Università degli Studi di Bari Aldo Moro
Istituto tumori "Giovanni Paolo II" - IRCCS
INFN Istituto Nazionale di Fisica Nucleare

Quantum-inspired machine learning

Domenico Pomarico
Research fellowship QuantHEP

Paper: MDPI Mathematics 2021, 9(4), 410
DOI: 10.3390/math9040410

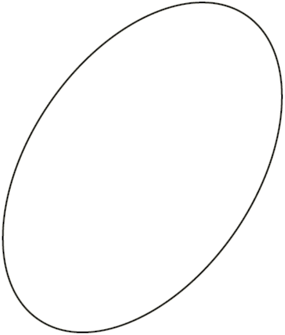
in collaboration with A. Fanizzi, N. Amoroso, R. Bellotti, A. Biafora, S. Bove, V. Didonna, D. La Forgia, M. I. Pastena, P. Tamborra, A. Zito, V. Lorusso, R. Massafra.

Contents

- Classification systems theory
  - Combine devices, data elaboration & classification
  - SVM feature maps
- Tree tensor networks
  - Building up a circuit with qubits
  - Variational quantum algorithms
- Our application case
  - Summary about breast cancer biology
  - Feature selection & performances

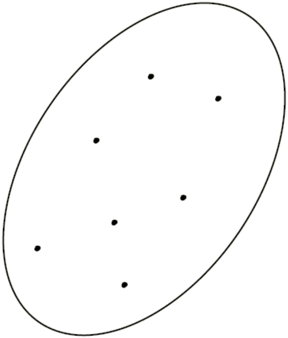
Classification systems theory

population set
of outcomes Ω



Classification systems theory

population set
of outcomes Ω



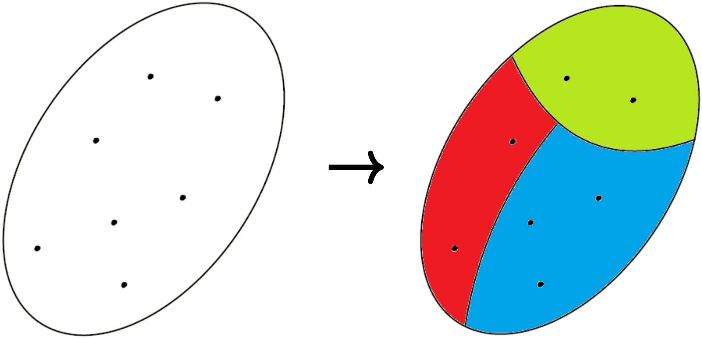
data set given
by some sensors $\quad \Omega \xrightarrow{s} \mathcal{D}$
$\mathcal{D} = \{x_1, \ldots, x_N\}$

Classification systems theory

population set of outcomes Ω

$$\Omega \xrightarrow{\boldsymbol{T}} \mathcal{L}$$

n-label set $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_n\}$



data set given by some sensors $\mathcal{D} = \{x_1, \ldots, x_N\}$

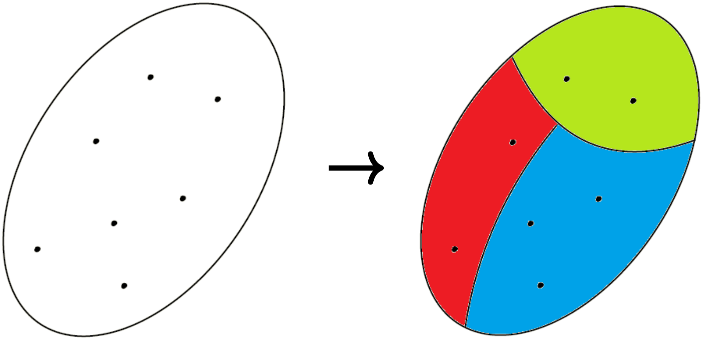
$$\Omega \xrightarrow{\boldsymbol{s}} \mathcal{D}$$

Classification systems theory

population set
of outcomes Ω

$$\Omega \xrightarrow{\boldsymbol{T}} \mathcal{L}$$

n-label set
$\mathcal{L} = \{\ell_1, \ell_2, \dots, \ell_n\}$



data set given
by some sensors
$\mathcal{D} = \{x_1, \dots, x_N\}$

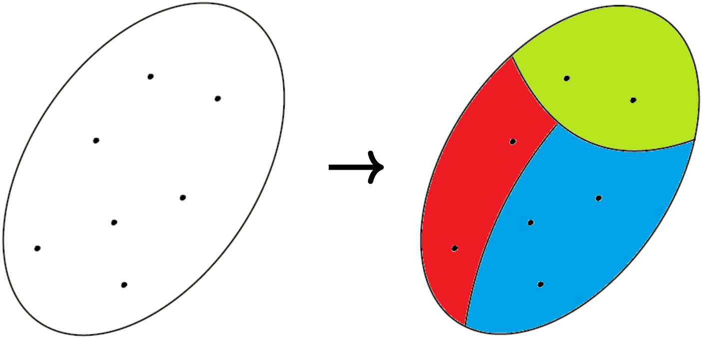
$$\Omega \xrightarrow{\boldsymbol{s}} \mathcal{D} \xrightarrow{\Psi} \mathcal{F}$$

Classification systems theory

population set of outcomes Ω

$$\Omega \xrightarrow{\boldsymbol{T}} \mathcal{L}$$

n-label set $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_n\}$



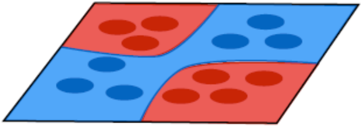
data set given by some sensors $\mathcal{D} = \{x_1, \ldots, x_N\}$

$$\Omega \xrightarrow{\boldsymbol{s}} \mathcal{D} \xrightarrow{\Psi} \mathcal{F} \xrightarrow{\boldsymbol{c_\theta}} \mathcal{L}$$

$$\boxed{A_{\boldsymbol{\theta}} = \boldsymbol{c_\theta} \circ \Psi \circ \boldsymbol{s}}$$

classification system

Linear/kernel supervised learning in least-squares SVM

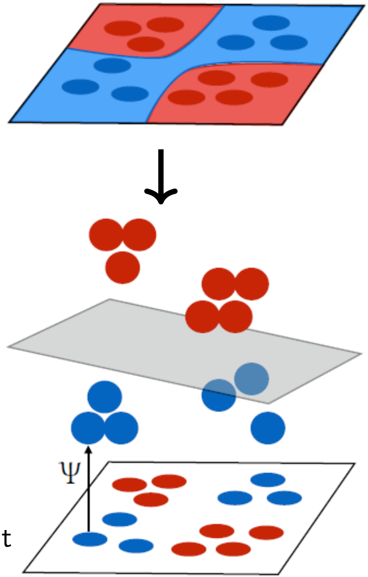


Labeled dataset: $$y_i = \begin{cases} +1 \text{ if } x_i \in A \\ -1 \text{ if } x_i \in B \end{cases}$$

Cost function: $$\mathcal{J}(\boldsymbol{\theta}) = \left|\left|\mathcal{M}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{y}\right|\right|^2$$

A linear decision function: $\mathcal{M}_{\boldsymbol{\theta}}(\boldsymbol{x}) = W_{\boldsymbol{\theta}} \cdot \boldsymbol{x}$ fails the classification in this case.

Linear/kernel supervised learning in least-squares SVM



Labeled dataset: $y_i = \begin{cases} +1 \text{ if } x_i \in A \\ -1 \text{ if } x_i \in B \end{cases}$

Cost function: $\mathcal{J}(\boldsymbol{\theta}) = \left|\left|\mathcal{M}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{y}\right|\right|^2$

A linear decision function: $\mathcal{M}_{\boldsymbol{\theta}}(\boldsymbol{x}) = W_{\boldsymbol{\theta}} \cdot \boldsymbol{x}$ fails the classification in this case.

↓

Non-linear feature map $\Psi(\boldsymbol{x})$, e.g.

$$\Psi(x_i) = \left(x_{i,1}, x_{i,2}, x_{i,1}^2 + x_{i,2}^2\right)$$

with $\mathcal{M}_{\boldsymbol{\theta}}(\Psi) = W_{\boldsymbol{\theta}} \cdot \Psi(\boldsymbol{x})$ able to classify, but involving high computational complexity!

Feature map

Let's label features with x_i^j for $i = 1, \ldots, N$ and for patient $j = 1, \ldots, M$.

Rescale: $x_i^j \longrightarrow \frac{x_i^j}{\max_j x_i^j} \xrightarrow{\psi} \begin{pmatrix} \sin\left(\frac{\pi}{2} x_i^j\right) \\ \cos\left(\frac{\pi}{2} x_i^j\right) \end{pmatrix} = \psi_i^j \in \mathbb{R}^2$ single variable feature map

Feature map

Let's label features with x_i^j for $i = 1, \ldots, N$ and for patient $j = 1, \ldots, M$.

Rescale: $x_i^j \longrightarrow \frac{x_i^j}{\max_j x_i^j} \xrightarrow{\psi} \begin{pmatrix} \sin\left(\frac{\pi}{2} x_i^j\right) \\ \cos\left(\frac{\pi}{2} x_i^j\right) \end{pmatrix} = \psi_i^j \in \mathbb{R}^2$ single variable feature map

For each patient the global feature vector is: $\Psi^j = \bigotimes_{j=1}^{N} \psi_i^j \in \mathbb{R}^{2^N}$

Example: $N = 4$

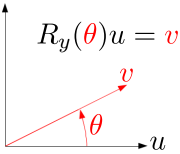
Feature map

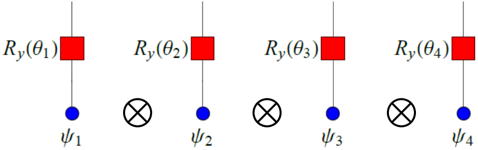
Let's label features with x_i^j for $i = 1, \ldots, N$ and for patient $j = 1, \ldots, M$.

Rescale: $x_i^j \longrightarrow \dfrac{x_i^j}{\max_j x_i^j} \xrightarrow{\psi} \begin{pmatrix} \sin\left(\frac{\pi}{2} x_i^j\right) \\ \cos\left(\frac{\pi}{2} x_i^j\right) \end{pmatrix} = \psi_i^j \in \mathbb{R}^2$ single variable feature map

For each patient the global feature vector is: $\Psi^j = \bigotimes_{j=1}^{N} \psi_i^j \in \mathbb{R}^{2^N}$

Example: $N = 4$





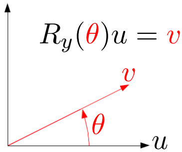
Feature map

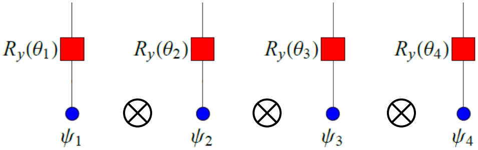
Let's label features with x_i^j for $i = 1, \ldots, N$ and for patient $j = 1, \ldots, M$.

Rescale: $x_i^j \longrightarrow \dfrac{x_i^j}{\max_j x_i^j} \xrightarrow{\psi} \begin{pmatrix} \sin\left(\frac{\pi}{2} x_i^j\right) \\ \cos\left(\frac{\pi}{2} x_i^j\right) \end{pmatrix} = \psi_i^j \in \mathbb{R}^2$ single variable feature map

For each patient the global feature vector is: $\Psi^j = \bigotimes_{j=1}^{N} \psi_i^j \in \mathbb{R}^{2^N}$

Example: $N = 4$





$$R_y(\theta) = \mathrm{e}^{-\mathrm{i}\sigma_y \theta} \in \mathrm{SO}(2)$$

$$R_y(\theta)\psi_i = (\mathbb{1}\cos(\theta) - \mathrm{i}\sigma_y \sin(\theta))\,\psi_i = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \sin(\frac{\pi}{2} x_i) \\ \cos(\frac{\pi}{2} x_i) \end{pmatrix}$$

A crash course on spins (aka qubits!)

$$\Psi = a\psi_\uparrow + b\psi_\downarrow, \quad \psi_\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \psi_\downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad a, b \in \mathbb{C}$$

We can act on such a spin state by means of Pauli and identity matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

A crash course on spins (aka qubits!)

$$\Psi = a\psi_\uparrow + b\psi_\downarrow, \quad \psi_\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \psi_\downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad a, b \in \mathbb{C}$$

We can act on such a spin state by means of Pauli and identity matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example: spin flip — up projection — down projection

$$\sigma_x \psi_\uparrow = \psi_\downarrow, \qquad P_\uparrow \Psi = a\psi_\uparrow, \qquad P_\downarrow \Psi = b\psi_\downarrow$$

A crash course on spins (aka qubits!)

$$\Psi = a\psi_\uparrow + b\psi_\downarrow, \quad \psi_\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \psi_\downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad a, b \in \mathbb{C}$$

We can act on such a spin state by means of Pauli and identity matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example: spin flip | up projection | down projection

$$\sigma_x \psi_\uparrow = \psi_\downarrow, \qquad P_\uparrow \Psi = a\psi_\uparrow, \qquad P_\downarrow \Psi = b\psi_\downarrow$$

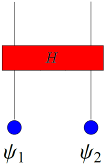
Spin interactions

Example: $H = J\sigma_z \otimes \sigma_z$

$$H\psi_\uparrow \otimes \psi_\uparrow = J\psi_\uparrow \otimes \psi_\uparrow$$

$$H\psi_\downarrow \otimes \psi_\downarrow = J\psi_\downarrow \otimes \psi_\downarrow$$

$$H\psi_\uparrow \otimes \psi_\downarrow = -J\psi_\uparrow \otimes \psi_\downarrow$$



A crash course on spins (aka qubits!)

$$\Psi = a\psi_\uparrow + b\psi_\downarrow, \quad \psi_\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \psi_\downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad a, b \in \mathbb{C}$$

We can act on such a spin state by means of Pauli and identity matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example: spin flip — up projection — down projection

$$\sigma_x \psi_\uparrow = \psi_\downarrow, \qquad P_\uparrow \Psi = a\psi_\uparrow, \qquad P_\downarrow \Psi = b\psi_\downarrow$$

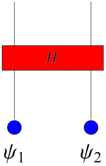
Spin interactions

Example: $H = J\sigma_z \otimes \sigma_z$

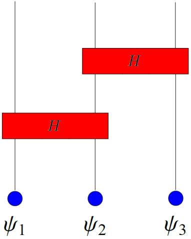
$$H\psi_\uparrow \otimes \psi_\uparrow = J\psi_\uparrow \otimes \psi_\uparrow$$

$$H\psi_\downarrow \otimes \psi_\downarrow = J\psi_\downarrow \otimes \psi_\downarrow$$

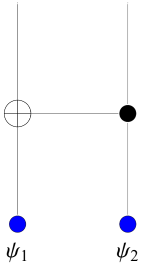
$$H\psi_\uparrow \otimes \psi_\downarrow = -J\psi_\uparrow \otimes \psi_\downarrow$$



Multiple spins may interact subsequently



Controlled NOT gates let data interact!



Classically XOR: $\mathrm{CNOT} = \sigma_x \otimes P_\uparrow + \mathbb{1} \otimes P_\downarrow$

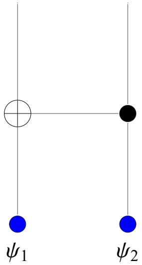
Example:

$$\mathrm{CNOT}\psi_\downarrow \otimes \psi_\downarrow = \psi_\downarrow \otimes \psi_\downarrow, \quad \mathrm{CNOT}\psi_\downarrow \otimes \psi_\uparrow = \psi_\uparrow \otimes \psi_\uparrow$$

In our case: $\mathrm{CNOT}\ \psi_1 \otimes \psi_2$

the interaction of ψ_1 depends on the control qubit ψ_2.

Controlled NOT gates let data interact!



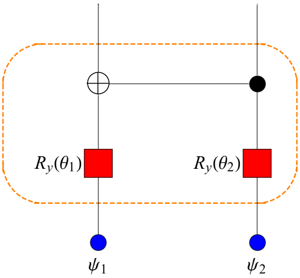
Classically XOR: $\mathrm{CNOT} = \sigma_x \otimes P_\uparrow + \mathbb{1} \otimes P_\downarrow$

Example:

$$\mathrm{CNOT}\psi_\downarrow \otimes \psi_\downarrow = \psi_\downarrow \otimes \psi_\downarrow, \quad \mathrm{CNOT}\psi_\downarrow \otimes \psi_\uparrow = \psi_\uparrow \otimes \psi_\uparrow$$

In our case: $\mathrm{CNOT}\ \psi_1 \otimes \psi_2$

the interaction of ψ_1 depends on the control qubit ψ_2.



$$H = \mathrm{CNOT}\ R_y(\theta_1) \otimes R_y(\theta_2)$$

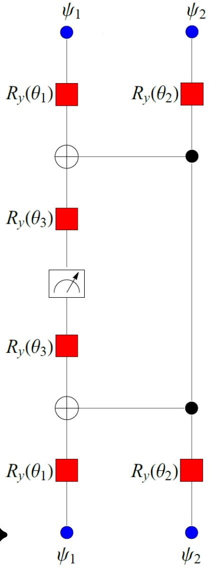
Example: binary features

$$x_1 = 1 \longrightarrow \psi_\uparrow$$

$$x_2 = 0 \longrightarrow \psi_\downarrow$$

$$H\psi_\uparrow \otimes \psi_\downarrow = \mathrm{CNOT} \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} \otimes \begin{pmatrix} -\sin(\theta_2) \\ \cos(\theta_2) \end{pmatrix}$$

Least-squares SVM via measurements

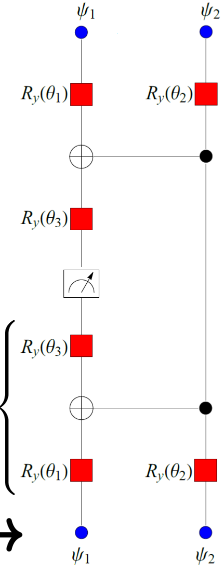


$\Psi = \psi_1 \otimes \psi_2$ feature vector $\longrightarrow$

Least-squares SVM via measurements

$$R_y(\theta_3) \otimes \mathbb{1} \ \mathrm{CNOT} \ R_y(\theta_1) \otimes R_y(\theta_2) = \mathcal{U}_{\boldsymbol{\theta}}$$

$\Psi = \psi_1 \otimes \psi_2$ feature vector →

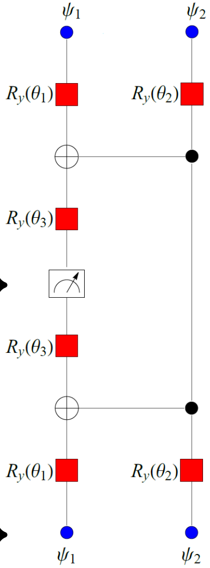


Least-squares SVM via measurements

$$P_{\uparrow} \otimes \mathbb{1} = \mathcal{M}$$ measurement →

$$R_y(\theta_3) \otimes \mathbb{1} \text{ CNOT } R_y(\theta_1) \otimes R_y(\theta_2) = \mathcal{U}_{\boldsymbol{\theta}}$$

$$\Psi = \psi_1 \otimes \psi_2$$ feature vector →



Least-squares SVM via measurements

Index $j = 1, \ldots, M$ for the patients sample:

score $\qquad \mathcal{M}_{\boldsymbol{\theta}}(\Psi^j) = \langle \mathcal{U}_{\boldsymbol{\theta}} \Psi^j, \mathcal{M}\, \mathcal{U}_{\boldsymbol{\theta}} \Psi^j \rangle$

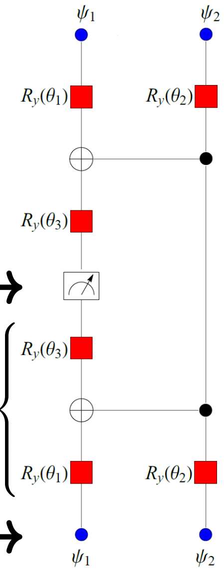
and minimize the mean squared error cost function:

$$\mathcal{J}(\boldsymbol{\theta}) = \frac{1}{M} \sum_{j=1}^{M} (\mathcal{M}_{\boldsymbol{\theta}}(\Psi^j) - y^j)^2$$

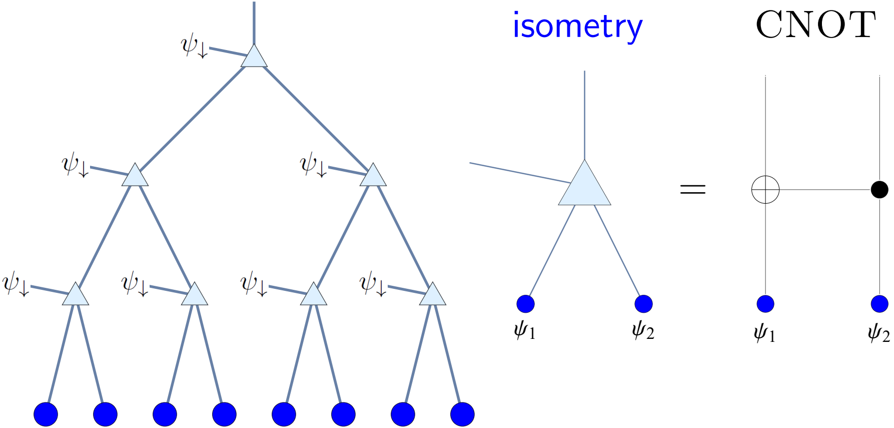
$P_\uparrow \otimes \mathbb{1} = \mathcal{M}$ measurement →

$R_y(\theta_3) \otimes \mathbb{1}\ \mathrm{CNOT}\ R_y(\theta_1) \otimes R_y(\theta_2) = \mathcal{U}_{\boldsymbol{\theta}}$

$\Psi = \psi_1 \otimes \psi_2$ feature vector →



Tree tensor network encoding

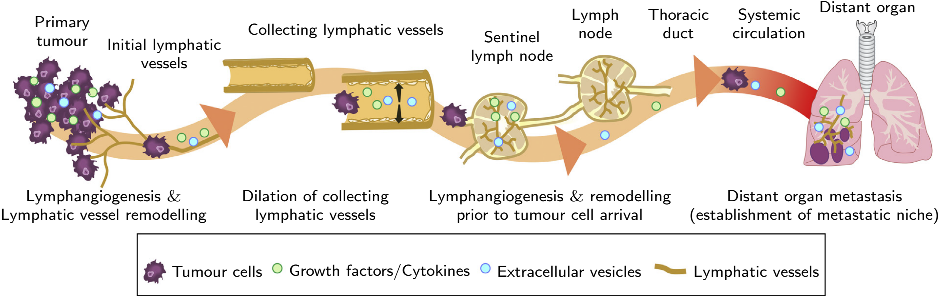


Ising model ground state:

$$H = -\sum_j \sigma_z^{(j)} \sigma_z^{(j+1)}$$

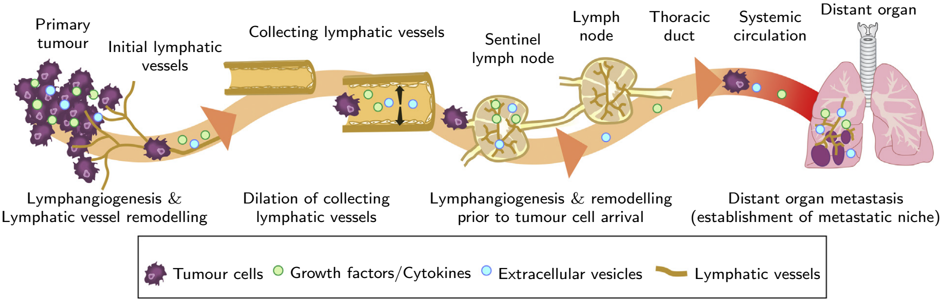
$$a\ \psi_\uparrow + b\ \psi_\downarrow \xrightarrow{\text{encoding}} a\ \psi_\uparrow \otimes \psi_\uparrow \cdots \otimes \psi_\uparrow + b\ \psi_\downarrow \otimes \psi_\downarrow \cdots \otimes \psi_\downarrow$$

Our case study: breast cancer lymph node metastasis



Farnsworth, R.H.; Achen, M.G.; Stacker, S.A. The evolving role of lymphatics in cancer metastasis. Curr. Opin. Immunol. **2018**, 53, 64–73.

Our case study: breast cancer lymph node metastasis



Our dataset:
634 patients with 214 lymph nodes metastasis

Purpose of the study: detection of lymph nodes metastasis in a pre-operative stage

Patients' features:

- age

Tumour features:

- diameter
- grading
- in situ component
- multiplicity
- histologic type
- ER
- PgR
- Ki67
- HER2/neu

Farnsworth, R.H.; Achen, M.G.; Stacker, S.A. The evolving role of lymphatics in cancer metastasis. Curr. Opin. Immunol. **2018**, 53, 64–73.

Performance evaluation

Given the number of negative n and positive p patients, with classifier results involving true positive tp and true negative tn, we define

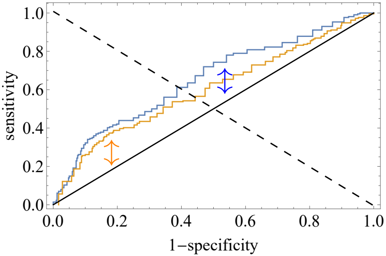
$$\text{accuracy} = \frac{tn + tp}{n + p}, \qquad \text{specificity} = \frac{tn}{n}, \qquad \text{sensitivity} = \frac{tp}{p}.$$

Performance evaluation

Given the number of negative n and positive p patients, with classifier results involving true positive tp and true negative tn, we define

$$\text{accuracy} = \frac{tn + tp}{n + p}, \qquad \text{specificity} = \frac{tn}{n}, \qquad \text{sensitivity} = \frac{tp}{p}.$$

The score threshold characterizes a single classifier and its variation in the interval $[0, 1]$ defines a family of classifiers, whose performances are summarized by receiver operating characteristic (ROC) curves, using the area under the ROC curve (AUC).



Classifier selection according to Youden test through the maximization of

$$J = \text{sensitivity} - \text{specificity} + 1$$

A hard decision making process

Our best classical (not quantum!) classifier with diameter, grading, histologic type, multifocality, in situ component, PgR:

AUC (%)	**Accuracy** (%)	**Specificity** (%)	**Sensitivity** (%)
70.8 (70.3-71.1)	69.8 (69.3-70.2)	74.8 (72.9-75.1)	61.0 (60.3-61.7)

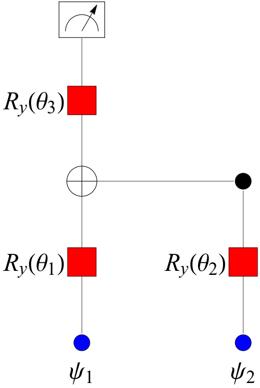
reporting the 1^{st}-3^{rd} interquartile range after 10 ten-fold cross-validations.

A hard decision making process

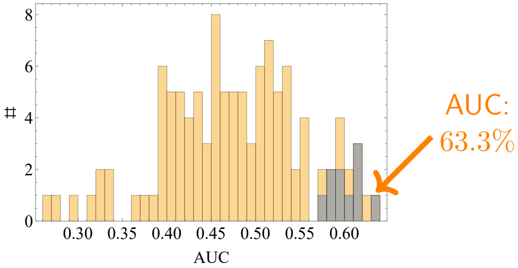
Our best classical (not quantum!) classifier with diameter, grading, histologic type, multifocality, in situ component, PgR:

AUC (%)	**Accuracy** (%)	**Specificity** (%)	**Sensitivity** (%)
70.8 (70.3-71.1)	69.8 (69.3-70.2)	74.8 (72.9-75.1)	61.0 (60.3-61.7)

reporting the 1^{st}-3^{rd} interquartile range after 10 ten-fold cross-validations.



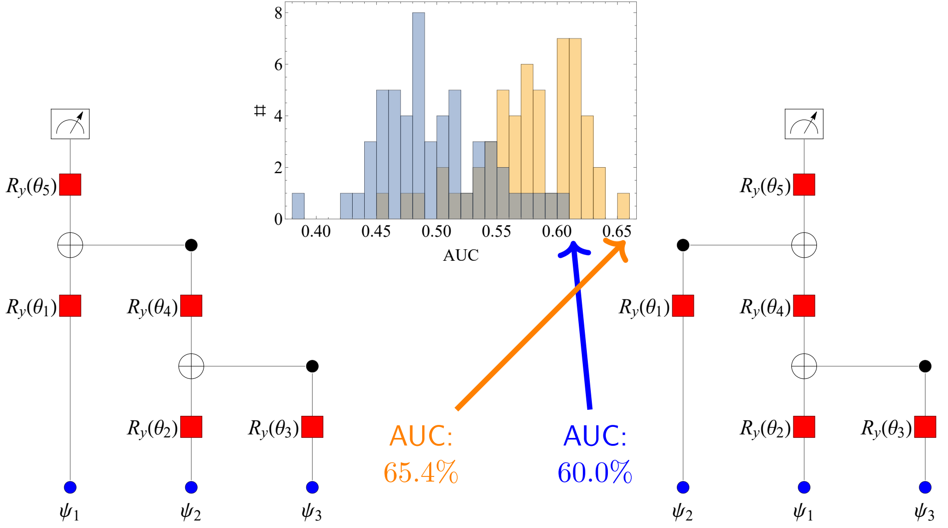
A feature selection is implemented for any variables pair in our quantum classifier:



diameter feature is mapped in ψ_1 for shaded bars.

Features selection

Maintaining the best features pair (diameter, PgR), we compare two not equivalent scheme for three feature, yielding grading and Ki67:



Summary & outlook

	AUC (%)	**Accuracy** (%)
Classical (6 features)	70.8 (70.3-71.1)	69.8 (69.3-70.2)
Classical (3 features)	67.4 (67.4-67.5)	61.7 (60.6-62.8)
Quantum (2 features)	63.1 (62.7-63.3)	65.5 (65.1-65.8)
Quantum (3 features)	64.7 (64.1-65.1)	69.5 (61.8-70.2)
Quantum (3 features)	59.5 (59.1-59.8)	65.3 (64.7-65.9)

- Quantum circuits offer a promising model for complex systems;
- classification performances of our tutorial model still miss a possible clinical application;
- further improvements are related with computational complexity reduction to manage quantum states in higher dimensions.

Thank you!

Thank you!

Questions/Comments?