

IMPLEMENTAZIONE DI UN ALGORITMO KNN MULTICLASSE SU HARDWARE QUANTISTICO

TESI DI LAUREA IN FISICA

MARIANO MOLLO N85000880

RELATORI:

GIOVANNI ACAMPORA

AUTILIA VITIELLO

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

15 OTTOBRE 2019



1. Introduzione
2. Machine learning
3. Quantum computing
4. Metodi
5. Quantum machine learning
6. Risultati
7. Conclusione

INTRODUZIONE



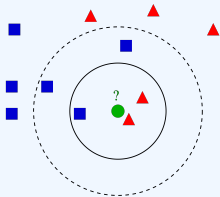
- Generalizzazione di un algoritmo KNN quantistico per la classificazione multiclasse
- Simulare l'usabilità degli algoritmi di machine learning quantistico in contesti reali
- Applicazione al data set Iris (3 classi, 4 caratteristiche, 150 vettori)

Il machine learning è una branca dell'IA che dà l'abilità ai computer di apprendere dai dati

Gli algoritmi di machine learning permettono di apprendere modelli matematici da insiemi di dati per

- effettuare previsioni su dati sconosciuti
- individuare regolarità nascoste
- compiere decisioni che danno il risultato migliore

ALGORITMO K-NEAREST NEIGHBOURS



- k è un numero naturale
- Dato un dataset $D = v_0, \dots, v_n$
 $v_i \in \{\text{setosa}, \text{versicolor}, \text{virginica}\}$
- Dato un nuovo vettore x
- Si considerano i k vettori più vicini ad x
- Si classifica x con un voto a maggioranza
- Si assegnano pesi dipendenti dall'inverso della distanza per aumentare l'influenza di quelli più vicini

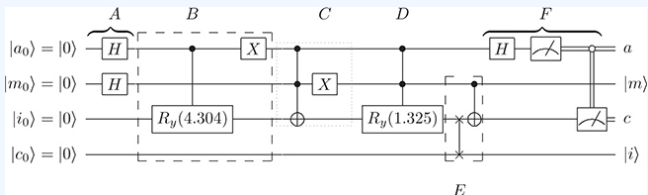
Usare strumenti di calcolo che sfruttano le proprietà quantistiche della materia

- Il qubit è la controparte quantistica del bit. Un bit può assumere un singolo valore tra 0 e 1. Un qubit può esistere in una sovrapposizione dei due stati $|0\rangle$ e $|1\rangle$.
- Un qubit vive in uno spazio di Hilbert bidimensionale.
- Un sistema di n qubit è descritto da uno spazio di dimensione 2^n
- Il quantum computing permette la parallelizzazione del lavoro e la memorizzazione di grandi quantità di dati.

Un computer quantistico con n qubit possiede 2^n ampiezze di probabilità

Queste ampiezze possono essere usate per immagazzinare quantità enormi di informazioni

Numero di qubit	RAM classica richiesta
5	256 byte
25	2 gigabyte
50	8000 terabyte
275	numero di atomi nell'universo osservabile



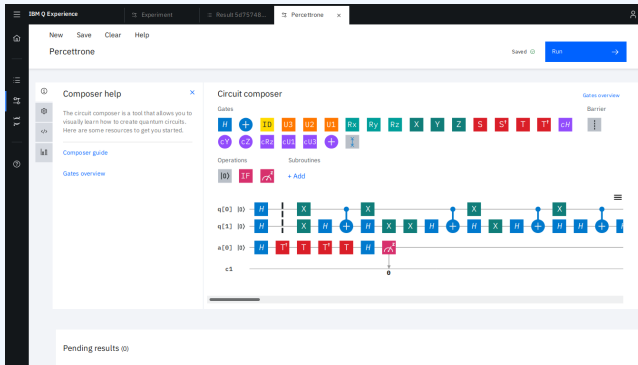
In un articolo di ricerca, Schuld et al. propongono un algoritmo KNN quantistico che permette di ridurre esponenzialmente le risorse necessarie per il riconoscimento.

L'algoritmo riesce a comparare due classi alla volta, usa come k il numero totale di vettori e usa una metrica che va come $1/\text{distanza}$. L'implementazione viene fatta su un computer quantistico a 5 qubit.

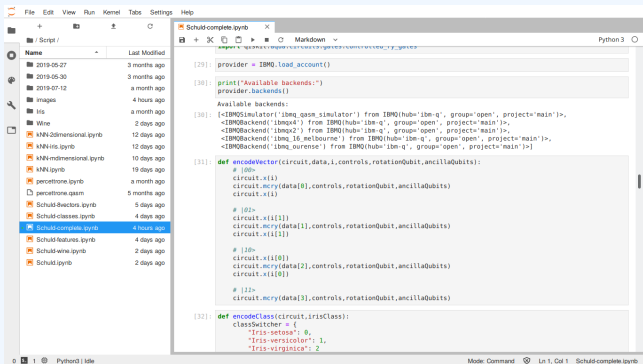
- A partire dall'algoritmo KNN quantistico proposto da Schuld et al., è possibile implementarne una versione multiclasse?
- Quali sono le performance di classificazione all'aumentare del numero di qubit?

METODI

IBM Q EXPERIENCE



- Accessibile al pubblico
- Permette simulazioni con o senza rumore
- Fino a 14 qubit superconduttivi
- Fino a 32 qubit simulati



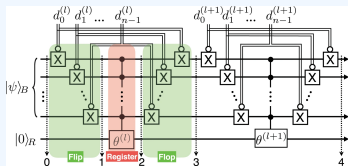
Piccoli circuiti quantistici possono essere simulati su computer convenzionali

Qiskit permette di

- ideare circuiti quantistici
- simularne l'esecuzione sul proprio pc
- interfacciarsi con l'IBM QX

QUANTUM MACHINE LEARNING

CODIFICARE I DATI NELLE AMPIEZZE



Si è costruita una QRAM (Quantum RAM) attraverso l'algoritmo di costruzione di Petruccione et al. per codificare vettori classici nei vari stati quantistici a disposizione

$$\text{QRAM } |o\rangle = |\psi\rangle$$

Stato quantistico iniziale

$$|\psi_0\rangle = \frac{1}{\sqrt{2M}} \sum_{m=1}^M (|0\rangle |\psi_x\rangle + |1\rangle |\psi_{t^m}\rangle) |c^m\rangle |m\rangle$$

Calcolo della distanza con interferenza quantistica

$$|\psi_1\rangle = \frac{1}{2\sqrt{M}} \sum_{m=1}^M (|0\rangle (|\psi_x\rangle + |\psi_{t^m}\rangle) + |1\rangle (|\psi_x\rangle - |\psi_{t^m}\rangle)) |c^m\rangle |m\rangle$$

Misura condizionale

$$|\psi_2\rangle = \frac{1}{2\sqrt{M}} \sum_{m=1}^M \sum_{i=1}^N (x_i + t_i^m) |0\rangle |i\rangle |c^m\rangle |m\rangle$$

Probabilità di misurare una data classe

$$P(|c^m\rangle = |s\rangle) = \sum_{m|c^m=s} 1 - \frac{1}{4M} |x - t^m|^2$$

Classificazione

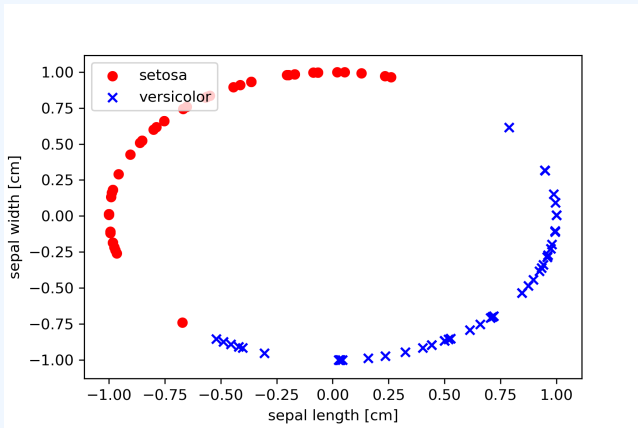
$$c = \begin{cases} 0 & \text{se } P(|c^0\rangle) \text{ maggiore} \\ 1 & \text{se } P(|c^1\rangle) \text{ maggiore} \\ \text{etc} \dots \end{cases}$$

RISULTATI

DATA SET IRIS

L'insieme dati usato per far girare l'algoritmo è stato il ben noto Iris Data Set

Si è prima standardizzato e normalizzato i vettori di training e poi si è proceduto alla classificazione



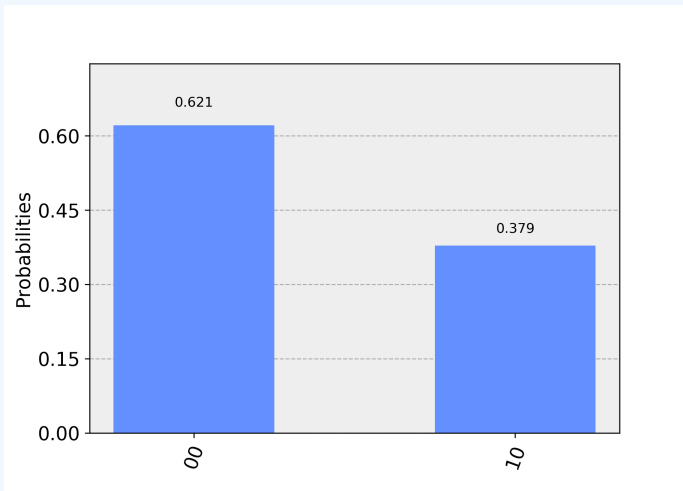


Figure: Misura della classe setosa

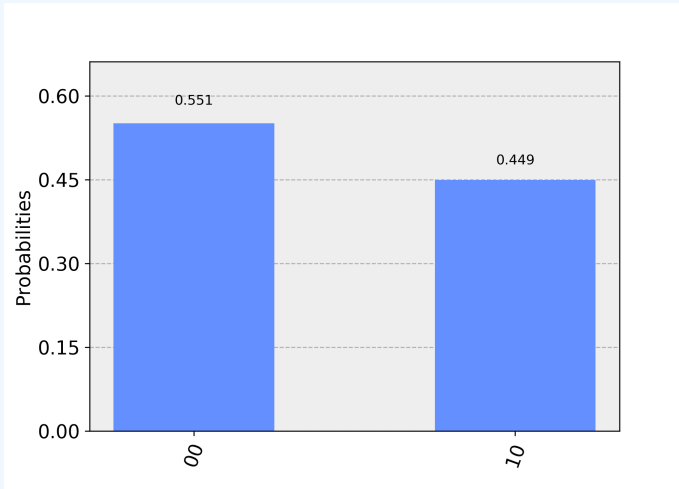


Figure: Misura della classe setosa

CLASSIFICAZIONE MULTICLASSE

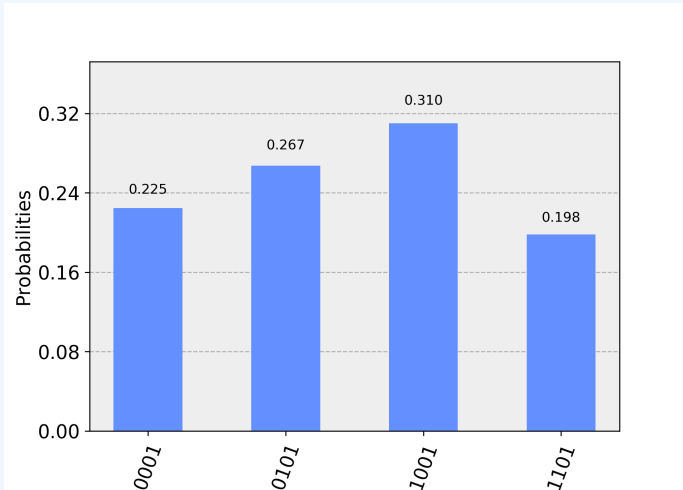


Figure: Misura multiclasse virginica

L'implementazione rispecchia i risultati aspettati di classificazione e permette di ridurre esponenzialmente le risorse necessarie in termini di memoria usata e di numero di elaborazioni.

L'algoritmo KNN non è l'unico che può beneficiare di un'approccio quantistico e si aspettano ancora molte evoluzioni nel campo del quantum machine learning.

DOMANDE?