

Building Machine Learning Systems on Microsoft Azure Cloud Virtual Machines

Cheng-Hao Cai, Jing Sun, Gillian Dobbie

School of Computer Science
University of Auckland
chenghao.cai@auckland.ac.nz

This project is supported by the Microsoft Asia Cloud Research Software Fellowships (CRSF) Program

Contents

- ▶ Introduction
- ▶ Microsoft Azure Cloud Virtual Machines
- ▶ Project 1: Automatic Voice Recognition
- ▶ Project 2: Automated Program Repair
- ▶ Conclusion
- ▶ References

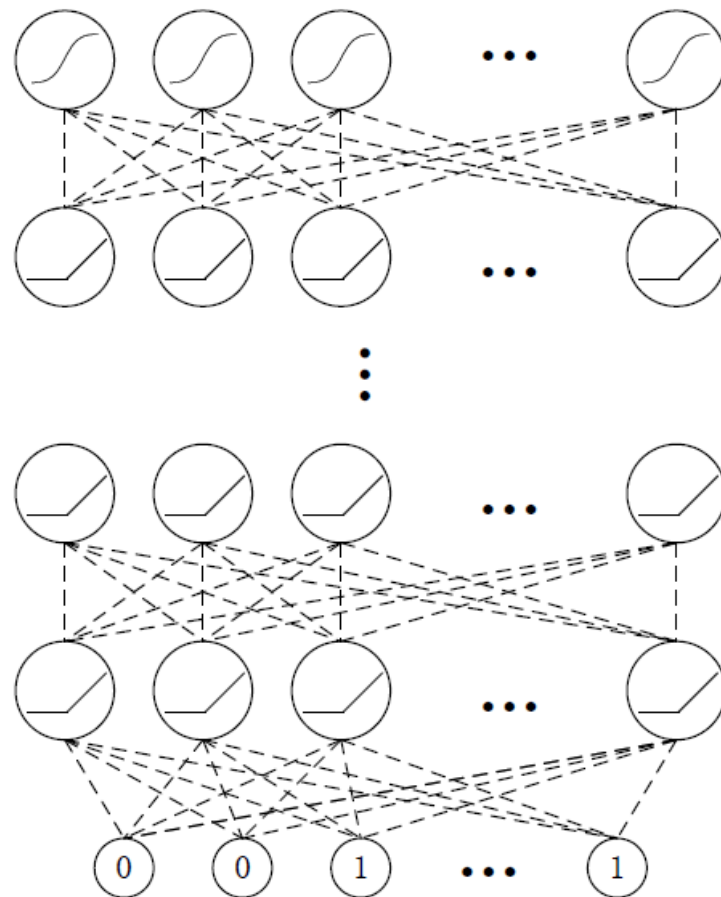
Introduction

Motivation

- ▶ **Machine Learning (ML)** techniques are widely used in industry.
- ▶ Popular ML techniques include:
 - Neural Networks
 - Decision Trees
 - Support Vector Machines
 - Hidden Markov Models
 - etc.
- ▶ ML applications: voice recognition, face recognition, machine translation, weather prediction, etc.

Motivation

- ▶ Most ML techniques require very powerful computers to run.
- ▶ For example, to build a voice recognition system, you may need a neural network with thousands of neurons and millions of neural connections.
- ▶ Using a computer with hundreds of CPUs, you still need several days to train the neural network.



Research Questions

- ▶ How to build large ML systems?
- ▶ How to use large ML models when hardware resources are limited?
- ▶ Applications of ML models?

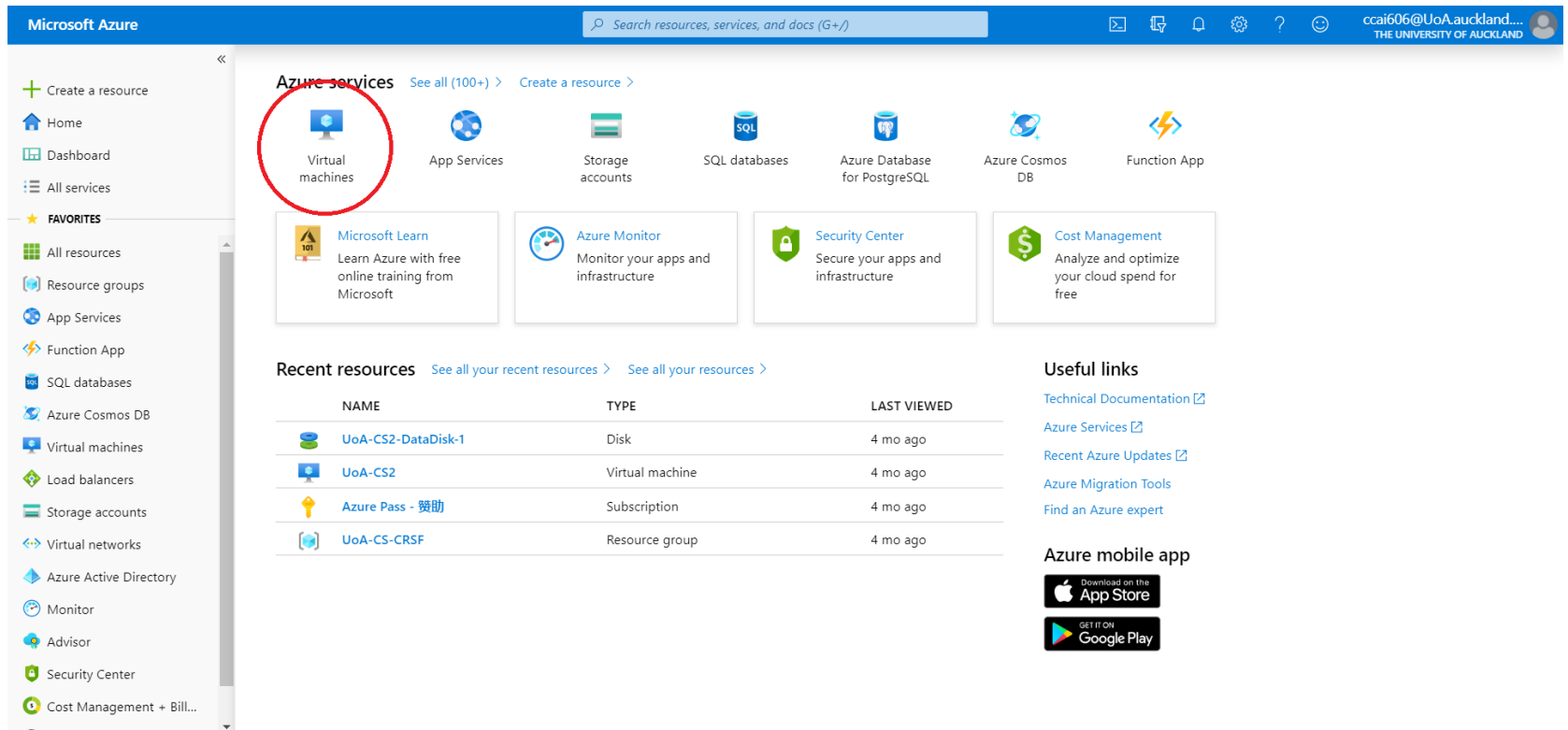
Our Solutions

- ▶ We run large ML models on Microsoft Azure Virtual Machines (VM).
- ▶ We build ML-based voice recognition systems.
- ▶ We apply ML techniques to automatic program repair.

Microsoft Azure Cloud VMs

Microsoft Azure Cloud VMs

- In Microsoft Azure Cloud, VMs can be set up:



The screenshot shows the Microsoft Azure portal interface. The 'Virtual machines' service is highlighted with a red circle in the 'Azure services' section. The left sidebar contains a navigation menu with options like 'Create a resource', 'Home', 'Dashboard', 'All services', and 'FAVORITES'. The main content area displays various Azure services and a table of recent resources.

Azure services See all (100+) > Create a resource >

- Virtual machines
- App Services
- Storage accounts
- SQL databases
- Azure Database for PostgreSQL
- Azure Cosmos DB
- Function App

Microsoft Learn
Learn Azure with free online training from Microsoft

Azure Monitor
Monitor your apps and infrastructure

Security Center
Secure your apps and infrastructure

Cost Management
Analyze and optimize your cloud spend for free

Recent resources See all your recent resources > See all your resources >

NAME	TYPE	LAST VIEWED
UoA-CS2-DataDisk-1	Disk	4 mo ago
UoA-CS2	Virtual machine	4 mo ago
Azure Pass - 赞助	Subscription	4 mo ago
UoA-CS-CRSF	Resource group	4 mo ago

Useful links

- Technical Documentation
- Azure Services
- Recent Azure Updates
- Azure Migration Tools
- Find an Azure expert

Azure mobile app

Download on the App Store

GET IT ON Google Play

Microsoft Azure Cloud VMs

- Azure VMs include different operating systems:

The screenshot shows the 'Create a virtual machine' wizard in the Microsoft Azure portal. The 'Image' dropdown menu is open, displaying a list of operating systems. The 'Image' field is highlighted with a red box.

Instance details

- Virtual machine name: UoA-CS-3
- Region: (US) East US 2
- Availability options: No infrastructure redundancy required
- Image: Ubuntu Server 16.04 LTS
- Size: (not specified)

Administrator account

- Authentication type: (not specified)
- Username: (not specified)
- SSH public key: (not specified)

[Learn more about creating and using SSH keys in Azure](#)

[Review + create](#) [< Previous](#) [Next : Disks >](#)

Microsoft Azure Cloud VMs

- Configurations of CPUs and memory can be changed:

Microsoft Azure

Search resources, services, and docs (G+)

Home > Virtual machines

Create a virtual machine

Instance details

Virtual machine name *

Region *

Availability options

Image *

Size *

Administrator account

Authentication type

Username *

SSH public key *

Review + create

Select a VM size

Browse available virtual machine sizes and their features

Search by VM size... Clear all filters

Size: Medium (7-16) Generation: 2 selected Family: General purpose Premium disk: Supported Add filter

Showing 9 of 251 VM sizes. | Subscription: Azure Pass - 赞助 | Region: West US 2 | Current size: Standard_D2s_v3

VM Si...	Offering	Family	vCPUs	RAM (GiB)	Data disks	Max IOPS	Temporary storage (...)	Premium disk ...	Cost/month	esti
B8ms	Standard	General purpose	8	32	16	4320	64	Yes	NZ\$374.59	
D8s_v3	Standard	General purpose	8	32	16	12800	64	Yes	NZ\$431.96	
DS4_v2	Standard	General purpose	8	28	32	25600	56	Yes	NZ\$515.20	
B12ms	Standard	General purpose	12	48	16	6480	96	Yes	NZ\$561.32	
B16ms	Standard	General purpose	16	64	32	8640	128	Yes	NZ\$749.17	
D16asv3	Standard	General purpose	16	64	32	25600	128	Yes	Unavailable	
D16s_v3	Standard	General purpose	16	64	32	25600	128	Yes	NZ\$863.91	
D8asv3	Standard	General purpose	8	32	16	12800	64	Yes	Unavailable	
DS5_v2	Standard	General purpose	16	56	64	51200	112	Yes	NZ\$1,030.40	

Select

Prices presented are estimates in your local currency that include only Azure infrastructure costs and any discounts for the subscription and location. The prices don't include any applicable software costs. [View Azure pricing calculator](#). Final charges will appear in your local currency in cost analysis and billing views.

Microsoft Azure Cloud VMs

- ▶ Extra disk spaces can be added into VMs:

The screenshot shows the Microsoft Azure portal interface. On the left is a navigation sidebar with options like 'Create a resource', 'Home', 'Dashboard', 'All services', and 'FAVORITES'. The main area is divided into two panels. The left panel, titled 'Create a new disk', contains fields for 'Name' (UoA-CS-3_DataDisk_0), 'Source type' (None (empty disk)), and 'Size' (1023 GiB, Premium SSD). The right panel, titled 'Select a disk size', shows a table of available disk sizes and their features. A red box highlights this table. Below the table is a section for 'Create a custom size' with explanatory text.

Create a new disk

Create a new disk to store applications and data on your VM. Disk pricing varies by storage type, and number of transactions. [Learn more about Azure Managed Disks](#)

Name *

Source type * ⓘ

Size * ⓘ **1023 GiB**
Premium SSD
[Change size](#)

Select a disk size

Browse available disk sizes and their features.

Account type ⓘ
Premium SSD

Size	Disk tier	Max IOPS	Max throughput
32 GiB	P4	120	25
64 GiB	P6	240	50
128 GiB	P10	500	100
256 GiB	P15	1100	125
512 GiB	P20	2300	150
1024 GiB	P30	5000	200
2048 GiB	P40	7500	250
4096 GiB	P50	7500	250
8192 GiB	P60	16000	500
16384 GiB	P70	18000	750
32767 GiB	P80	20000	900

Create a custom size

Enter the size of the disk you would like to create. You will be charged the same rate for your provisioned disk, regardless of how much of the disk space is being used. For example, a 200 GiB disk is provisioned on a 256 GiB disk, so you would be billed for the 256 GiB.

Microsoft Azure Cloud VMs

- For each Azure VM, an IP of remote access is automatically created:

The screenshot displays the Microsoft Azure portal interface. On the left, the 'All resources' link in the 'FAVORITES' section is highlighted with a red box. The main content area shows the details for a virtual machine named 'UoA-CS2'. The 'Overview' tab is selected, displaying various properties. The 'Public IP address' is highlighted with a red box and shows the value '52.167.60.10'. Below the properties, there are two line charts: 'CPU (average)' and 'Network (total)', both showing data for the last 1 hour.

Microsoft Azure

Search resources, services, and docs (G+)

ccaif606@UoAuckland...
THE UNIVERSITY OF AUCKLAND

Home > All resources > UoA-CS2

UoA-CS2
Virtual machine

Search (Ctrl+/)

Connect Start Restart Stop Capture Delete Refresh

Resource group (change) : UoA-CS-CRSF

Status : Running

Location : East US 2

Subscription (change) : Azure Pass - 赞助

Subscription ID : dcdaf7d3-4b01-4d0f-9c51-4c5179addca5

Computer name : UoA-CS2

Operating system : Linux (ubuntu 16.04)

Size : Standard D8s v3 (8 vcpus, 32 GiB memory)

Ephemeral OS disk : N/A

Public IP address : 52.167.60.10

Private IP address : 10.0.0.4

Virtual network/subnet : UoA-CS-CRSF-vnet/default

DNS name : Configure

Tags (change) : Click here to add tags

Show data for last: 1 hour 6 hours 12 hours 1 day 7 days 30 days

CPU (average)

Network (total)

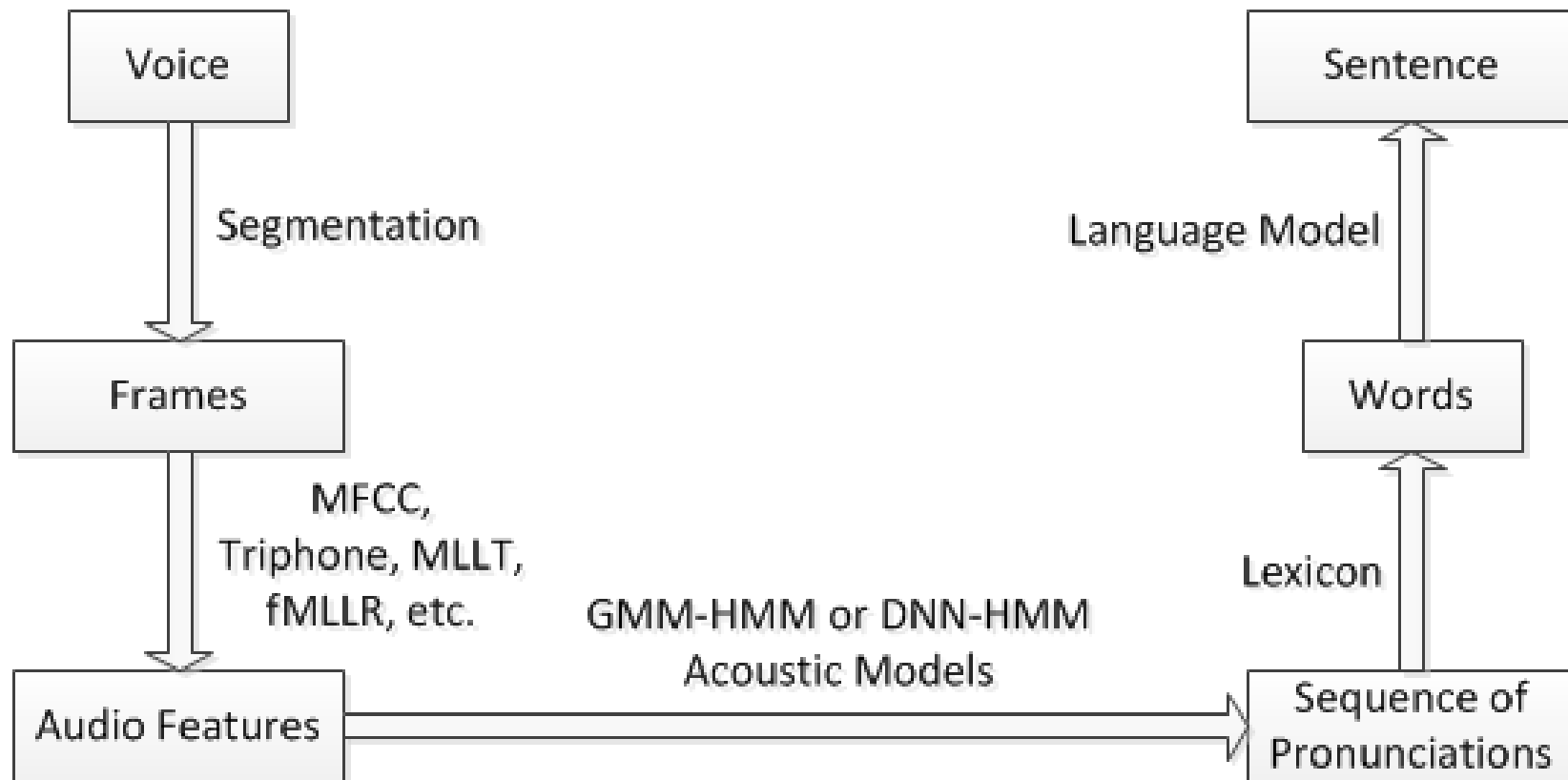
Project 1

Automatic Voice Recognition

Automatic Speech Recognition

- ▶ **Automatic Speech Recognition (ASR)** is the use of machines to convert human speeches to texts.
- ▶ ASR has been used in many applications:
 - Siri
 - Amazon Alexa
 - automatic subtitle generator
 - etc.
- ▶ Problems with ASR: accents, background noise, professional words, etc.

Voice Recognition System with the Kaldi toolkit



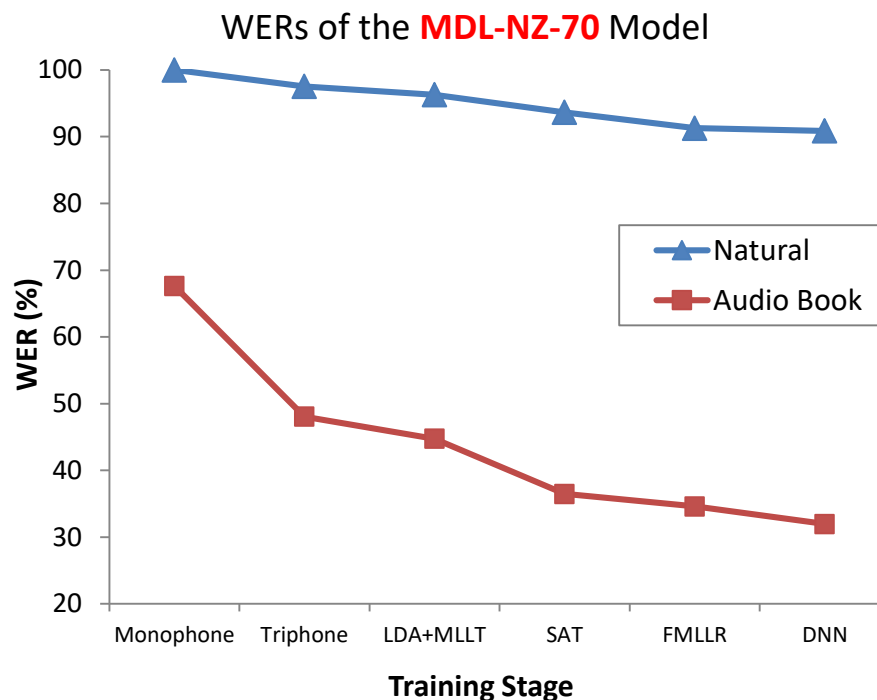
Example Results - Noisy Test Data

Kaldi's Output: "DENOUNCE **STANDING AND** FEATURE HAS BEEN TANTALIZED IT'S ALWAYS BEEN IMPORTANT HOW PEOPLE IT IS A TROUBLE I CALMLY AND SEVERAL KINDS OF OCCUPIED THE **PEACH** IT IS IMPORTANT TO ALL OF THEM EVERY TROUBLE GROUP WITHIN THIS ROCK TO BELIEVE THAT IT IS THAT **TAKEN** PLACE COULD DO SO BUT THERE WHILE HE CUT TOO HAS BEEN USED TO KILOMETER WALKING TRACK WHICH ZIGZAGS TO THE PEAK AS POLARIZED **THE COMMUNITY AND THOSE WHO WANT TO GET GOOD AND THOSE DEFIANTLY AND PACED A A LAUGHS** A GREEDY RANGE IS ONE OF THE HORSE BAYS BEST KNOWN MINORIES LAST YEAR THE VINEYARD SPENT THREE HUNDRED THOUSAND DOLLARS CREATING THE TRACK ON NEIGHBOURING LAND OLD CONSENT FROM THE HASTINGS DISTRICT COUNSEL TO CUT THE FACE OF THE MOTHER BUT EVERY LETTER THAT HE WE PAULLINUS SAYS NO ONE HAD BOTHERED CONSULTING WITH LOCAL **MONEY**"

Standard Answer: "DEEMED AN **OUTSTANDING LAND** FEATURE HAS BEEN VANDALISED IT S ALWAYS BEEN IMPORTANT TO OUR PEOPLE IT IS A TRIBAL ICON AND SEVERAL TRIBES HAVE OCCUPIED THE **PEAK** AND IT IS IMPORTANT TO ALL OF THEM EVERY TRIBAL GROUP WITHIN THIS ROHE BELIEVES THAT IT IS A **TAPU** PLACE IT IS SACRED BUT THEIR WAHI TAPU HAS A NEW KM WALKING TRACK WHICH ZIGZAGS TO THE PEAK IT HAS POLARISED **THE COMMUNITY THOSE WHO WANT TO KEEP IT AND THOSE DEFIANTLY OPPOSED MAN TIKA TONU ALL U E TIKA TONU U E TIKA TONU ATU KI A KOE E TAMA** CRAGGY RANGE IS ONE OF THE HAWKE S BAY S BEST KNOWN WINERIES LAST YEAR THE VINEYARD SPENT CREATING THE TRACK ON NEIGHBOURING LAND IT SOUGHT CONSENT FROM THE HASTINGS DISTRICT COUNCIL TO CARVE THE FACE OF TE MATA BUT IWI LEADER NGAHIWI TOMOANA **SAYS NO** ONE BOTHERED CONSULTING WITH LOCAL **MAORI**"



Natural Voice vs Audio Book Voice



- ▶ Natural --- The test set of natural voice of movies, games, etc.
- ▶ Audio Book --- The test set of audio book voice from LibriSpeech.
- ▶ Finding: Natural voice is difficult to be recognised.

Project 2

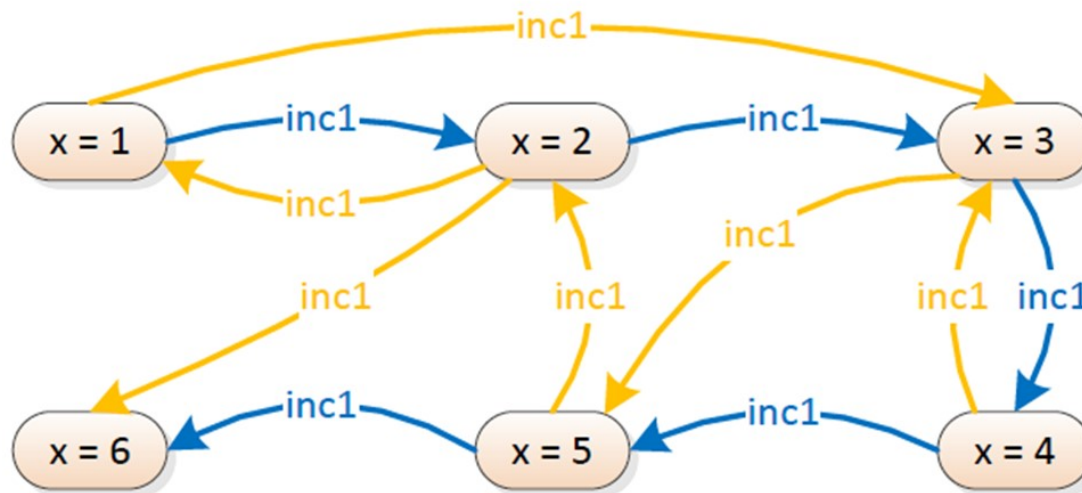
Automated Program Repair

Automated Program Repair

- ▶ Software engineers write computer programs every day.
- ▶ After writing programs, engineers are responsible for repairing faulty programs.
- ▶ The manual repair of faulty programs often reduces the efficiency of software development.
- ▶ Can we automate program repair processes?

Automated Program Repair

- ▶ The first step of automated program repair is to make computers understand how programs run.
- ▶ State transitions can represent program running processes.
- ▶ For example, if we have a program “inc1”: PRE $x > 0 \ \& \ x < 6$ THEN $x := x + 1$ END. **Valid executions** and **invalid executions** are shown in the following state diagram:



- ▶ ML models can learn to identify valid executions and invalid executions.

Automated Program Repair

► Results of state transition learning:

Subject	# Examples	ROC-AUC				
		<i>BNB</i>	<i>LR</i>	<i>SVM</i>	<i>RF</i>	<i>Silas</i>
M01	24530	0.615	0.619	1.000	1.000	1.000
M02	22500	0.722	0.736	0.984	0.996	0.990
M03	4608	0.729	0.691	0.965	1.000	1.000
M04	10202	0.660	0.668	0.936	1.000	1.000
M05	15360	0.640	0.645	0.997	0.999	0.999
M06	14160	0.584	0.593	0.999	0.999	0.999
M07	4550	0.663	0.654	0.725	0.965	0.995
M08	2024	0.673	0.709	0.905	0.998	0.988
M09	14580	0.536	0.539	0.994	0.999	0.999
M10	6592	0.625	0.636	0.993	1.000	0.998
M11	15860	0.848	0.868	0.930	0.997	0.995
M12	20976	0.566	0.563	0.996	1.000	1.000
M13	10210	0.536	0.532	0.621	0.999	1.000
M14	11918	0.802	0.814	0.997	1.000	0.997
M15	1944	0.777	0.837	0.844	0.999	0.988
M16	2456	0.642	0.666	0.850	0.999	0.999
M17	9056	0.679	0.735	0.956	0.998	0.996
M18	51384	0.783	0.822	0.995	1.000	0.998
Average	242910	0.682	0.697	0.962	0.998	0.998
Classification Accuracy		0.636	0.652	0.902	0.981	0.983

- M01 - 18: programs
- ROC-AUC: a degree of ML model performance
- BNB - Bernoulli naive Bayes classifiers,
- LR - logistic regression classifiers
- SVM - support vector machines
- RF - random forests (normal)
- Silas - random forests (high performance)

Conclusion

- ▶ ML techniques require considerable computational resources to run.
- ▶ We used Microsoft Azure Cloud VMs to build ML systems.
- ▶ We built an automatic voice recognition system on an Azure VM.
- ▶ We built an ML-based automated program repair system on an Azure VM.
- ▶ In the future, we look forward to using better ML techniques to improve the two projects.

References

- [1] Jean-Raymond Abrial. The B-book – assigning programs to meanings. Cambridge University Press, 2005.
- [2] Cheng-Hao Cai, Jing Sun and Gillian Dobbie. Automatic B-model Repair using Model Checking and Machine Learning, in Automated Software Engineering, Volume 26, Issue 3, Pages 653-704, Springer, August 2019.
- [3] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, [...]. The Kaldi Speech Recognition Toolkit. IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.
- [4] Vassil Panayotov, Guoguo Chen, Daniel Povey, Sanjeev Khudanpur. Librispeech: An ASR corpus based on public domain audio books. ICASSP 2015: 5206-5210.