Supervised Learning with Quantum Enhanced Feature Spaces

Alejandro Rosales

Research Paper Reference

Title: Supervised Learning with Quantum Enhanced Feature Spaces

Authors: Vojtech Havlicek, Antonio D. Corcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta.

Research Centers: IBM T.J. Watson Research Center and Center for Theoretical Physics, Massachusetts Institute of Technology

Dated: June 7, 2018

Link: https://arxiv.org/pdf/1804.11326.pdf

Motivation/Goal

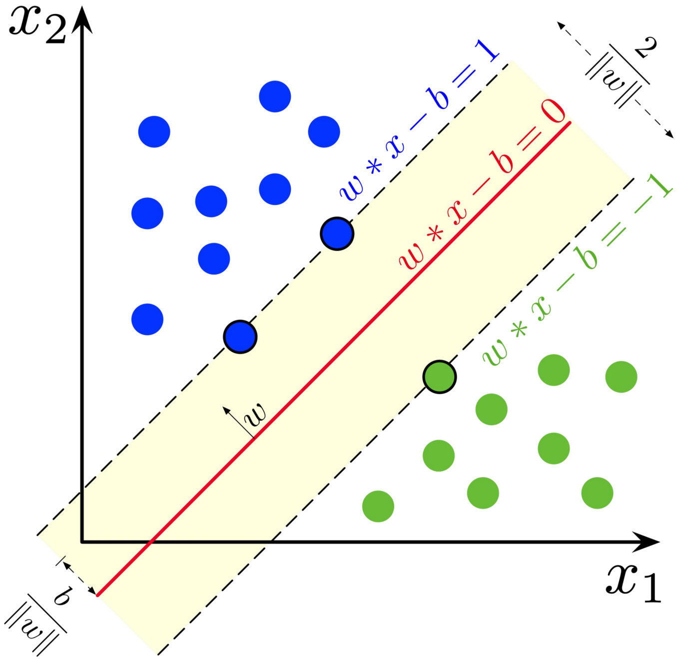
Current problem: limitations on successful solution for problems when feature space becomes large high-dimensional

Suggested solution: utilize controlled entanglement and interference to exploit exponentially growing quantum state space

Goal: present new class of tools for exploring the applications of noisy intermediate scale quantum computers to machine learning for improved computational power and efficacy

Preliminaries

Classical Linear Support Vector Machine Visual



Classical Linear Support Vector Machine Definitions

Formal Definitions:

- binary classification: $C = \{+1, -1\}$
- data: $\vec{x}_i \in T \subset \mathbb{R}^d$
- true data label: $y_i \in \{+1, -1\}$
- true map: $m : T \cup S \rightarrow \{+1, -1\}$
- approximation map: $\tilde{m} : S \rightarrow \{+1, -1\}$
- linear separability using hyperplane $(\mathbf{w}, b),\ \mathbf{w} \in \mathbb{R}^d$

Classical Linear SVM and High Dimensional Classification

Higher dimension feature space improves fidelity of classifications but increases computational resources

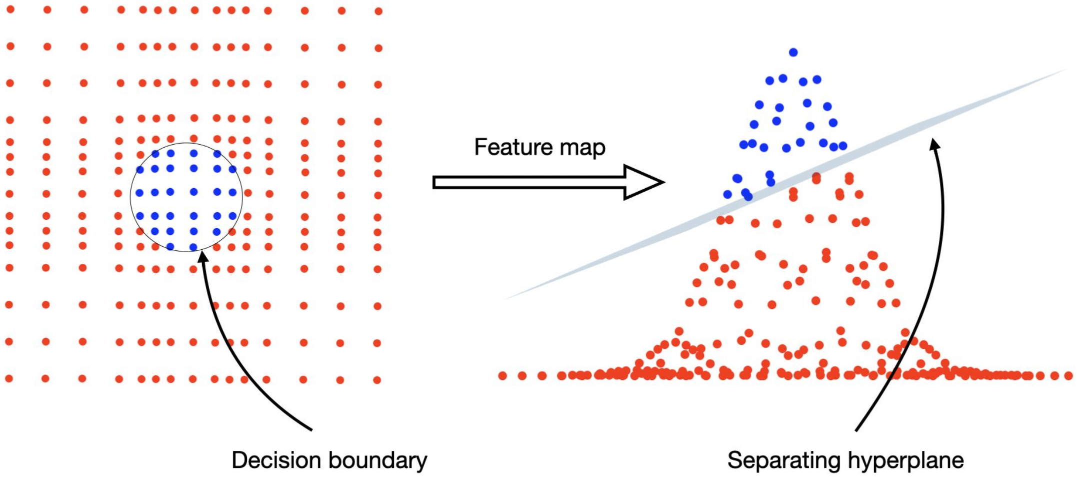


Image source: https://www.hashpi.com/the-intuition-behind-kernel-methods

Quantum Variational Classification

Data Classification

training set T and a test set S, where T, S are a subset $\Omega \subset \mathbb{R}^d$

true map: $m : T \cup S \to \{+1, -1\}$

approximated map: $\tilde{m} : S \to \{+1, -1\}$

Goal for optimized data classification:

$$m(\vec{s}) = \tilde{m}(\vec{s}) \text{ with high probability for some given test data } \vec{s} \in S$$

Test Data

data is generated of dimension $n = d = 2$ for a 2-qubit system

data vector labels: $\vec{x} \in T \cup S \subset (0, 2\pi]^2$

$\mathbf{f}V$

parity function $\mathbf{f} = Z_1 Z_2$

Note: delta values for supports

Given $\Delta = 0.3$, if:

$$\langle \Phi(\vec{x}) | V^\dagger \mathbf{f} V | \Phi(\vec{x}) \rangle \geq \Delta$$

then $m(\vec{x}) = +1$. If:

$$\langle \Phi(\vec{x}) | V^\dagger \mathbf{f} V | \Phi(\vec{x}) \rangle \leq -\Delta$$

then $m(\vec{x}) = -1$.

Overview: Quantum Variational Classification

Training Protocol
While Classifier Not Converged:
- Quantum Feature Mapping on Training Data
- Variational Optimization
- Data Measurement
- Calculating Error
- Optimize Classifier

Testing Protocol
Test Optimized Classifier

How We Will Proceed

Training Protocol

For Each Step in Protocol:

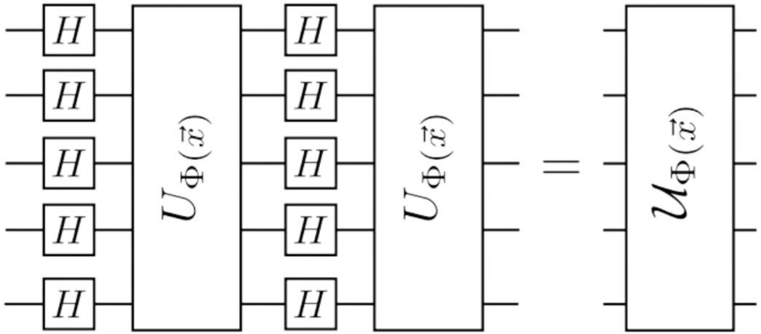
- Define Unitary Function
- Circuit Diagram
- Implementation

Quantum Feature Mapping

Unitary

$$\mathcal{U}_{\Phi(\vec{x})} = U_{\Phi(\vec{x})} H^{\otimes n} U_{\Phi(\vec{x})} H^{\otimes n}$$

Circuit Diagram



Quantum Feature Mapping Continued

Implementation

data $\vec{x} \in \Omega$ is mapped to a reference state $|0\rangle^n$ using the feature map circut $\mathcal{U}_{\Phi(\vec{x})}$

individual qubit:

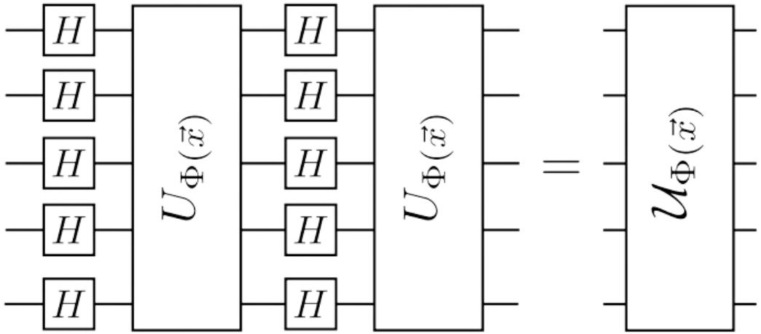
$$\vec{x} \mapsto |\,\phi_i(\vec{x})\rangle = U(\varphi_i(\vec{x}))|\,0\rangle$$

where: $\varphi : \vec{x} \to (0, 2\pi]^2 \times [0, \pi]$

full qubit state:

$$\Phi : \vec{x} \mapsto |\,\Phi(\vec{x})\rangle\langle\Phi(\vec{x})\,| = \bigotimes_{i=1}^{n} |\,\phi_i(x)\rangle\langle\phi_i(x)\,|$$

Circuit Diagram



Quantum Variational Classification

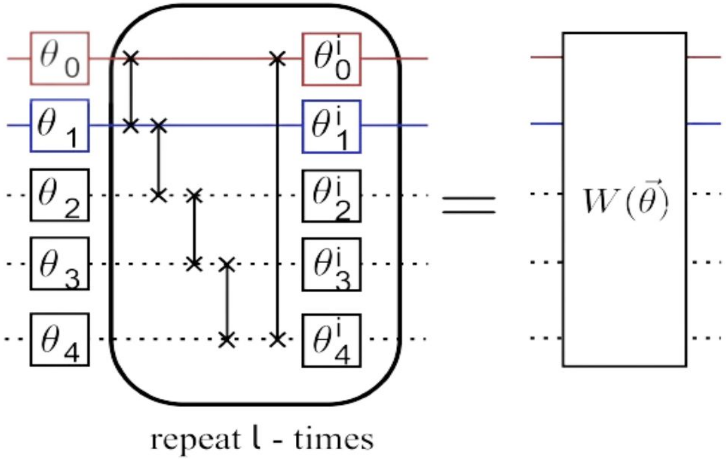
Implementation

$W(\vec{\theta})$

defined as:

$$U_{loc}^{(l)}(\theta_l)U_{ent}\dots U_{loc}^{(2)}(\theta_2)U_{ent}U_{loc}^{(1)}(\theta_1)$$

Circuit Diagram: Short Depth Quantum Circuit



variational circut is parametrized by $\vec{\theta} \in \mathbb{R}^{2n(l+1)}$ and it is what is optimized during training

Quantum Variational Classification

Implementation (Continued)

$$W(\vec{\theta}) = U_{loc}^{(l)}(\theta_l)U_{ent}\ldots U_{loc}^{(2)}(\theta_2)U_{ent}U_{loc}^{(1)}(\theta_1)$$

Where, $U_{loc}^{(l)}(\theta_l)$ is full layers single qubit rotations defined as:

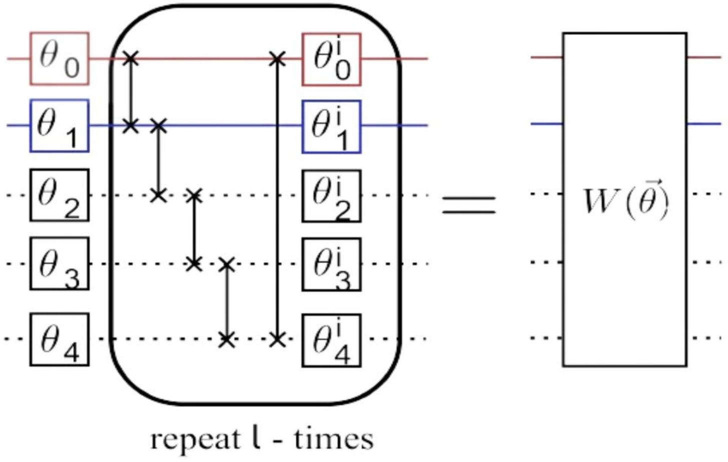
$$U_{loc}^{(t)}(\theta_t) = \bigotimes_{i=1}^{n} U(\theta_{i,t})$$

and $U(\theta_{i,t}) \in SU(2)$.

For the variational unitary, U_{ent} is an alternating layers of entangling gates and is defined as:

$$U_{ent} = \prod_{(i,j)\in E} \mathbf{CZ}(i,j)$$

Circuit Diagram: Short Depth Quantum Circuit



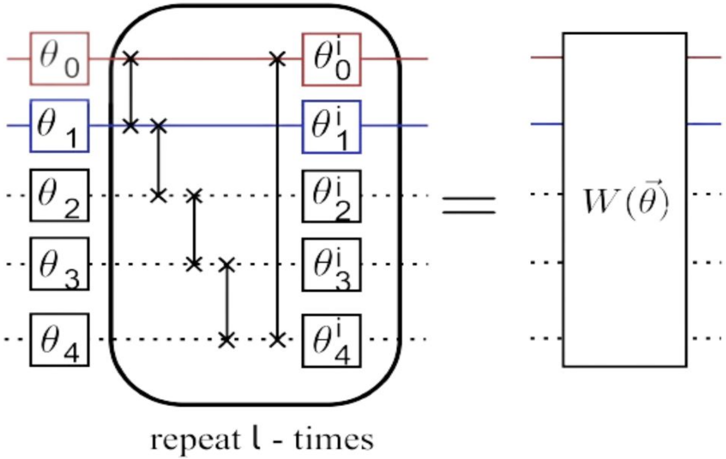
Measurement

Unitary

$$p_y = \frac{1}{2}\left(1 + y\langle\Phi(\vec{x})\,|W^\dagger(\theta)\ \mathbf{f}\ W(\theta)|\ \Phi(\vec{x})\rangle\right)$$

binary measurement $\{M_y\}$ to the state $W(\vec{\theta})\mathcal{U}_{\Phi(\vec{x})}|0\rangle^n$

Circuit Diagram: Short Depth Quantum Circuit

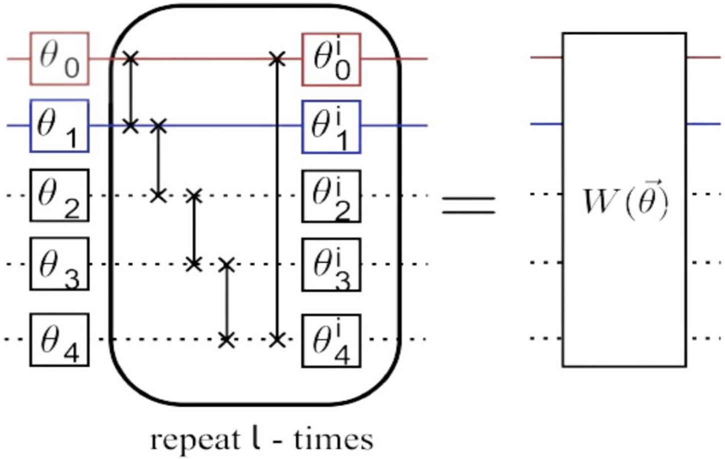


Measurement

Implementation

$$\tilde{m}(\vec{x}) = \text{sign}\left(\frac{1}{2^n}\sum_{\alpha} w_\alpha(\theta)\Phi_\alpha(\vec{x}) + b\right)$$

Circuit Diagram: Short Depth Quantum Circuit

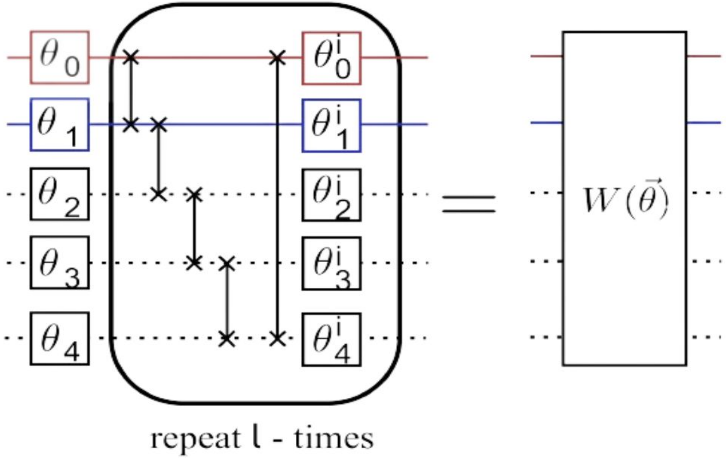


Calculating Error and Optimization

Implementation

$$R_{\text{emp}}(\vec{\theta}) = \frac{1}{|T|} \sum_{\vec{x} \in T} \Pr\left(\tilde{m}(\vec{x}) \neq m(\vec{x})\right)$$

Circuit Diagram: Short Depth Quantum Circuit



Calculating Error and Optimization

Implementation

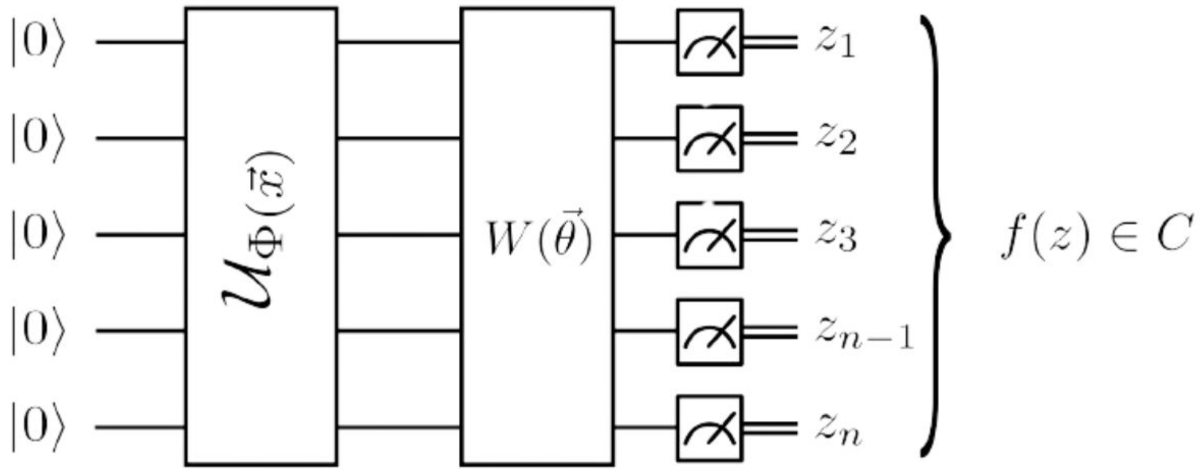
binomial cumulative density function CDF of the empirical distribution $\hat{p}_y(\vec{x})$

Binomial CDF can be approximated for large sample size (R shots) with sigmoid function

$$\Pr\left(\tilde{m}(\vec{x}) \neq m(\vec{x})\right) \approx \operatorname{sig}\left(\frac{\sqrt{R}\left(\frac{1}{2} - \left(\hat{p}_y(\vec{x}) - \frac{yb}{2}\right)\right)}{\sqrt{2(1-\hat{p}_y(\vec{x}))\hat{p}_y(\vec{x})}}\right)$$

Quantum Variational Classification Recap

$$p_y = \langle \Phi(\vec{x}) | W^\dagger(\vec{\theta}) M_y W(\vec{\theta}) | \Phi(\vec{x}) \rangle$$



Algorithm 1 Quantum variational classification: the training phase

Input Labeled training samples $T = \{\vec{x} \in \Omega \subset \mathbb{R}^n\} \times \{y \in C\}$, Optimization routine,
Parameters Number of measurement shots R, and initial parameter $\vec{\theta}_0$.
Calibrate the quantum Hardware to generate short depth trial circuits.
Set initial values of the variational parameters $\vec{\theta} = \vec{\theta}_0$ for the short-depth circuit $W(\vec{\theta})$
while Optimization (e.g. SPSA) of $R_{\mathrm{emp}}(\vec{\theta})$ has not converged **do**
  for $i = 1$ `to` $|T|$ **do**
    Set the counter $\mathrm{r}_y = 0$ for every $y \in C$.
    for *shot* $= 1$ `to` R **do**
      Use $\mathcal{U}_{\Phi(\vec{x}_i)}$ to prepare initial feature-map state $|\Phi(\vec{x}_i)\rangle\langle\Phi(\vec{x}_i)|$
      Apply discriminator circuit $W(\vec{\theta})$ to the initial feature-map state .
      Apply $|C|$ - outcome measurement $\{M_y\}_{y\in C}$
      Record measurement outcome label y by setting $\mathrm{r}_y \to \mathrm{r}_y + 1$
    end for
    Construct empirical distribution $\hat{p}_y(\vec{x}_i) = \mathrm{r}_y R^{-1}$.
    Evaluate $\Pr\left(\tilde{m}(\vec{x_i}) \neq y_i | m(\vec{x}) = y_i\right)$ with $\hat{p}_y(\vec{x}_i)$ and y_i
    Add contribution $\Pr\left(\tilde{m}(\vec{x_i}) \neq y_i | m(\vec{x}) = y_i\right)$ to cost function $R_{\mathrm{emp}}(\vec{\theta})$.
  end for
  Use optimization routine to propose new $\vec{\theta}$ with information from $R_{\mathrm{emp}}(\vec{\theta})$
end while
return the final parameter $\vec{\theta}^*$ and value of the cost function $R_{\mathrm{emp}}(\theta^*)$

Optimization Algorithm

Spall's SPSA stochastic gradient descent

"algorithm performs well in the noisy experimental setting"

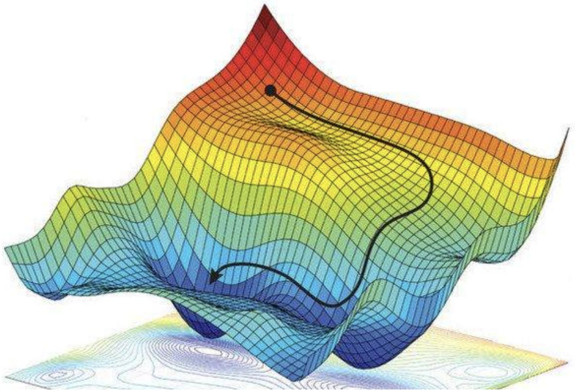


Image source: https://dtmvamahs40ux.cloudfront.net/gl-academy/course/course-1281-Non-convex-optimization-We-utilize-stochastic-gradient-descent-to-find-a-local-optimum.jpg

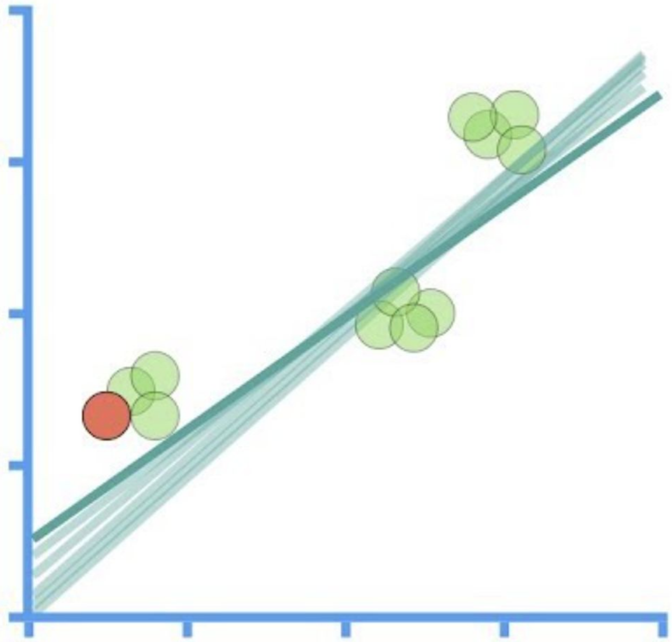


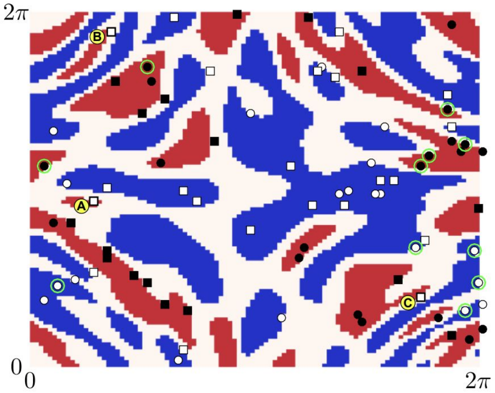
Image source: https://i.ytimg.com/vi/vMh0zPT0tLI/maxresdefault.jpg

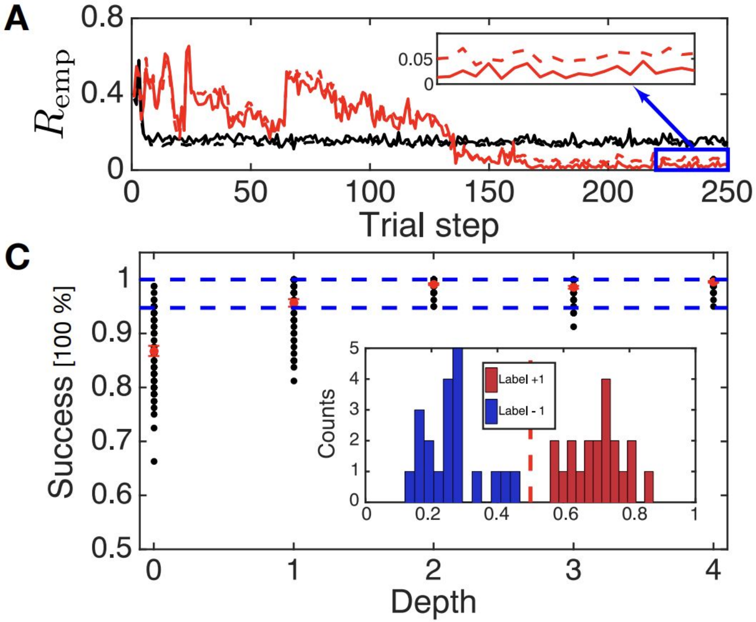
Results

Histogram shows probability of measuring label +1 for a test set of 20 points per label obtained

Red: +1
Blue: −1

Generated with a separation gap of magnitude 0.3 between them.





Results Continued

“clearly see an increase in classification success with increasing circuit depth… reaching values very close to 100% for depths larger than 1”

Results Interpretation

Quantum advantage can only be obtained by classically hard to estimate kernel.

Experimentally demonstrated a classifier that exploits a quantum feature space.

In the experiment we find that even in the presence of noise, we are capable of achieving success rates up to 100%.

Future: improvement on real world data sets

Extra Resources:

https://context-switching.com/tcs/quantumenhancedfeaturespace.html

Thank you for coming to my TED talk

No questions!!

Are there any questions?

Thank you again!

End scene.