Machine learning for Quantum Noise Discrimination

Machine Learning at GGI

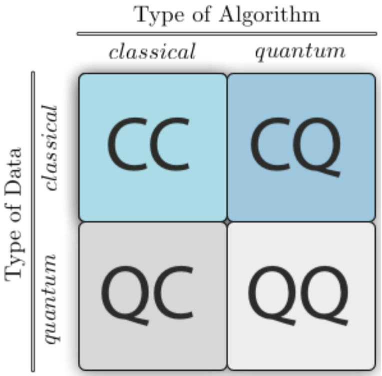
Stefano Martina
stefano.martina@unifi.it

Quantum Driving And
Bio-complexity

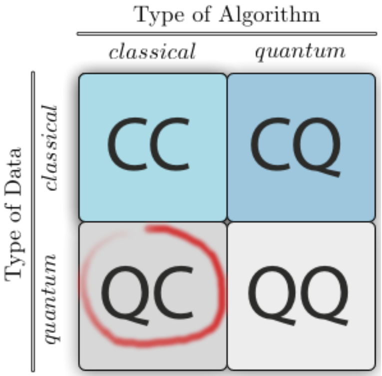
UNIVERSITÀ
DEGLI STUDI
FIRENZE

September 5, 2022

Quantum Machine Learning



Quantum Machine Learning



Introduction

Paper in review: Stefano Martina, Stefano Gherardini and Filippo Caruso. Machine learning classification of non-Markovian noise disturbing quantum dynamics

- ✔ We consider a Quantum Random Walker on a complex graph
  - perturbed by noise
- ✔ Discriminate quantum noise, by measuring only walker populations
  - Support Vector Machines
  - Neural Networks and Recurrent Neural Networks
- ✔ The dynamic parameters are crucial to the classification capacity
  - short evolution time / high frequency ⟶ easy
  - long evolution time / low frequency ⟶ hard
- ✔ Over 90% accuracy in classification between
  - two IID noises
  - two coloured noises
  - one IID VS one coloured noises

easier

Introduction

Paper in review: Stefano Martina, Stefano Gherardini and Filippo Caruso. Machine learning classification of non-Markovian noise disturbing quantum dynamics

- ✔ We consider a Quantum Random Walker on a complex graph
  - perturbed by noise
- ✔ Discriminate quantum noise, by measuring only walker populations
  - Support Vector Machines
  - Neural Networks and Recurrent Neural Networks
- ✔ The dynamic parameters are crucial to the classification capacity
  - short evolution time / high frequency ⟶ easy
  - long evolution time / low frequency ⟶ hard
- ✔ Over 90% accuracy in classification between
  - two IID noises
  - two coloured noises
  - one IID VS one coloured noises

easier

Introduction

Paper in review: Stefano Martina, Stefano Gherardini and Filippo Caruso. Machine learning classification of non-Markovian noise disturbing quantum dynamics

- ✔ We consider a Quantum Random Walker on a complex graph
  - perturbed by noise
- ✔ Discriminate quantum noise, by measuring only walker populations
  - Support Vector Machines
  - Neural Networks and Recurrent Neural Networks
- ✔ The dynamic parameters are crucial to the classification capacity
  - short evolution time / high frequency ⟶ easy
  - long evolution time / low frequency ⟶ hard
- ✔ Over 90% accuracy in classification between
  - two IID noises
  - two coloured noises
  - one IID VS one coloured noises

easier

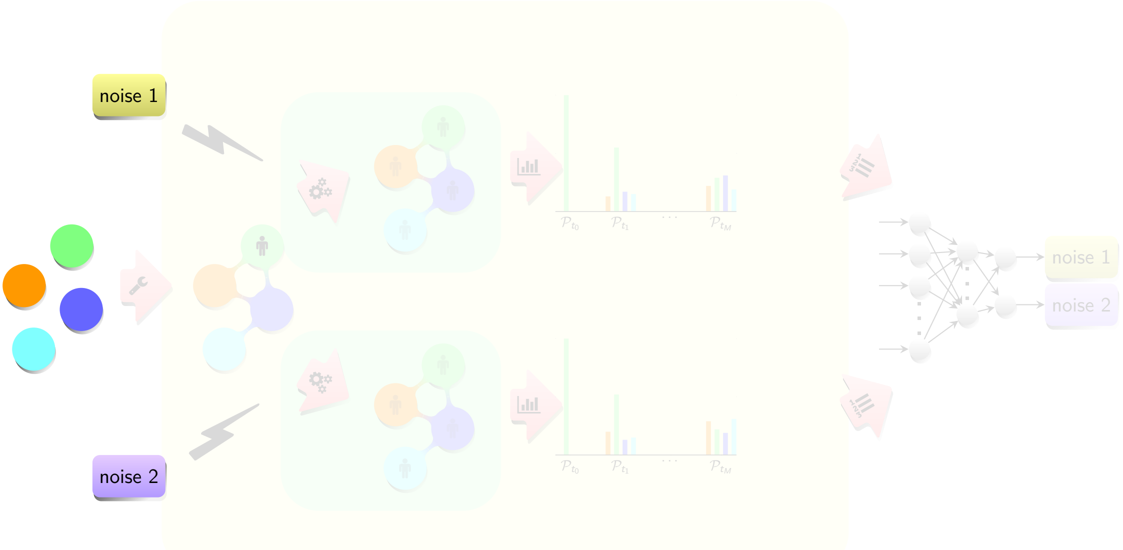
Introduction

Paper in review: Stefano Martina, Stefano Gherardini and Filippo Caruso. Machine learning classification of non-Markovian noise disturbing quantum dynamics

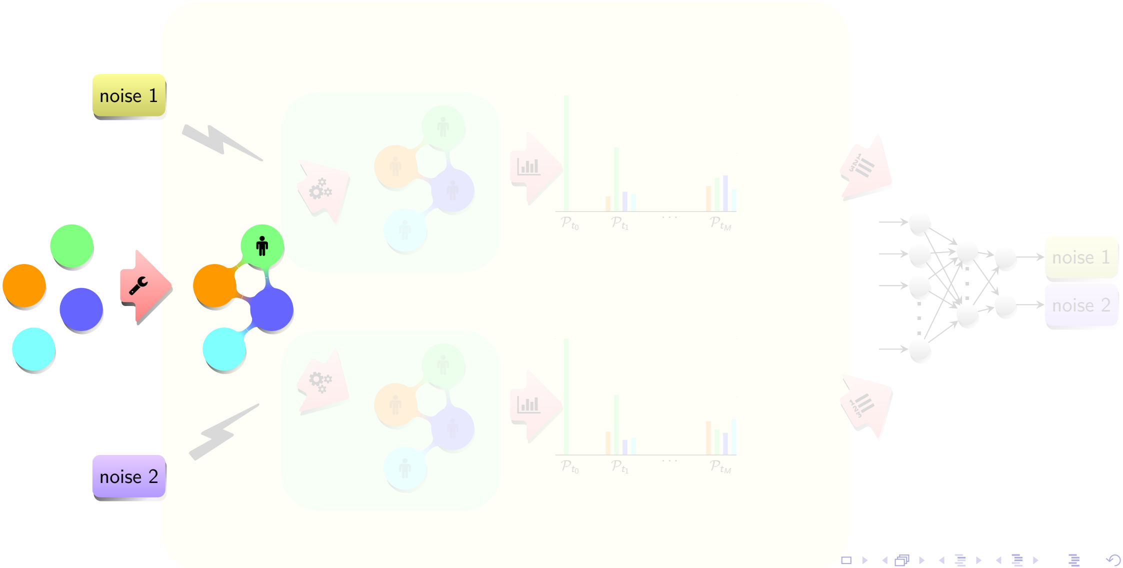
- ✔ We consider a Quantum Random Walker on a complex graph
  - perturbed by noise
- ✔ Discriminate quantum noise, by measuring only walker populations
  - Support Vector Machines
  - Neural Networks and Recurrent Neural Networks
- ✔ The dynamic parameters are crucial to the classification capacity
  - short evolution time / high frequency ⟶ easy
  - long evolution time / low frequency ⟶ hard
- ✔ Over 90% accuracy in classification between
  - two IID noises
  - two coloured noises
  - one IID VS one coloured noises

easier

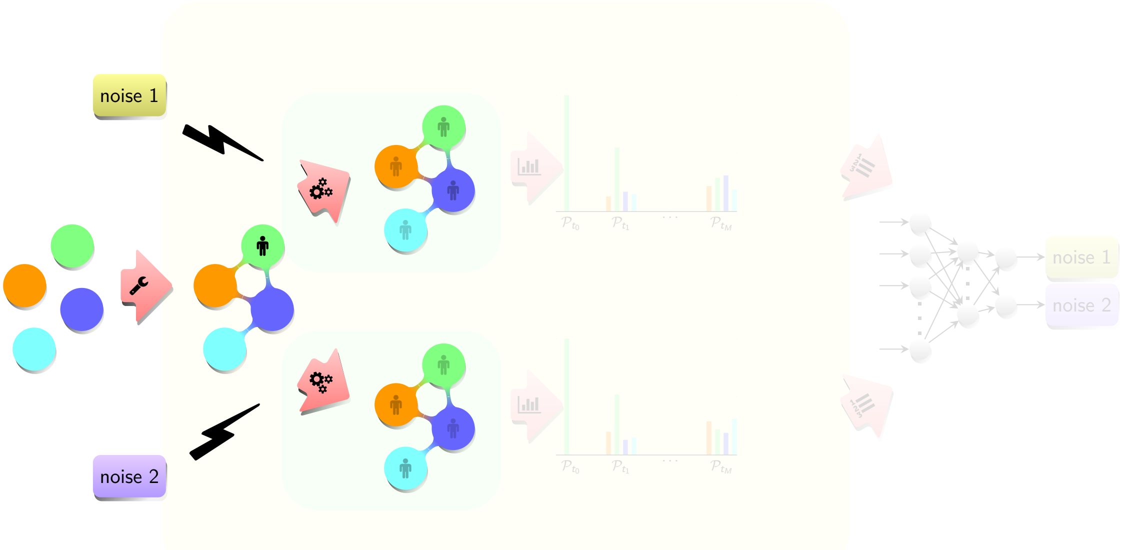
Setting definition



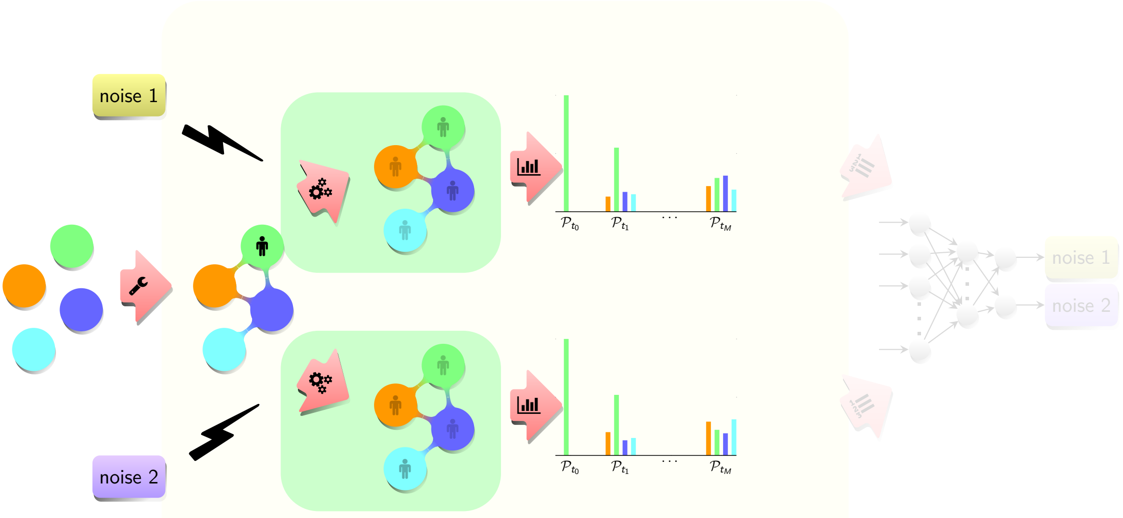
Setting definition



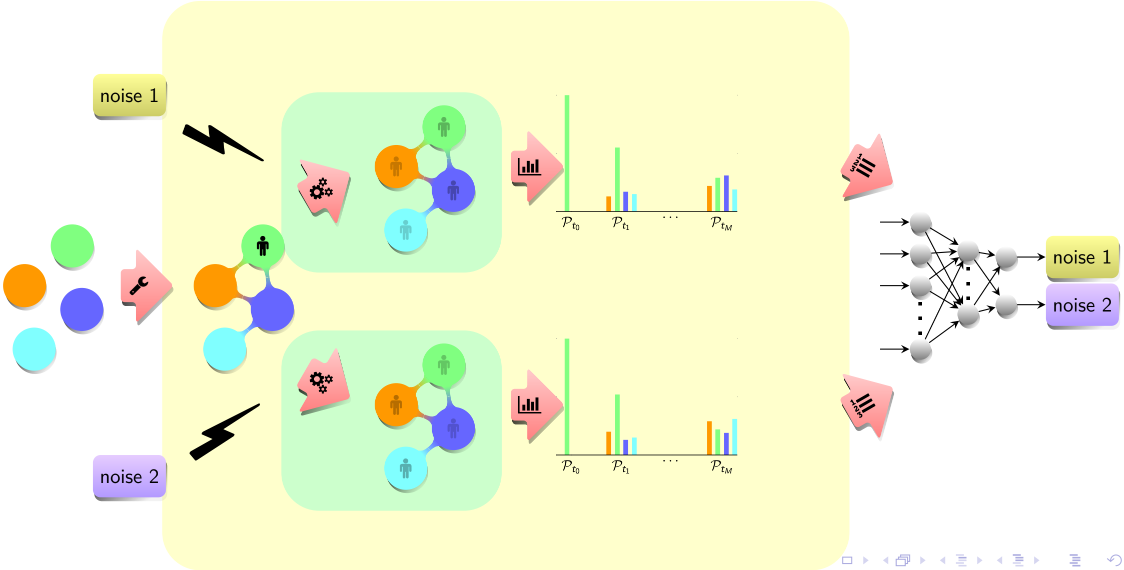
Setting definition



Setting definition

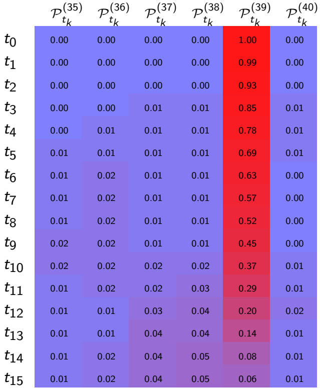


Setting definition



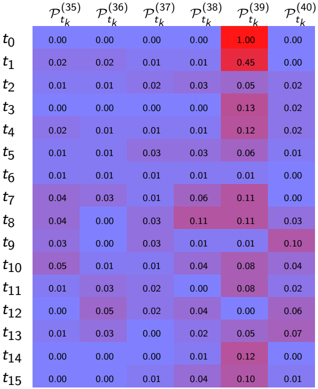
Example

- ✔ Example of populations with $t_{15} = 0.1$



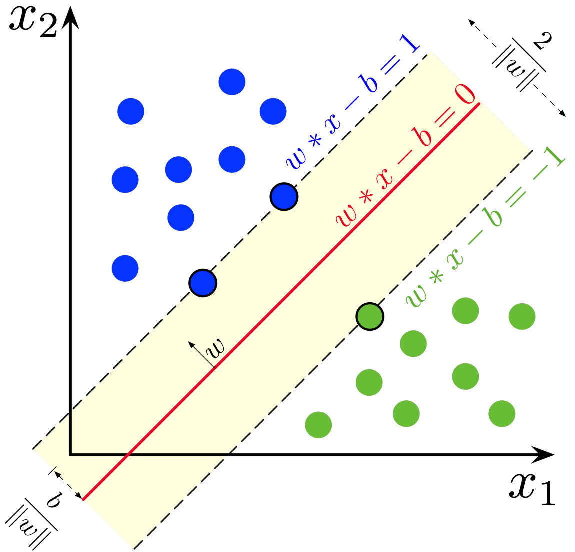
	$\mathcal{P}_{t_k}^{(35)}$	$\mathcal{P}_{t_k}^{(36)}$	$\mathcal{P}_{t_k}^{(37)}$	$\mathcal{P}_{t_k}^{(38)}$	$\mathcal{P}_{t_k}^{(39)}$	$\mathcal{P}_{t_k}^{(40)}$
t_0	0.00	0.00	0.00	0.00	1.00	0.00
t_1	0.00	0.00	0.00	0.00	0.99	0.00
t_2	0.00	0.00	0.00	0.00	0.93	0.00
t_3	0.00	0.00	0.01	0.01	0.85	0.01
t_4	0.00	0.01	0.01	0.01	0.78	0.01
t_5	0.01	0.01	0.01	0.01	0.69	0.01
t_6	0.01	0.02	0.01	0.01	0.63	0.00
t_7	0.01	0.02	0.01	0.01	0.57	0.00
t_8	0.01	0.02	0.01	0.01	0.52	0.00
t_9	0.02	0.02	0.01	0.01	0.45	0.00
t_{10}	0.02	0.02	0.02	0.02	0.37	0.01
t_{11}	0.01	0.02	0.02	0.03	0.29	0.01
t_{12}	0.01	0.01	0.03	0.04	0.20	0.02
t_{13}	0.01	0.01	0.04	0.04	0.14	0.01
t_{14}	0.01	0.02	0.04	0.05	0.08	0.01
t_{15}	0.01	0.02	0.04	0.05	0.06	0.01

- ✔ Example of populations with $t_{15} = 1$

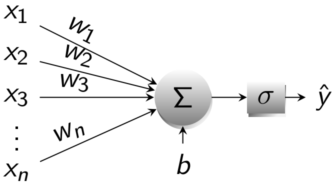


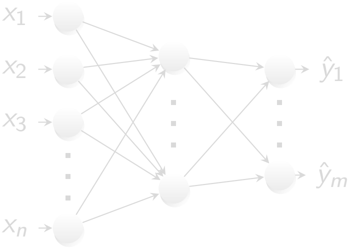
	$\mathcal{P}_{t_k}^{(35)}$	$\mathcal{P}_{t_k}^{(36)}$	$\mathcal{P}_{t_k}^{(37)}$	$\mathcal{P}_{t_k}^{(38)}$	$\mathcal{P}_{t_k}^{(39)}$	$\mathcal{P}_{t_k}^{(40)}$
t_0	0.00	0.00	0.00	0.00	1.00	0.00
t_1	0.02	0.02	0.01	0.01	0.45	0.00
t_2	0.01	0.01	0.02	0.03	0.05	0.02
t_3	0.00	0.00	0.00	0.00	0.13	0.02
t_4	0.02	0.01	0.01	0.01	0.12	0.02
t_5	0.01	0.01	0.03	0.03	0.06	0.01
t_6	0.01	0.01	0.01	0.01	0.01	0.00
t_7	0.04	0.03	0.01	0.06	0.11	0.00
t_8	0.04	0.00	0.03	0.11	0.11	0.03
t_9	0.03	0.00	0.03	0.01	0.01	0.10
t_{10}	0.05	0.01	0.01	0.04	0.08	0.04
t_{11}	0.01	0.03	0.02	0.00	0.08	0.02
t_{12}	0.00	0.05	0.02	0.04	0.00	0.06
t_{13}	0.01	0.03	0.00	0.02	0.05	0.07
t_{14}	0.00	0.00	0.00	0.01	0.12	0.00
t_{15}	0.00	0.00	0.01	0.04	0.10	0.01

Support Vector Machines



Multi Layer Perceptron





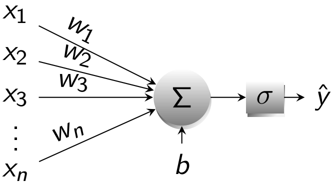
$$\hat{y} \equiv \sigma(\mathbf{w}^T \cdot \mathbf{x} + b)$$

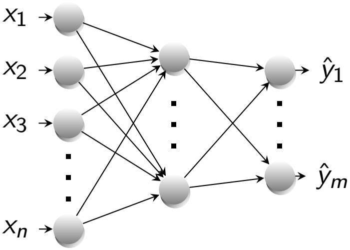
$$\mathbf{h}[0] \equiv \mathbf{x}$$

$$\mathbf{h}[l] \equiv \sigma\left(W[l]^T \cdot \mathbf{h}[l-1] + \mathbf{b}[l]\right)$$

$$\hat{\mathbf{y}} \equiv \mathbf{h}[L]$$

Multi Layer Perceptron





$$\hat{y} \equiv \sigma(\mathbf{w}^T \cdot \mathbf{x} + b)$$

$$\mathbf{h}[0] \equiv \mathbf{x}$$

$$\mathbf{h}[l] \equiv \sigma \left(W[l]^T \cdot \mathbf{h}[l-1] + \mathbf{b}[l] \right)$$

$$\hat{\mathbf{y}} \equiv \mathbf{h}[L]$$

Loss

Categorical Cross Entropy

$$\ell(\hat{\mathbf{y}}, \mathbf{y}) = -\sum_{j=1}^{O} y^{(j)} \log \hat{y}^{(j)}$$

- ✔ Common used in classification tasks
- ✔ Measure distance between two probability distributions
  - $\hat{\mathbf{y}}$ needs to be a probability distributions
  - obtained with softmax function:

$$\sigma^{(i)}(\mathbf{z}) = \frac{e^{z^{(i)}}}{\sum_{j=1}^{O} e^{z^{(j)}}}$$

Loss

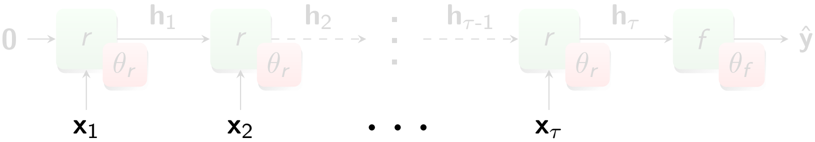
Categorical Cross Entropy

$$\ell(\hat{\mathbf{y}}, \mathbf{y}) = -\sum_{j=1}^{O} y^{(j)} \log \hat{y}^{(j)}$$

- ✔ Common used in classification tasks
- ✔ Measure distance between two probability distributions
  - $\hat{\mathbf{y}}$ needs to be a probability distributions
  - obtained with softmax function:

$$\sigma^{(i)}(\mathbf{z}) \equiv \frac{e^{z^{(i)}}}{\sum_{j=1}^{O} e^{z^{(j)}}}$$

Unidirectional Recurrent Neural Network (RNN)

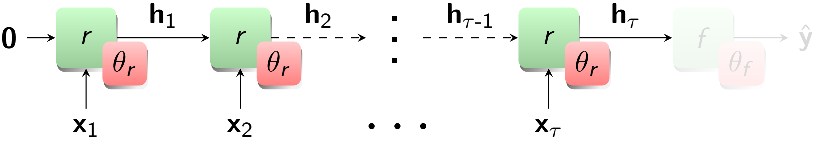


$$\mathbf{h}_t = r(\mathbf{x}_t, \mathbf{h}_{t-1};\, \theta_r)$$

$$\hat{\mathbf{y}} = f(\mathbf{h}_{\mathcal{T}};\, \theta_f)$$

- Used on sequential data
- Processed iteratively by non-linear function r
  - r parametrized with shared set of weights θ_r
  - $\mathbf{h}_t$ sort of memory
- $\mathbf{h}_{\mathcal{T}}$ representation of all the sequence
  - In classification $\mathbf{h}_{\mathcal{T}}$ can be processed by MLP f

Unidirectional RNN

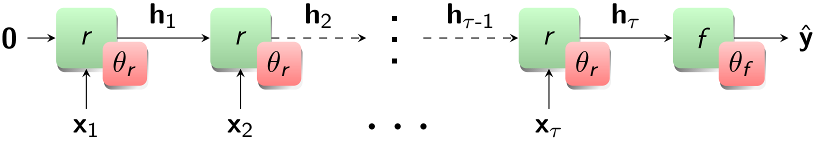


$$\mathbf{h}_t = r(\mathbf{x}_t, \mathbf{h}_{t-1};\, \theta_r)$$

$$\hat{\mathbf{y}} = f(\mathbf{h}_\tau;\, \theta_f)$$

- ✔ Used on sequential data
- ✔ Processed iteratively by non-linear function r
  - ▶ r parametrized with shared set of weights θ_r
  - ▶ $\mathbf{h}_t$ sort of memory
- ✔ $\mathbf{h}_\tau$ representation of all the sequence
  - ▶ In classification $\mathbf{h}_\tau$ can be processed by MLP f

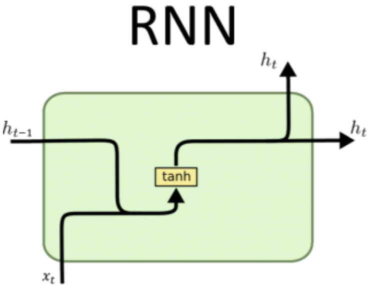
Unidirectional RNN

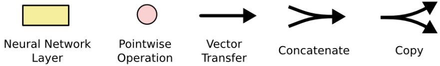


$$\mathbf{h}_t = r(\mathbf{x}_t, \mathbf{h}_{t-1};\, \theta_r)$$
$$\hat{\mathbf{y}} = f(\mathbf{h}_\tau;\, \theta_f)$$

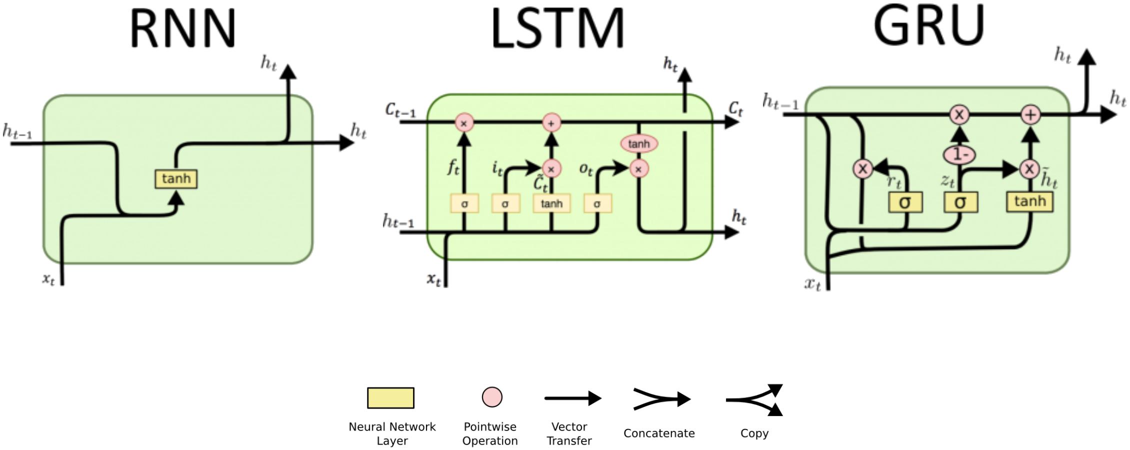
- ✔ Used on sequential data
- ✔ Processed iteratively by non-linear function r
  - r parametrized with shared set of weights θ_r
  - $\mathbf{h}_t$ sort of memory
- ✔ $\mathbf{h}_\tau$ representation of all the sequence
  - In classification $\mathbf{h}_\tau$ can be processed by MLP f

GRU/LSTM

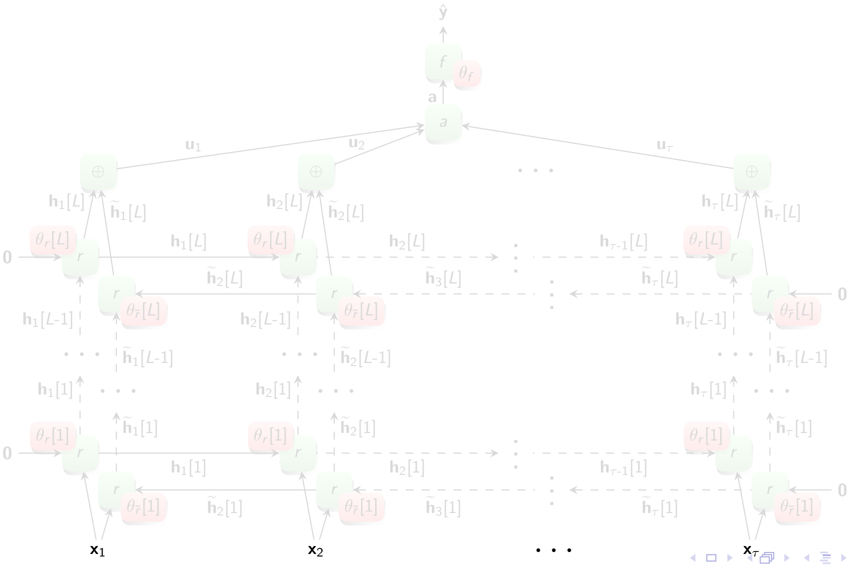

RNN
h_t
h_{t-1}
tanh
h_t
x_t


Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

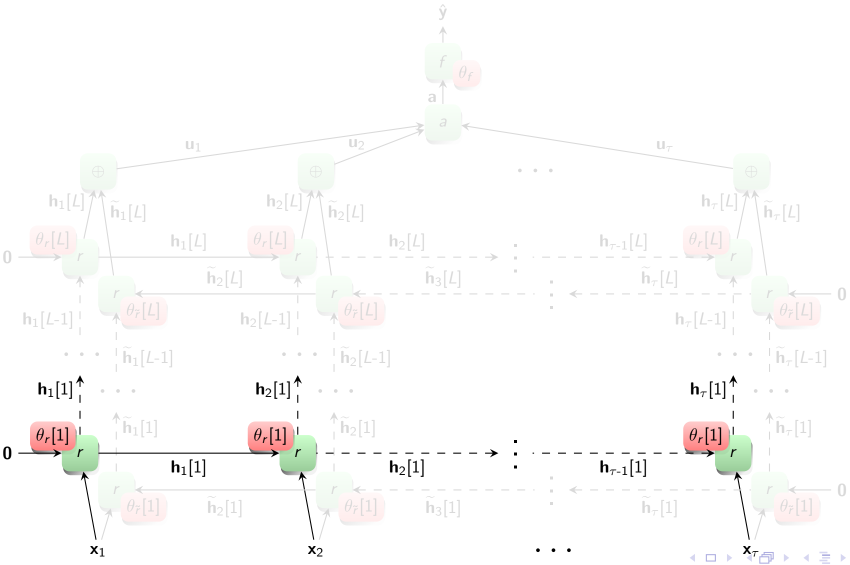
GRU/LSTM


RNN
h_{t-1}
tanh
h_t
h_t
x_t
LSTM
C_{t-1}
C_t
h_t
×
+
tanh
f_t
i_t
$\tilde{C}_t$
o_t
σ
σ
tanh
σ
h_{t-1}
h_t
x_t
GRU
h_t
h_{t-1}
h_t
X
+
1-
r_t
z_t
$\tilde{h}_t$
σ
σ
tanh
x_t
Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

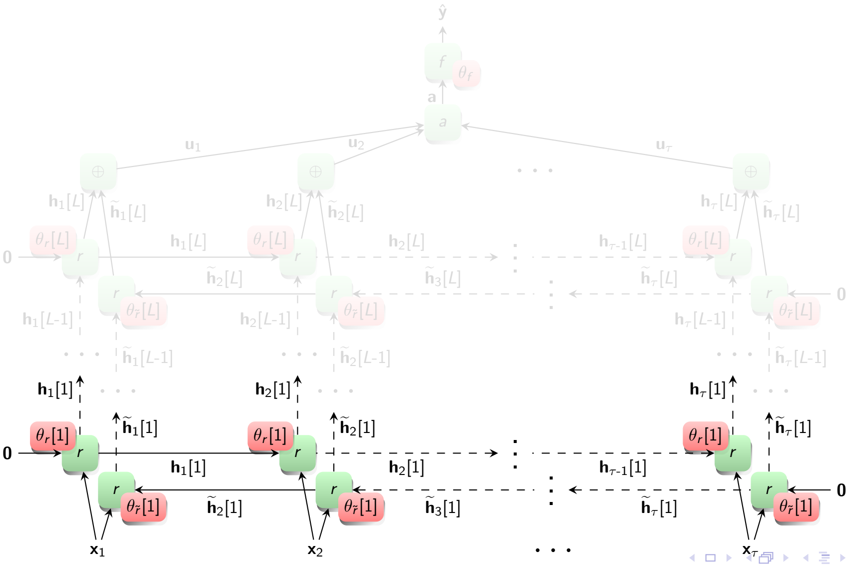
Bidirectional RNN with aggregation



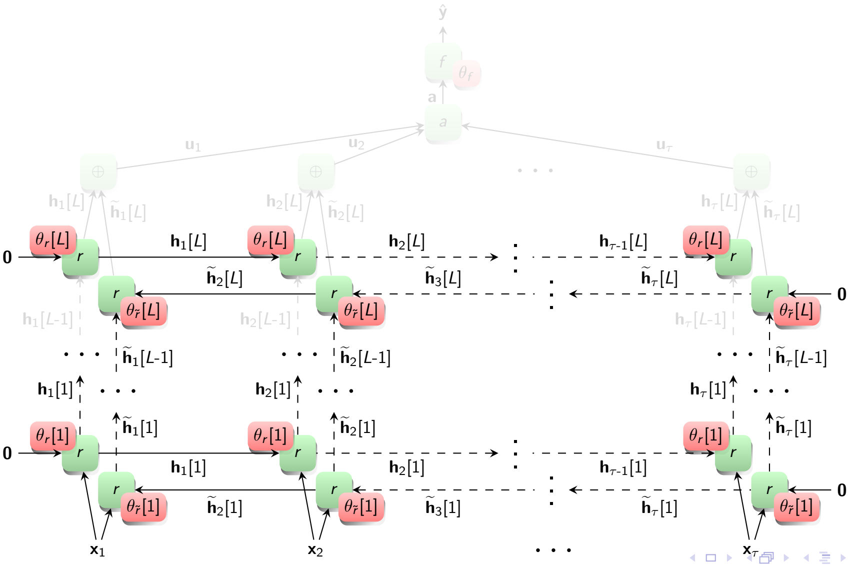
Bidirectional RNN with aggregation



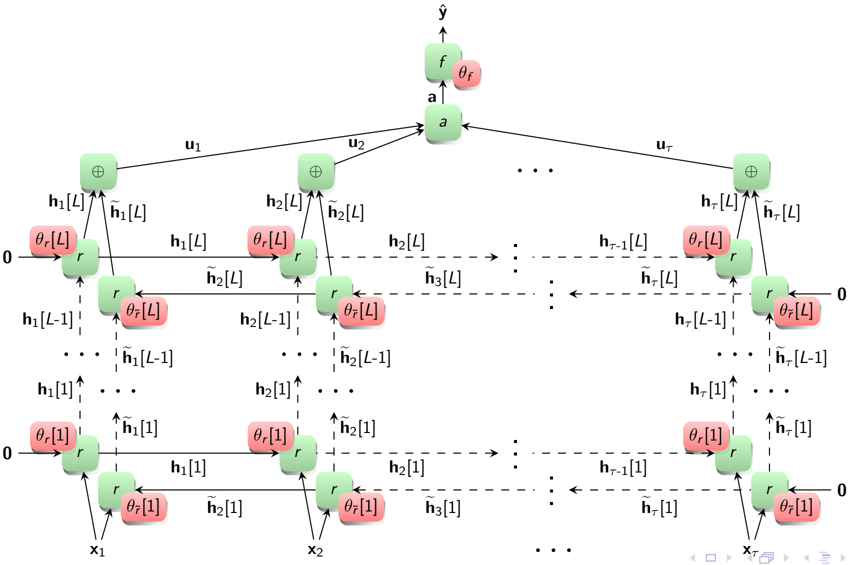
Bidirectional RNN with aggregation



Bidirectional RNN with aggregation



Bidirectional RNN with aggregation



Aggregation

Standard

$$\mathbf{a} = \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_1[L]$$

Max pooling

$$\mathbf{u}_t = \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_t[L]$$

$$\mathbf{a}^{(j)} = \max_t \mathbf{u}_t^{(j)}$$

Attention mechanism

$$\mathbf{u}_t = \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_t[L]$$

$$\mathbf{v}_t = \tanh\left(\mathbf{W}^T \cdot \mathbf{u}_t + \mathbf{b}\right)$$

$$\alpha_t = \frac{e^{\langle \mathbf{v}_t, \mathbf{c} \rangle}}{\sum_{j=1}^{T} e^{\langle \mathbf{v}_j, \mathbf{c} \rangle}}$$

$$\mathbf{a} = \sum_{t=1}^{T} \alpha_t \mathbf{u}_t$$

Aggregation

Standard

$$\mathbf{a} = \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_1[L]$$

Max pooling

$$\begin{aligned} \mathbf{u}_t &= \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_t[L] \\ \mathbf{a}^{(j)} &= \max_t \mathbf{u}_t^{(j)} \end{aligned}$$

Attention mechanism

$$\begin{aligned} \mathbf{u}_t &= \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_t[L] \\ \mathbf{v}_t &= \tanh\left(\mathbf{W}^T \cdot \mathbf{u}_t + \mathbf{b}\right) \\ \alpha_t &= \frac{e^{\langle \mathbf{v}_t, \mathbf{c} \rangle}}{\sum\limits_{j=1}^{T} e^{\langle \mathbf{v}_j, \mathbf{c} \rangle}} \\ \mathbf{a} &= \sum_{t=1}^{T} \alpha_t \mathbf{u}_t \end{aligned}$$

Aggregation

Standard

$$\mathbf{a} = \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_1[L]$$

Max pooling

$$\mathbf{u}_t = \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_t[L]$$

$$\mathbf{a}^{(j)} = \max_t \mathbf{u}_t^{(j)}$$

Attention mechanism

$$\mathbf{u}_t = \mathbf{h}_t[L] \oplus \widetilde{\mathbf{h}}_t[L]$$

$$\mathbf{v}_t = \tanh\left(\mathbf{W}^T \cdot \mathbf{u}_t + \mathbf{b}\right)$$

$$\alpha_t \equiv \frac{e^{\langle \mathbf{v}_t, \mathbf{c} \rangle}}{\sum_{j=1}^{T} e^{\langle \mathbf{v}_j, \mathbf{c} \rangle}}$$

$$\mathbf{a} \equiv \sum_{t=1}^{T} \alpha_t \mathbf{u}_t$$

Results

		$t_{15} = 0.1$			$t_{15} = 1$		
		IID	**NM**	**VS**	**IID**	**NM**	**VS**
$\mathcal{P}_{t_{15}}$	**SVM**	97.0	82.3	96.5	50.3	51.2	49.5
	MLP	96.9	80.7	96.6	49.5	50.7	50.2
$\mathcal{P}_{t_0}, \dots, \mathcal{P}_{t_{15}}$	**SVM**	96.4	80.1	96.3	73.6	61.9	75.0
	GRU	96.5	91.5	96.7	90.5	73.3	88.2
	LSTM	96.8	90.4	96.4	88.6	70.3	86.3
	bi-GRU	96.6	92.2	96.6	91.0	74.6	90.6
	bi-LSTM	96.7	89.7	96.5	90.8	70.6	87.2
	bi-GRU-att	97.0	91.6	96.1	90.9	73.4	87.9
	bi-LSTM-att	96.9	87.9	96.3	89.0	71.6	87.4
	bi-GRU-max	96.6	92.6	96.6	91.8	76.1	90.4
	bi-LSTM-max	96.6	91.4	96.3	91.4	74.9	89.0

Results

		$t_{15} = 0.1$			$t_{15} = 1$		
		IID	**NM**	**VS**	**IID**	**NM**	**VS**
$\mathcal{P}_{t_{15}}$	**SVM**	97.0	82.3	96.5	50.3	51.2	49.5
	MLP	96.9	80.7	96.6	49.5	50.7	50.2
$\mathcal{P}_{t_0}, \dots, \mathcal{P}_{t_{15}}$	**SVM**	96.4	80.1	96.3	73.6	61.9	75.0
	GRU	96.5	91.5	96.7	90.5	73.3	88.2
	LSTM	96.8	90.4	96.4	88.6	70.3	86.3
	bi-GRU	96.6	92.2	96.6	91.0	74.6	90.6
	bi-LSTM	96.7	89.7	96.5	90.8	70.6	87.2
	bi-GRU-att	97.0	91.6	96.1	90.9	73.4	87.9
	bi-LSTM-att	96.9	87.9	96.3	89.0	71.6	87.4
	bi-GRU-max	96.6	92.6	96.6	91.8	76.1	90.4
	bi-LSTM-max	96.6	91.4	96.3	91.4	74.9	89.0

Results

		$t_{15} = 0.1$			$t_{15} = 1$		
		IID	**NM**	**VS**	**IID**	**NM**	**VS**
$\mathcal{P}_{t_{15}}$	**SVM**	97.0	82.3	96.5	50.3	51.2	49.5
	MLP	96.9	80.7	96.6	49.5	50.7	50.2
$\mathcal{P}_{t_0}, \ldots, \mathcal{P}_{t_{15}}$	**SVM**	96.4	80.1	96.3	73.6	61.9	75.0
	GRU	96.5	91.5	96.7	90.5	73.3	88.2
	LSTM	96.8	90.4	96.4	88.6	70.3	86.3
	bi-GRU	96.6	92.2	96.6	91.0	74.6	90.6
	bi-LSTM	96.7	89.7	96.5	90.8	70.6	87.2
	bi-GRU-att	97.0	91.6	96.1	90.9	73.4	87.9
	bi-LSTM-att	96.9	87.9	96.3	89.0	71.6	87.4
	bi-GRU-max	96.6	92.6	96.6	91.8	76.1	90.4
	bi-LSTM-max	96.6	91.4	96.3	91.4	74.9	89.0

Results

		$t_{15} = 0.1$			$t_{15} = 1$		
		IID	**NM**	**VS**	**IID**	**NM**	**VS**
$\mathcal{P}_{t_{15}}$	**SVM**	97.0	82.3	96.5	50.3	51.2	49.5
	MLP	96.9	80.7	96.6	49.5	50.7	50.2
$\mathcal{P}_{t_0}, \dots, \mathcal{P}_{t_{15}}$	**SVM**	96.4	80.1	96.3	73.6	61.9	75.0
	GRU	96.5	91.5	96.7	90.5	73.3	88.2
	LSTM	96.8	90.4	96.4	88.6	70.3	86.3
	bi-GRU	96.6	92.2	96.6	91.0	74.6	90.6
	bi-LSTM	96.7	89.7	96.5	90.8	70.6	87.2
	bi-GRU-att	97.0	91.6	96.1	90.9	73.4	87.9
	bi-LSTM-att	96.9	87.9	96.3	89.0	71.6	87.4
	bi-GRU-max	96.6	92.6	96.6	91.8	76.1	90.4
	bi-LSTM-max	96.6	91.4	96.3	91.4	74.9	89.0

Results

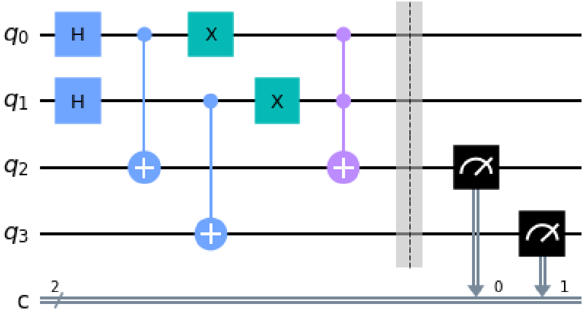
		$t_{15} = 0.1$			$t_{15} = 1$		
		IID	**NM**	**VS**	**IID**	**NM**	**VS**
$\mathcal{P}_{t_{15}}$	**SVM**	97.0	82.3	96.5	50.3	51.2	49.5
	MLP	96.9	80.7	96.6	49.5	50.7	50.2
$\mathcal{P}_{t_0}, \ldots, \mathcal{P}_{t_{15}}$	**SVM**	96.4	80.1	96.3	73.6	61.9	75.0
	GRU	96.5	91.5	96.7	90.5	73.3	88.2
	LSTM	96.8	90.4	96.4	88.6	70.3	86.3
	bi-GRU	96.6	92.2	96.6	91.0	74.6	90.6
	bi-LSTM	96.7	89.7	96.5	90.8	70.6	87.2
	bi-GRU-att	97.0	91.6	96.1	90.9	73.4	87.9
	bi-LSTM-att	96.9	87.9	96.3	89.0	71.6	87.4
	bi-GRU-max	96.6	92.6	96.6	91.8	76.1	90.4
	bi-LSTM-max	96.6	91.4	96.3	91.4	74.9	89.0

Results

		$t_{15} = 0.1$			$t_{15} = 1$		
		IID	**NM**	**VS**	**IID**	**NM**	**VS**
$\mathcal{P}_{t_{15}}$	**SVM**	97.0	82.3	96.5	50.3	51.2	49.5
	MLP	96.9	80.7	96.6	49.5	50.7	50.2
$\mathcal{P}_{t_0}, \dots, \mathcal{P}_{t_{15}}$	**SVM**	96.4	80.1	96.3	73.6	61.9	75.0
	GRU	96.5	91.5	96.7	90.5	73.3	88.2
	LSTM	96.8	90.4	96.4	88.6	70.3	86.3
	bi-GRU	96.6	92.2	96.6	91.0	74.6	90.6
	bi-LSTM	96.7	89.7	96.5	90.8	70.6	87.2
	bi-GRU-att	97.0	91.6	96.1	90.9	73.4	87.9
	bi-LSTM-att	96.9	87.9	96.3	89.0	71.6	87.4
	bi-GRU-max	96.6	92.6	96.6	91.8	76.1	90.4
	bi-LSTM-max	96.6	91.4	96.3	91.4	74.9	89.0

Used circuit

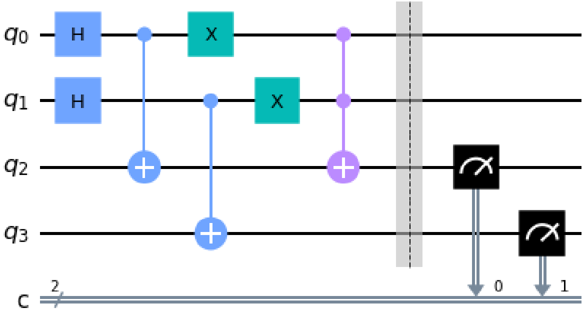
Martina, S., Buffoni, L., Gherardini, S., & Caruso, F. (2022). Learning the noise fingerprint of quantum devices. Quantum Machine Intelligence, 4(1), 1-12.



- ✔ We add a temporal dimension
- ✔ The circuit is repeated 3 times
- ✔ The measurements are after each CNOT and Toffoli for a total of 9 steps

Used circuit

Martina, S., Buffoni, L., Gherardini, S., & Caruso, F. (2022). Learning the noise fingerprint of quantum devices. Quantum Machine Intelligence, 4(1), 1-12.



- ✔ We add a temporal dimension
- ✔ The circuit is repeated 3 times
- ✔ The measurements are after each CNOT and Toffoli for a total of 9 steps

FAST dataset

- ✓ Temporally close executions
- ✓ 7 different IBM NISQ devices
- ✓ For each device 2000 sequences of 9 steps
  - each one is a distribution probability obtained running 1000 shots of the circuit

FAST dataset

- ✓ Temporally close executions
- ✓ 7 different IBM NISQ devices
- ✓ For each device 2000 sequences of 9 steps
  - each one is a distribution probability obtained running 1000 shots of the circuit

FAST dataset

- ✓ Temporally close executions
- ✓ 7 different IBM NISQ devices
- ✓ For each device 2000 sequences of 9 steps
  - each one is a distribution probability obtained running 1000 shots of the circuit

Binary device classification

- ✓ Using FAST dataset
- ✓ For each pair of devices
  - the task is to find the used device from the measurement probabilities
- ✓ Both considering single and incremental steps
- ✓ Trained SVM classifiers
  - 60% – 20% – 20% train, validation and test split
  - kernels linear, RBF and polynomial with degree 2, 3 and 4

Binary device classification

- ✓ Using FAST dataset
- ✓ For each pair of devices
  - the task is to find the used device from the measurement probabilities
- ✓ Both considering single and incremental steps
- ✓ Trained SVM classifiers
  - 60% – 20% – 20% train, validation and test split
  - kernels linear, RBF and polynomial with degree 2, 3 and 4

Binary device classification

- ✓ Using FAST dataset
- ✓ For each pair of devices
  - ▶ the task is to find the used device from the measurement probabilities
- ✓ Both considering single and incremental steps
- ✓ Trained SVM classifiers
  - ▶ 60% – 20% – 20% train, validation and test split
  - ▶ kernels linear, RBF and polynomial with degree 2, 3 and 4

Binary device classification

- ✓ Using FAST dataset
- ✓ For each pair of devices
  - the task is to find the used device from the measurement probabilities
- ✓ Both considering single and incremental steps
- ✓ Trained SVM classifiers
  - 60% – 20% – 20% train, validation and test split
  - kernels linear, RBF and polynomial with degree 2, 3 and 4

Machines	Athens			Bogota			Casablanca			Lima			Quito			Santiago		
	k	$\alpha(k)$	$\alpha([1,k])$	k	$\alpha(k)$	$\alpha([1,k])$	k	$\alpha(k)$	$\alpha([1,k])$	k	$\alpha(k)$	$\alpha([1,k])$	k	$\alpha(k)$	$\alpha([1,k])$	k	$\alpha(k)$	$\alpha([1,k])$
Bogota	1	0.915	0.915															
	2	0.944	0.975															
	3	0.999	1.000															
	4	0.954	0.999															
	5	0.981	0.999															
	6	0.989	1.000															
	7	0.949	1.000															
	8	0.990	1.000															
	9	0.991	1.000															
Casablanca	1	0.895	0.895	1	0.831	0.831												
	2	0.740	0.921	2	0.932	0.959												
	3	0.968	0.983	3	0.943	0.995												
	4	0.994	0.998	4	0.960	0.999												
	5	0.969	0.999	5	0.889	0.998												
	6	0.988	1.000	6	0.811	1.000												
	7	0.869	1.000	7	0.830	0.999												
	8	0.906	1.000	8	0.818	0.999												
	9	0.927	1.000	9	0.782	1.000												
Lima	1	0.879	0.879	1	0.772	0.772	1	0.724	0.724									
	2	0.762	0.915	2	0.983	0.984	2	0.869	0.887									
	3	0.999	1.000	3	0.989	0.999	3	0.951	0.966									
	4	1.000	1.000	4	0.996	1.000	4	0.829	0.975									
	5	0.999	1.000	5	0.787	1.000	5	0.814	0.993									
	6	0.999	1.000	6	0.996	1.000	6	0.990	0.999									
	7	0.940	1.000	7	0.795	1.000	7	0.882	0.999									
	8	0.784	1.000	8	0.912	1.000	8	0.823	0.999									
	9	0.978	1.000	9	0.950	1.000	9	0.879	0.999									
Quito	1	0.685	0.685	1	0.815	0.815	1	0.834	0.834	1	0.725	0.725						
	2	1.000	1.000	2	1.000	1.000	2	1.000	1.000	2	1.000	1.000						
	3	1.000	1.000	3	0.990	1.000	3	1.000	1.000	3	1.000	1.000						
	4	0.998	1.000	4	1.000	1.000	4	1.000	1.000	4	1.000	1.000						
	5	0.993	1.000	5	0.787	1.000	5	0.881	1.000	5	0.714	1.000						
	6	0.966	1.000	6	0.983	1.000	6	0.979	1.000	6	1.000	1.000						
	7	0.948	1.000	7	0.965	1.000	7	0.940	1.000	7	0.978	1.000						
	8	0.998	1.000	8	0.969	1.000	8	0.959	1.000	8	0.991	1.000						
	9	0.988	1.000	9	0.891	1.000	9	0.864	1.000	9	0.953	1.000						

Multiclass device classification

- ✔ Not considering pairs, but a multiclass setting (one-vs-rest)

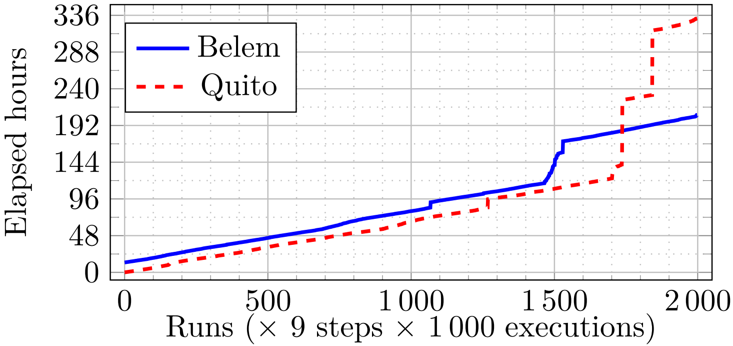
Machines	k	$\alpha(k)$	$\alpha([k-1,k])$	$\alpha([k-2,k])$	$\alpha([k-3,k])$	$\alpha([k-4,k])$	$\alpha([1,k])$
Athens & Bogota & Casablanca & Lima & Quito & Santiago & Yorktown	1	0.529					0.529
	2	0.691	0.850				0.850
	3	0.920	0.975	0.983			0.983
	4	0.896	0.983	0.991	0.992		0.992
	5	0.680	0.955	0.992	0.995	0.995	0.995
	6	0.789	0.946	0.988	0.997	0.998	0.998
	7	0.776	0.941	0.974	0.993	0.998	0.999
	8	0.703	0.911	0.960	0.982	0.994	0.999
	9	0.681	0.871	0.952	0.970	0.986	0.999
Average		0.740	0.929	0.977	0.988	0.994	

Time classification

- 2 runs of FAST dataset on the same machine but with 24 hours gap

Machine	k	$\alpha(k)$	$\alpha([1,k])$
Casablanca	1	0.882	0.882
	2	0.815	0.917
	3	0.757	0.948
	4	0.974	0.994
	5	0.969	1.000
	6	0.895	0.999
	7	0.917	0.999
	8	0.859	0.999
	9	0.721	0.999

SLOW dataset



- ✔ Temporally delayed executions (at least 2 minutes between each 1000 shots batch)
- ✔ 2 different IBM NISQ devices
- ✔ For each device 2000 sequences of 9 steps
  - each one is a distribution probability obtained running 1000 shots of the circuit

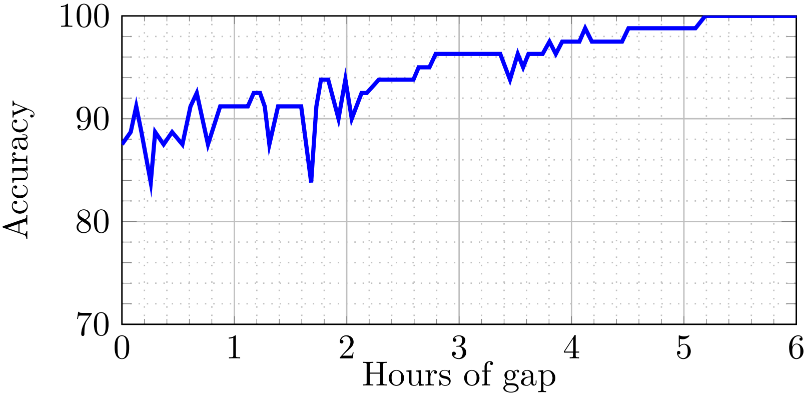
Time window classification

- ✔ in 1 machine of SLOW dataset
- ✔ discriminate between the first window of 200 runs and the subsequent windows

$[1, 200]$ vs		$[201, 400]$	$[401, 600]$	$[601, 800]$	$[801, 1000]$	$[1001, 1200]$	$[1201, 1400]$	$[1401, 1600]$	$[1601, 1800]$	$[1801, 2000]$
Machines	k	$\alpha(k)$	$\alpha(k)$	$\alpha(k)$	$\alpha(k)$	$\alpha(k)$	$\alpha(k)$	$\alpha(k)$	$\alpha(k)$	$\alpha(k)$
Belem	1	0.838	0.975	0.975	0.950	0.938	0.938	0.750	0.950	0.963
	2	0.812	0.850	0.912	0.875	0.975	0.925	0.800	0.863	0.875
	3	0.688	0.812	0.688	0.738	0.650	0.500	0.738	0.613	0.700
	4	0.738	0.800	0.700	0.750	0.700	0.713	0.863	0.875	0.875
	5	0.662	0.700	0.800	0.800	0.725	0.863	0.762	0.838	0.812
	6	0.700	0.700	0.938	0.950	0.838	0.762	0.800	0.750	0.800
	7	0.675	0.850	0.887	0.975	0.912	0.887	0.713	0.875	0.950
	8	0.775	0.800	0.900	0.912	0.938	0.988	0.787	0.938	0.938
	9	0.750	0.900	0.912	0.988	0.850	0.838	0.787	0.812	0.838
Average		0.738	0.821	0.857	0.882	0.837	0.824	0.778	0.835	0.861

$[1, 200]$ vs		$[201, 400]$	$[401, 600]$	$[601, 800]$	$[801, 1000]$	$[1001, 1200]$	$[1201, 1400]$	$[1401, 1600]$	$[1601, 1800]$	$[1801, 2000]$
Machines	k	$\alpha([1, k])$	$\alpha([1, k])$	$\alpha([1, k])$	$\alpha([1, k])$	$\alpha([1, k])$	$\alpha([1, k])$	$\alpha([1, k])$	$\alpha([1, k])$	$\alpha([1, k])$
Belem	1	0.838	0.975	0.975	0.950	0.938	0.938	0.750	0.950	0.963
	2	0.850	0.963	0.988	0.975	1.000	0.950	0.825	0.988	0.988
	3	0.887	0.975	0.988	0.975	0.988	0.988	0.850	1.000	0.988
	4	0.850	0.950	1.000	0.988	0.988	0.975	0.975	0.988	1.000
	5	0.850	0.963	1.000	0.988	0.988	0.975	0.963	1.000	1.000
	6	0.850	0.988	0.988	0.988	1.000	0.988	0.975	1.000	0.988
	7	0.863	0.988	1.000	0.988	1.000	1.000	0.988	1.000	1.000
	8	0.850	1.000	1.000	0.988	1.000	1.000	0.975	1.000	1.000
	9	0.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

- ✓ in same machine as previous slide
- ✓ $\alpha([1,9])$ discriminating the first window of 200 runs from another window sliding in time



Robustness in time

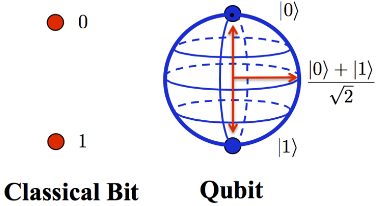
- Both machines of SLOW dataset
- $\alpha([1,9])$ discriminating the used machine
- Train on the window in row index; test on window in column index

	1	2	3	4	5	6	7	8	9	10
1	1.000	1.000	0.995	0.925	0.880	0.865	0.995	1.000	1.000	1.000
2	1.000	1.000	0.995	0.925	0.920	0.910	0.980	1.000	1.000	1.000
3	1.000	1.000	1.000	0.970	0.950	0.950	0.980	1.000	1.000	1.000
4	1.000	0.980	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.980	0.935	0.955	0.995	1.000	0.995	1.000	1.000	1.000	1.000
6	0.995	0.995	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1.000	1.000	0.995	0.985	1.000	0.990	1.000	1.000	1.000	1.000
8	1.000	1.000	0.995	0.995	1.000	0.990	0.995	1.000	1.000	1.000
9	1.000	1.000	0.995	0.995	0.970	0.960	1.000	1.000	1.000	1.000
10	1.000	1.000	0.995	0.995	0.995	0.995	0.995	1.000	1.000	1.000

Thank you! Questions?

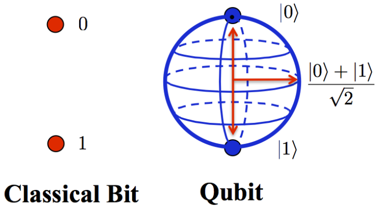
Qubit

- Classic bit can take one value between 0 and 1
- A qubit can take one of infinite values
  - in Hilbert vector space with basis of two elements $|0\rangle$ and $|1\rangle$
- A qubit is in superposition $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$
  - Where amplitudes α and β are complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$



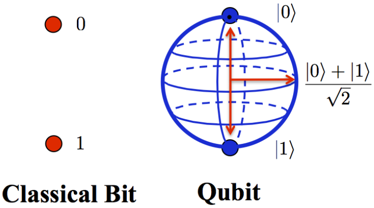
Qubit

- Classic bit can take one value between 0 and 1
- A qubit can take one of infinite values
  - in Hilbert vector space with basis of two elements $|0\rangle$ and $|1\rangle$
- A qubit is in superposition $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$
  - Where amplitudes α and β are complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$



Qubit

- ✔ Classic bit can take one value between 0 and 1
- ✔ A qubit can take one of infinite values
  - in Hilbert vector space with basis of two elements $|0\rangle$ and $|1\rangle$
- ✔ A qubit is in superposition $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$
  - Where amplitudes α and β are complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$



Measure

- ✔ The result of the measure is random
- ✔ When we measure a qubit we obtain a classical bit
- ✔ The measure of $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is
  - 0 with probability $|\alpha|^2$
  - 1 with probability $|\beta|^2$

Effect of measure

Wavefunction collapse the new state after the measurement will be $|0\rangle$ or $|1\rangle$ depending on the measurement result

No-cloning theorem We cannot perform several independent measurements of $|\psi\rangle$

Measure

- ✔ The result of the measure is random
- ✔ When we measure a qubit we obtain a classical bit
- ✔ The measure of $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is
  - ▸ 0 with probability $|\alpha|^2$
  - ▸ 1 with probability $|\beta|^2$

Effect of measure

Wavefunction collapse the new state after the measurement will be $|0\rangle$ or $|1\rangle$ depending on the measurement result

No-cloning theorem We cannot perform several independent measurements of $|\psi\rangle$

Measure

- The result of the measure is random
- When we measure a qubit we obtain a classical bit
- The measure of $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$ is
  - 0 with probability $|\alpha|^2$
  - 1 with probability $|\beta|^2$

Effect of measure

Wavefunction collapse the new state after the measurement will be $|0\rangle$ or $|1\rangle$ depending on the measurement result

No-cloning theorem We cannot perform several independent measurements of $|\psi\rangle$

Measure

- ✓ The result of the measure is random
- ✓ When we measure a qubit we obtain a classical bit
- ✓ The measure of $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is
  - 0 with probability $|\alpha|^2$
  - 1 with probability $|\beta|^2$

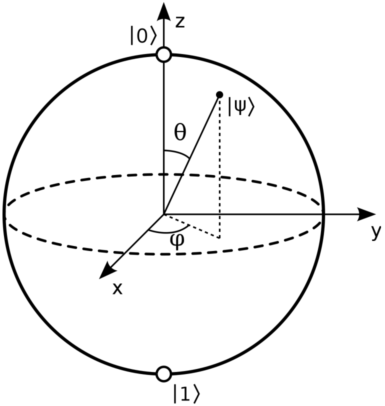
Effect of measure

Wavefunction collapse the new state after the measurement will be $|0\rangle$ or $|1\rangle$ depending on the measurement result

No-cloning theorem We cannot perform several independent measurements of $|\psi\rangle$

Bloch sphere

- ✔ We can rewrite $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle \rightarrow |\psi\rangle = \cos\frac{\theta}{2} |0\rangle + e^{i\varphi} \sin\frac{\theta}{2} |1\rangle$
  - with $0 \leq \theta \leq \pi$ and $0 \leq \varphi < 2\pi$



Quantum gates

- The evolution of a state is given by the Schrödinger equation

$$H(t)\left|\psi(t)\right\rangle = i\hbar\frac{\partial}{\partial t}\left|\psi(t)\right\rangle$$

- In quantum circuits, the operation given by complex unitary matrices, i.e. verifying

$$UU^\dagger = U^\dagger U = I$$

  where $U^\dagger$ is the complex conjugate transpose of U
- Each such matrix is a possible quantum gate in a quantum circuit

Application (for 1-qubit gate)

For $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\left|\psi\right\rangle = \alpha\left|0\right\rangle + \beta\left|1\right\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$$U\left|\psi\right\rangle = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} a\alpha + b\beta \\ c\alpha + d\beta \end{pmatrix} = (a\alpha + b\beta)\left|0\right\rangle + (c\alpha + d\beta)\left|1\right\rangle$$

Quantum gates

- The evolution of a state is given by the Schrödinger equation

$$H(t)\left|\psi(t)\right\rangle = i\hbar\frac{\partial}{\partial t}\left|\psi(t)\right\rangle$$

- In quantum circuits, the operation given by complex unitary matrices, i.e. verifying

$$UU^\dagger = U^\dagger U = I$$

  where $U^\dagger$ is the complex conjugate transpose of U
- Each such matrix is a possible quantum gate in a quantum circuit

Application (for 1-qubit gate)

For $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\left|\psi\right\rangle = \alpha\left|0\right\rangle + \beta\left|1\right\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$$U\left|\psi\right\rangle = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} a\alpha + b\beta \\ c\alpha + d\beta \end{pmatrix} = (a\alpha + b\beta)\left|0\right\rangle + (c\alpha + d\beta)\left|1\right\rangle$$

Quantum gates

- The evolution of a state is given by the Schrödinger equation

$$H(t)\,|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}\,|\psi(t)\rangle$$

- In quantum circuits, the operation given by complex unitary matrices, i.e. verifying

$$UU^\dagger = U^\dagger U = I$$

  where $U^\dagger$ is the complex conjugate transpose of U
- Each such matrix is a possible quantum gate in a quantum circuit

Application (for 1-qubit gate)

For $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$$U\,|\psi\rangle = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} a\alpha + b\beta \\ c\alpha + d\beta \end{pmatrix} = (a\alpha + b\beta)\,|0\rangle + (c\alpha + d\beta)\,|1\rangle$$

Pauli Gates (rotation of π along correspondig axis in Bloch)

X or NOT (Pauli σ_X)

$$X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\alpha\,|0\rangle + \beta\,|1\rangle \;—\boxed{X}—\; \beta\,|0\rangle + \alpha\,|1\rangle$$

Y (Pauli σ_Y)

$$Y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$\alpha\,|0\rangle + \beta\,|1\rangle \;—\boxed{Y}—\; -i\beta\,|0\rangle + i\alpha\,|1\rangle$$

Z (Pauli σ_Z)

$$Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\alpha\,|0\rangle + \beta\,|1\rangle \;—\boxed{Z}—\; \alpha\,|0\rangle - \beta\,|1\rangle$$

I

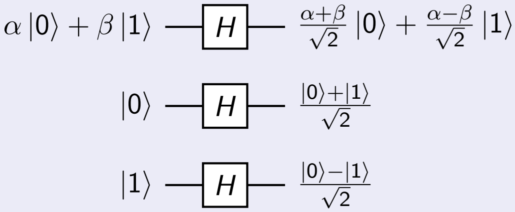
$$I := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\alpha\,|0\rangle + \beta\,|1\rangle \;—\boxed{I}—\; \alpha\,|0\rangle + \beta\,|1\rangle$$

Hadamard gate

H

$$H := \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$$



$$\alpha\left|0\right\rangle + \beta\left|1\right\rangle \; \text{---}\boxed{H}\text{---} \; \tfrac{\alpha+\beta}{\sqrt{2}}\left|0\right\rangle + \tfrac{\alpha-\beta}{\sqrt{2}}\left|1\right\rangle$$

$$\left|0\right\rangle \; \text{---}\boxed{H}\text{---} \; \tfrac{\left|0\right\rangle+\left|1\right\rangle}{\sqrt{2}}$$

$$\left|1\right\rangle \; \text{---}\boxed{H}\text{---} \; \tfrac{\left|0\right\rangle-\left|1\right\rangle}{\sqrt{2}}$$

$$\left|+\right\rangle := \frac{\left|0\right\rangle + \left|1\right\rangle}{\sqrt{2}} \qquad\qquad \left|-\right\rangle := \frac{\left|0\right\rangle - \left|1\right\rangle}{\sqrt{2}}$$

2 qubit systems

- ✔ Each qubit can be in state $|0\rangle$ or $|1\rangle$
- ✔ We have 4 possibilities, equivalently ($\otimes$ is Kroneker product)

$$|0\rangle \otimes |0\rangle, \quad |0\rangle \otimes |1\rangle, \quad |1\rangle \otimes |0\rangle, \quad |1\rangle \otimes |1\rangle$$
$$|0\rangle |0\rangle, \quad |0\rangle |1\rangle, \quad |1\rangle |0\rangle, \quad |1\rangle |1\rangle$$
$$|00\rangle, \quad |01\rangle, \quad |10\rangle, \quad |11\rangle$$

- ✔ We can have superposition

$$|\psi\rangle = \alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle$$

with amplitudes α_{xy} complex numbers such that $\sum_{x,y=0}^{1} |\alpha_{xy}|^2 = 1$

2 qubit systems

- Each qubit can be in state $|0\rangle$ or $|1\rangle$
- We have 4 possibilities, equivalently ($\otimes$ is Kroneker product)

$$|0\rangle \otimes |0\rangle\,, \quad |0\rangle \otimes |1\rangle\,, \quad |1\rangle \otimes |0\rangle\,, \quad |1\rangle \otimes |1\rangle$$

$$|0\rangle\,|0\rangle\,, \quad |0\rangle\,|1\rangle\,, \quad |1\rangle\,|0\rangle\,, \quad |1\rangle\,|1\rangle$$

$$|00\rangle\,, \quad |01\rangle\,, \quad |10\rangle\,, \quad |11\rangle$$

- We can have superposition

$$|\psi\rangle = \alpha_{00}\,|00\rangle + \alpha_{01}\,|01\rangle + \alpha_{10}\,|10\rangle + \alpha_{11}\,|11\rangle$$

with amplitudes α_{xy} complex numbers such that $\sum_{x,y=0}^{1} |\alpha_{xy}|^2 = 1$

2 qubit systems

- ✔ Each qubit can be in state $|0\rangle$ or $|1\rangle$
- ✔ We have 4 possibilities, equivalently ($\otimes$ is Kroneker product)

$$\begin{gathered} |0\rangle \otimes |0\rangle\,, \quad |0\rangle \otimes |1\rangle\,, \quad |1\rangle \otimes |0\rangle\,, \quad |1\rangle \otimes |1\rangle \\ |0\rangle\,|0\rangle\,, \quad |0\rangle\,|1\rangle\,, \quad |1\rangle\,|0\rangle\,, \quad |1\rangle\,|1\rangle \\ |00\rangle\,, \quad |01\rangle\,, \quad |10\rangle\,, \quad |11\rangle \end{gathered}$$

- ✔ We can have superposition

$$|\psi\rangle = \alpha_{00}\,|00\rangle + \alpha_{01}\,|01\rangle + \alpha_{10}\,|10\rangle + \alpha_{11}\,|11\rangle$$

with amplitudes α_{xy} complex numbers such that $\sum_{x,y=0}^{1} |\alpha_{xy}|^2 = 1$

Measuring 2 qubit systems

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$$

Measuring both qubits

- ✓ 00 with probability $|\alpha_{00}|^2$, new state $|00\rangle$
- ✓ 01 with probability $|\alpha_{01}|^2$, new state $|01\rangle$
- ✓ 10 with probability $|\alpha_{10}|^2$, new state $|10\rangle$
- ✓ 11 with probability $|\alpha_{11}|^2$, new state $|11\rangle$

Measuring only one qubit (the first in this case)

- ✓ 0 with probability $|\alpha_{00}|^2 + |\alpha_{01}|^2$, new state $\frac{\alpha_{00}|00\rangle+\alpha_{01}|01\rangle}{\sqrt{|\alpha_{00}|^2+|\alpha_{01}|^2}}$
- ✓ 1 with probability $|\alpha_{10}|^2 + |\alpha_{11}|^2$, new state $\frac{\alpha_{10}|10\rangle+\alpha_{11}|11\rangle}{\sqrt{|\alpha_{10}|^2+|\alpha_{11}|^2}}$

Measuring 2 qubit systems

$$|\psi\rangle = \alpha_{00}\,|00\rangle + \alpha_{01}\,|01\rangle + \alpha_{10}\,|10\rangle + \alpha_{11}\,|11\rangle$$

Measuring both qubits

- ✔ 00 with probability $|\alpha_{00}|^2$, new state $|00\rangle$
- ✔ 01 with probability $|\alpha_{01}|^2$, new state $|01\rangle$
- ✔ 10 with probability $|\alpha_{10}|^2$, new state $|10\rangle$
- ✔ 11 with probability $|\alpha_{11}|^2$, new state $|11\rangle$

Measuring only one qubit (the first in this case)

- ✔ 0 with probability $|\alpha_{00}|^2 + |\alpha_{01}|^2$, new state $\frac{\alpha_{00}|00\rangle + \alpha_{01}|01\rangle}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}$
- ✔ 1 with probability $|\alpha_{10}|^2 + |\alpha_{11}|^2$, new state $\frac{\alpha_{10}|10\rangle + \alpha_{11}|11\rangle}{\sqrt{|\alpha_{10}|^2 + |\alpha_{11}|^2}}$

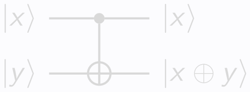
2 qubit gates

- ✔ If A and B are one-qubit gates acting on two different cubits, then on the two qubit $A \otimes B$
- ✔ In general all unitary matrices 4×4

CNOT

$$CNOT := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

for $x, y \in \{0, 1\}$



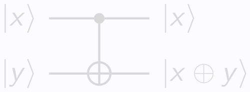
2 qubit gates

- ✔ If A and B are one-qubit gates acting on two different cubits, then on the two qubit $A \otimes B$
- ✔ In general all unitary matrices 4×4

CNOT

$$CNOT := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

for $x, y \in \{0, 1\}$



2 qubit gates

- ✔ If A and B are one-qubit gates acting on two different cubits, then on the two qubit $A \otimes B$
- ✔ In general all unitary matrices 4×4

CNOT

$$CNOT := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

for $x, y \in \{0, 1\}$