Supervised Learning with Quantum Measurements

Fabio A. González

QCP 2020-2

Universidad Nacional de Colombia

1. Supervised Machine Learning

Given $T = \{(x_i, y_i)\}_{i=1..n}$ (training data set) find $f: X \longrightarrow Y$
such that $\forall i \; f(x_i) \approx y_i$

$x_i \in X$ input

$y_i \in Y$ output

$$f(x) = g(w, x)$$

$w \longrightarrow$ parameters

Probabilistic approach Model $P(y \mid x)$ (Conditional prob. distribution)

$$f(x) = \arg\max_{k=1..m} P(y = y_k \mid x) \qquad Y = \{y_1, \dots, y_m\}$$

1. model the joint $P(y, x)$, from this get $P(y \mid x)$
2. $P(y \mid x) = \dfrac{P(x \mid y)\, P(y)}{P(x)}$
3. $P(y \mid x) = g(w \cdot x + w_0)$

$w, w_0 \longrightarrow$ Parameters

kernel approach

kernel function

$k: X \times X \longrightarrow \mathbb{R}$ $\qquad k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle_F$

implicitely induces a feature map $\Phi: X \longrightarrow F$

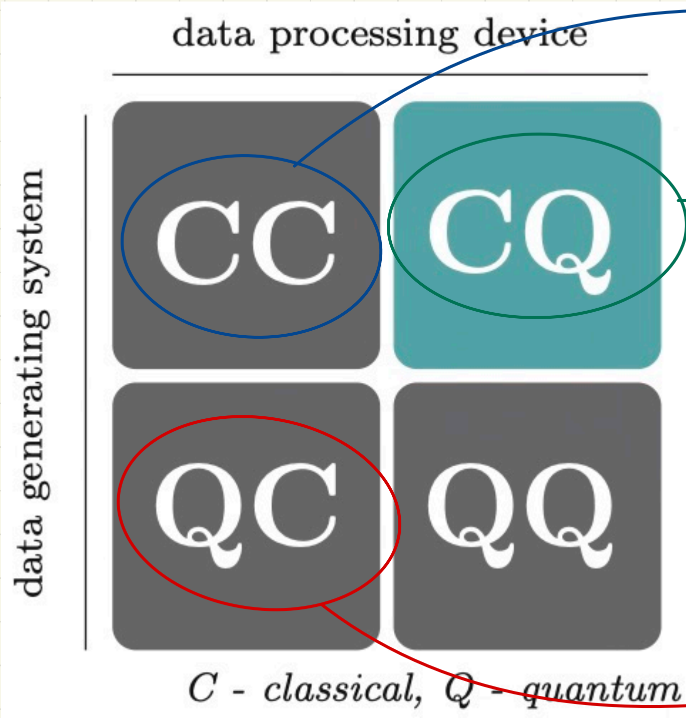
$$f(x) = \sum_{(x_i, y_i) \in S} \alpha_i \, k(x, x_i) \, y_i$$

$\alpha_i \longrightarrow$ Parameters

Learning

- Estimating probabilities
- Finding parameters using optimization.

2. Quantum Machine Learning



Classical ML algorithms with quantum inspiration

Quantum implementation of ML algorithms to deal with classical data.

Using ML algorithms to support quantum computing or research

3. Quantum measurement Classiffication (Training)

$$\Psi_x : X \longrightarrow H_x \qquad x_i \longmapsto |\Psi_x(x_i)\rangle$$

$$\Psi_y : Y \longrightarrow H_y \qquad y_i \longmapsto |\Psi_y(y_i)\rangle$$

Both inputs and outputs are represented as quantum states

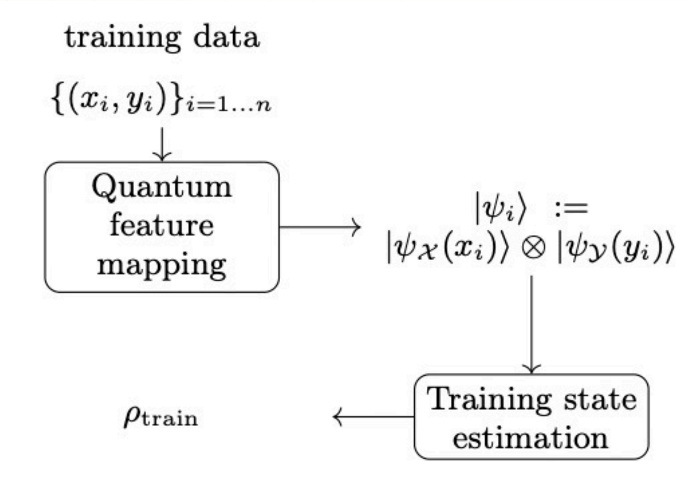


Figure 1. Training

1. Quantum feature mapping

$$\Psi : X \times Y \longrightarrow H_x \otimes H_y$$

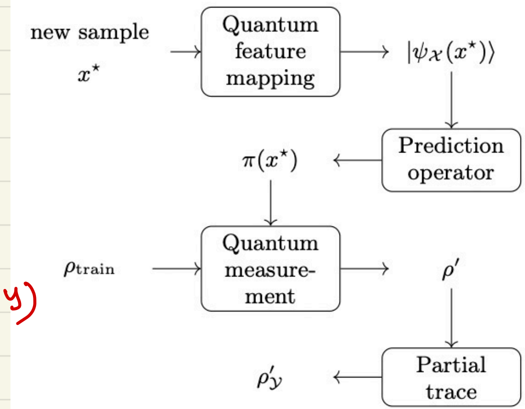
$$(x_i, y_i) \longmapsto \underbrace{|\Psi_x(x_i)\rangle \otimes |\Psi_z(z_i)\rangle}_{\Psi_i}$$

2. Training state estimation

$$\rho_{train} = \frac{1}{n} \sum_{i=1}^{n} |\Psi_i\rangle \langle \Psi_i|$$

This represents $P(x, y)$

4. QMC (Prediction)



$P(x,y)$

$P(y|x=x^*)$

Figure 2. Prediction

1. **Quantum feature mapping**

$$x^* \longmapsto |\psi_x(x^*)\rangle$$

2. **Prediction operator**

$$\pi(x^*) = |\psi_x(x^*)\rangle\langle\psi_x(x^*)| \otimes Id_{H_y}$$

3. **Quantum measurement**

$$\rho' = \frac{\pi(x^*)\rho_{train}\pi(x^*)}{tr[\pi(x^*)\rho_{train}\pi(x^*)]}$$

4. **Partial trace**

$$\rho'_y = Tr_X[\rho']$$

5. Relationship with Bayesian Learning

Proposition 1. *Let* $T = \{(x_i, y_i)\}_{i=1,\ldots,n}$ *be a set of training samples,* $x^\star$ *a sample to classify, with* $x_i, x^\star \in \{1, \ldots m\}$ *and* $y_i \in \{1, 2\}$*. Let* ρ_{train} *be the state calculated using the mixed state, eq. (8) or equivalently the classic mixture eq. (9), and a one-hot encoding feature map for both* x_i *and* y_i*. Then the diagonal elements of the density matrix* $\rho'_{\mathcal{Y}}$ *calculated using eq. (12) correspond to an estimation, using Bayesian inference, of the conditional probabilities* $P(y = i|x^\star)$*:*

$$\rho'_{\mathcal{Y}i,i} = P(y = i|x^\star) = \frac{P(x^\star|y = i)P(y = i)}{P(x^\star)}, \qquad (13)$$

where $P(x^\star|y = i)$*,* $P(y = i)$ *and* $P(x^\star)$ *are estimated from* T*.*

6. Relationship with Kernel learning

Proposition 2. *Let $T = \{(x_i, y_i)\}$ be a set of training samples, $x^\star$ a sample to classify, with $x_i, x^\star \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. Let ρ_{train} be the state calculated using a mixed state (eq. (8)) and quantum feature maps $\psi_\mathcal{X}$ and $\psi_\mathcal{Y}$. Then the density matrix $\rho'_\mathcal{Y}$, calculated with eq. (12), can be expressed as:*

$$\rho'_\mathcal{Y} = \mathcal{M} \sum_{i=1}^{N} |k(x^\star, x_i)|^2 \, |\psi_\mathcal{Y}(y_i)\rangle \langle \psi_\mathcal{Y}(y_i)| \,, \tag{14}$$

where $k(x^\star, x_i) = \langle \psi_\mathcal{X}(x^\star) | \psi_\mathcal{X}(x_i) \rangle$ and $\mathcal{M}^{-1} = \mathrm{Tr}[\pi(x^\star) \rho_{train} \pi(x^\star)]$.