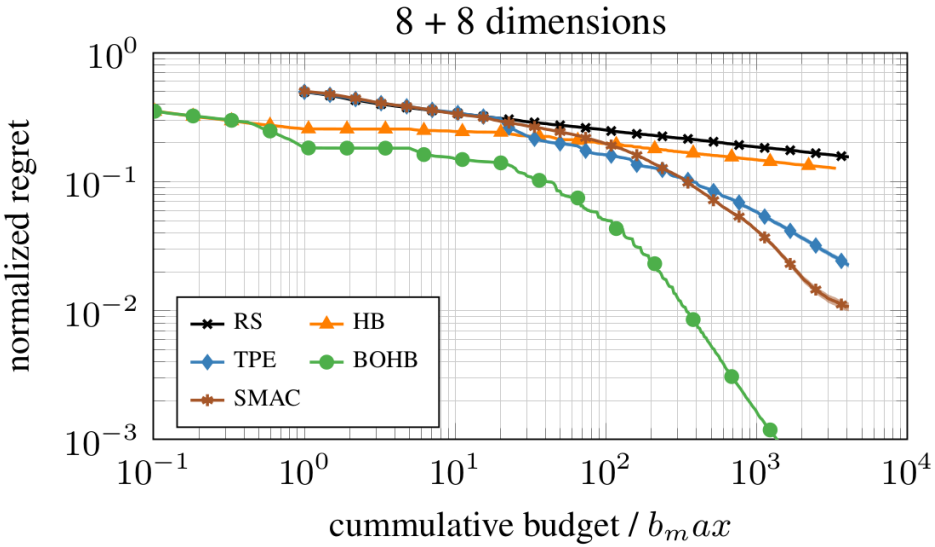


Motivation (The Why)

2 (similar) datasets with differing size, # of classes, class balance



BOHB outperforms SMAC in high dimensional, mixed data



Design Decisions (and BOHB)

CNN Structure

$INPUT \rightarrow [CONV \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT? \rightarrow MAXPOOL?]*M \rightarrow [FC \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT?]*K \rightarrow OUTPUT$

$M \in \{1, 2, 3\}$; $K \in \{0, 1, 2\}$; $? \rightarrow \top or \bot$

Design Decisions (and BOHB)

CNN Structure

$INPUT \rightarrow [CONV \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT? \rightarrow MAXPOOL?]*M \rightarrow [FC \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT?]*K \rightarrow OUTPUT$
$M \in \{1, 2, 3\}$; $K \in \{0, 1, 2\}$; $? \rightarrow \top or \bot$

- CONVolution layers
  - Kernel size
  - Padding
  - Stride
- ACTIVATION (relu/sigmoid/tanh)
- BATCHNORM, DROPOUT
  - True or False
- MAXPOOL (if True)
  - Kernel size (=stride)

Design Decisions (and BOHB)

CNN Structure

$INPUT \rightarrow [CONV \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT? \rightarrow MAXPOOL?]*M \rightarrow [FC \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT?]*K \rightarrow OUTPUT$
$M \in \{1, 2, 3\}$; $K \in \{0, 1, 2\}$; $? \rightarrow \top or \bot$

- CONVolution layers
  - Kernel size
  - Padding
  - Stride
- ACTIVATION (relu/sigmoid/tanh)
- BATCHNORM, DROPOUT
  - True or False
- MAXPOOL (if True)
  - Kernel size (=stride)
- Learning rate
- Optimizer
- Batch size

Design Decisions (and BOHB)

CNN Structure

$INPUT \rightarrow [CONV \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT? \rightarrow MAXPOOL?]*M \rightarrow [FC \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT?]*K \rightarrow OUTPUT$
$M \in \{1, 2, 3\}$; $K \in \{0, 1, 2\}$; $? \rightarrow \top or \bot$

- CONVolution layers
  - Kernel size
  - Padding
  - Stride
- ACTIVATION (relu/sigmoid/tanh)
- BATCHNORM, DROPOUT
  - True or False
- MAXPOOL (if True)
  - Kernel size (=stride)
- Learning rate
- Optimizer
- Batch size

- >30 hyperparameters
- (Budget = Epochs) $\rightarrow$ expensive runs

Design Decisions (and BOHB)

CNN Structure

$INPUT \rightarrow [CONV \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT? \rightarrow MAXPOOL?]*M \rightarrow [FC \rightarrow BATCHNORM? \rightarrow ACTIVATION \rightarrow DROPOUT?]*K \rightarrow OUTPUT$

$M \in \{1, 2, 3\}$; $K \in \{0, 1, 2\}$; $? \rightarrow \top or \bot$

- CONVolution layers
  - Kernel size
  - Padding
  - Stride
- ACTIVATION (relu/sigmoid/tanh)
- BATCHNORM, DROPOUT
  - True or False
- MAXPOOL (if True)
  - Kernel size (=stride)
- Learning rate
- Optimizer
- Batch size

- >30 hyperparameters
- (Budget = Epochs) $\rightarrow$ expensive runs

BOHB params:

eta	min_budget	max_budget
2	1	16
4	1	16
3	1	9
2	1	10

BOHB on KMNIST (and K49)

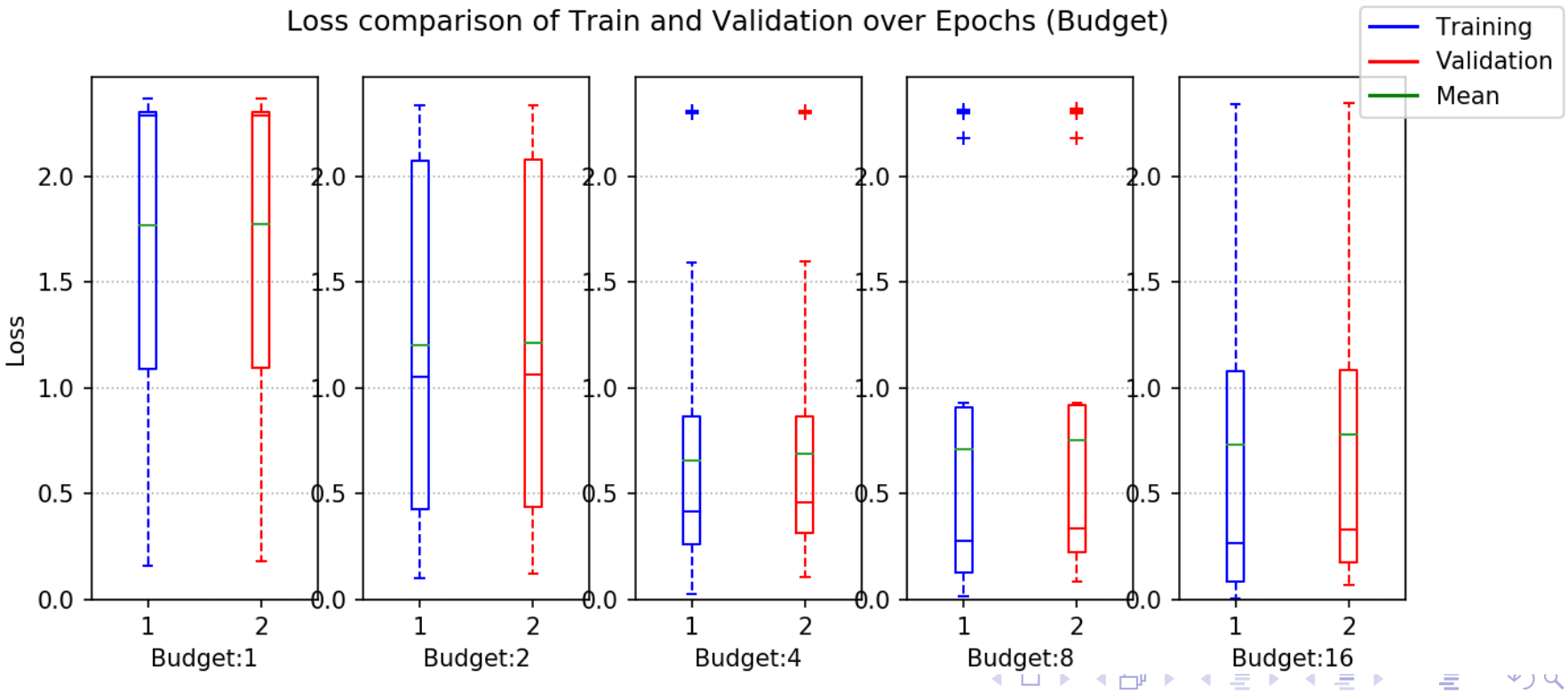
Dataset	eta_min_max_iter (BOHB)	Validation Accuracy	Train Accuracy	Test Accuracy	BOHB Runtime
KMNIST	2_1_16_10	98.08%	99.58%	94.69%	<3 hrs
KMNIST	3_1_9_20	97.17%	98.23%	93.26%	<3 hrs
KMNIST	4_1_16_20	98.26%	99.95%	95.98%	<4 hrs
KMNIST	2_1_10_14	97.77%	99.98%	94.09%	<4 hrs
KMNIST	3_2_20_20	98.31%	99.57%	95.51%	<6 hrs

BOHB on KMNIST (and K49)

Dataset	eta_min_max_iter (BOHB)	Validation Accuracy	Train Accuracy	Test Accuracy	BOHB Runtime
KMNIST	2_1_16_10	98.08%	99.58%	94.69%	<3 hrs
KMNIST	3_1_9_20	97.17%	98.23%	93.26%	<3 hrs
KMNIST	4_1_16_20	98.26%	99.95%	95.98%	<4 hrs
KMNIST	2_1_10_14	97.77%	99.98%	94.09%	<4 hrs
KMNIST	3_2_20_20	98.31%	99.57%	95.51%	<6 hrs
K49	3_1_9_10	89.06%	99.93%	88.12%	<5 hrs
K49	2_1_10_10	91.19%	99.25%	88.25%	<12 hrs

BOHB on KMNIST (and K49)

Dataset	eta_min_max_iter (BOHB)	Validation Accuracy	Train Accuracy	Test Accuracy	BOHB Runtime
KMNIST	2_1_16_10	98.08%	99.58%	94.69%	<3 hrs
KMNIST	3_1_9_20	97.17%	98.23%	93.26%	<3 hrs
KMNIST	4_1_16_20	98.26%	99.95%	95.98%	<4 hrs
KMNIST	2_1_10_14	97.77%	99.98%	94.09%	<4 hrs
KMNIST	3_2_20_20	98.31%	99.57%	95.51%	<6 hrs
K49	3_1_9_10	89.06%	99.93%	88.12%	<5 hrs
K49	2_1_10_10	91.19%	99.25%	88.25%	<12 hrs



Extracting juice (Can K49 leverage KMNIST?)

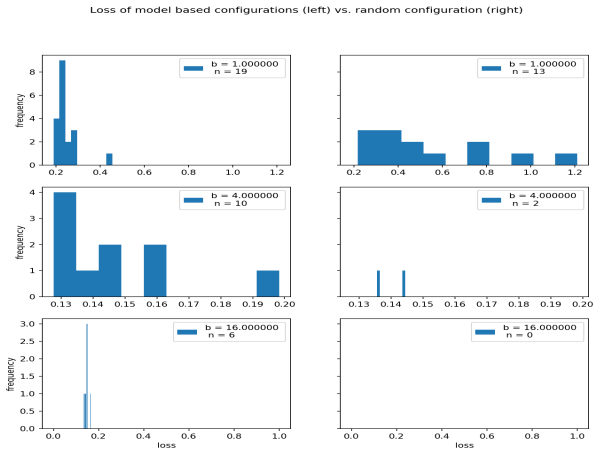
Model: KMNIST $\rightarrow$ K49

KMNIST Test	K49 Train	K49 Test	Run-time
94.69%	96.45%	89.91%	<1hr
95.51%	95.62%	90.20%	<1hr
95.98%	99.63%	93.07%	<1hr

Extracting juice (Can K49 leverage KMNIST?)

Model: KMNIST $\rightarrow$ K49

KMNIST Test	K49 Train	K49 Test	Run-time
94.69%	96.45%	89.91%	<1hr
95.51%	95.62%	90.20%	<1hr
95.98%	99.63%	93.07%	<1hr



Configuration: KMNIST $\rightarrow$ K49
Hyperparameters for BOHB:

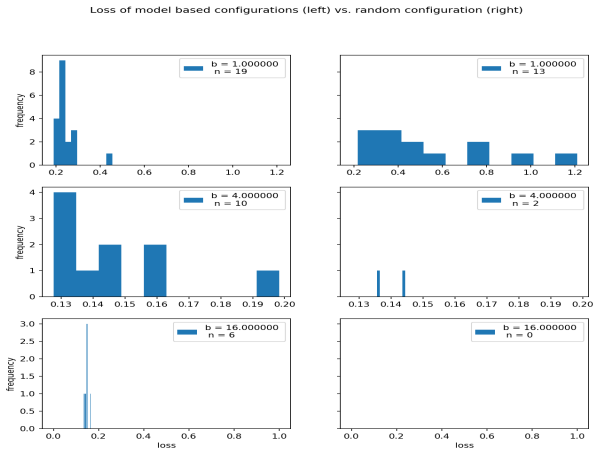
- batch size
- # of channels
- # of FC layers and neurons

KMNIST Test	K49 Train	K49 Test	Run-time
95.51%	99.86%	93.85%	<6hrs
95.98%	99.93%	94.19%	<20hrs

Extracting juice (Can K49 leverage KMNIST?)

Model: KMNIST $\rightarrow$ K49

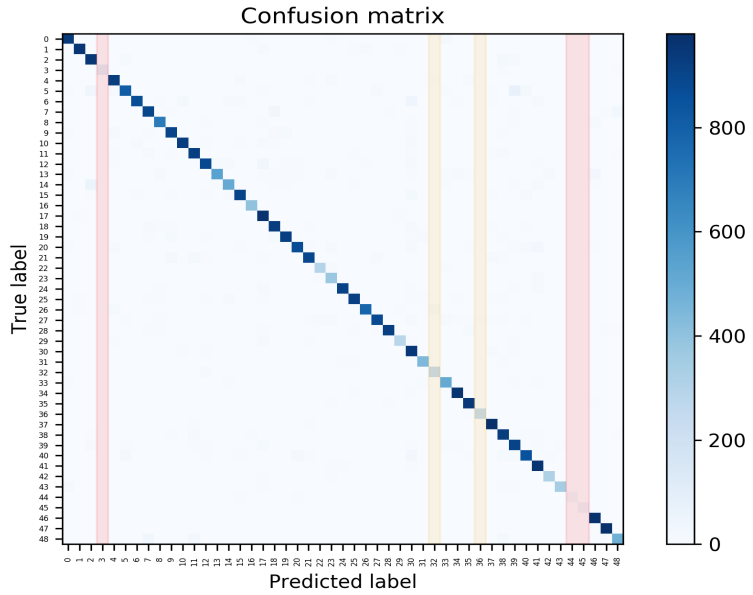
KMNIST Test	K49 Train	K49 Test	Run-time
94.69%	96.45%	89.91%	<1hr
95.51%	95.62%	90.20%	<1hr
95.98%	99.63%	93.07%	<1hr



Configuration: KMNIST $\rightarrow$ K49
Hyperparameters for BOHB:

- batch size
- # of channels
- # of FC layers and neurons

KMNIST Test	K49 Train	K49 Test	Run-time
95.51%	99.86%	93.85%	<6hrs
95.98%	99.93%	94.19%	<20hrs

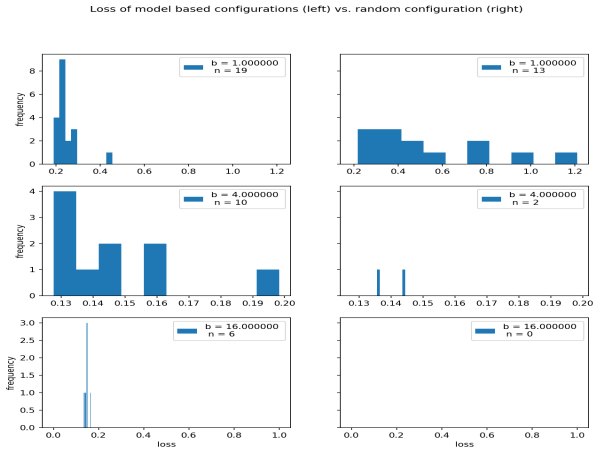


Issues: (i) Under-represented classes; (ii) Slow experiments;

Extracting juice (Can K49 leverage KMNIST?)

Model: KMNIST $\rightarrow$ K49

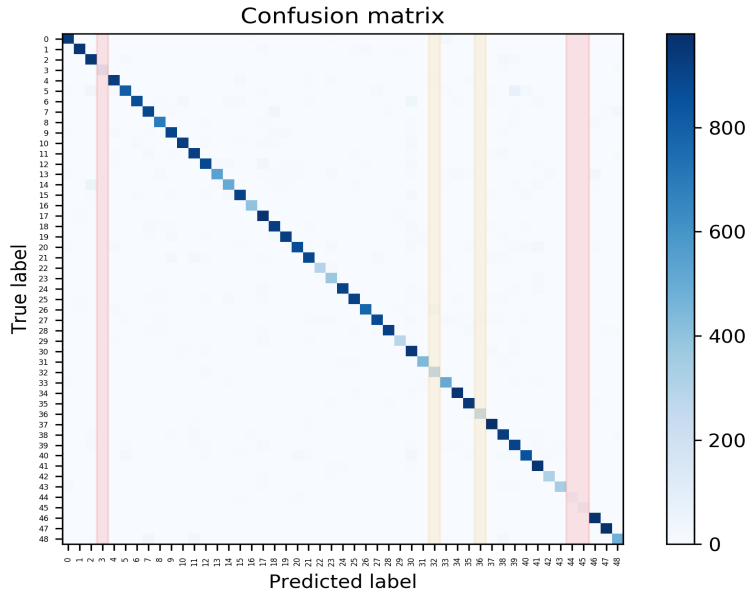
KMNIST Test	K49 Train	K49 Test	Run-time
94.69%	96.45%	89.91%	<1hr
95.51%	95.62%	90.20%	<1hr
95.98%	99.63%	93.07%	<1hr



Configuration: KMNIST $\rightarrow$ K49
Hyperparameters for BOHB:

- batch size
- # of channels
- # of FC layers and neurons

KMNIST Test	K49 Train	K49 Test	Run-time
95.51%	99.86%	93.85%	<6hrs
95.98%	99.93%	94.19%	<20hrs



Issues: (i) Under-represented classes; (ii) Slow experiments; [(iii) What if only K49?]

Final Results

Final numbers to report.