Deep quantum neural networks form Gaussian processes

Diego García-Martín

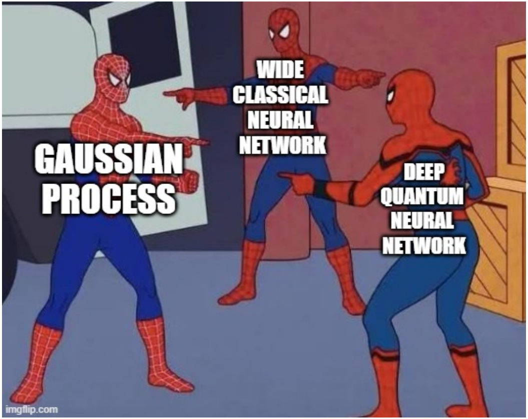
Los Alamos National Laboratory

dgarciamartin@lanl.gov

November 21st

Overview

1. Gaussian processes
2. GPs in classical ML
3. GPs in QML
4. Proof technique
5. Conclusions

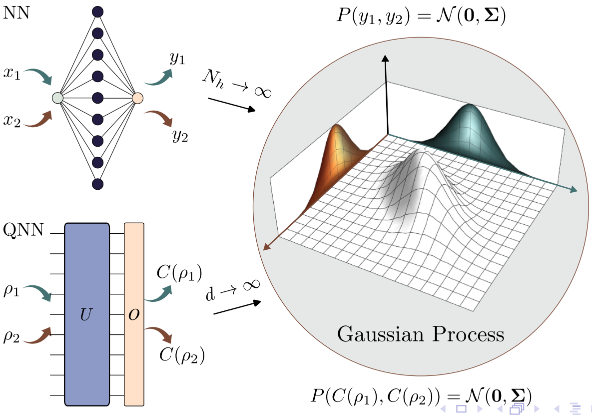


Gaussian processes

Gaussian processes

Definition (Gaussian process)

A collection of random variables $\{X_1, X_2, \dots\}$ is a GP if and only if, for every finite set of indices $\{1, 2, \dots, m\}$, the vector $(X_1, X_2, \dots, X_m)$ follows a multivariate Gaussian distribution, $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$.



GPs in classical ML

GPs in classical ML

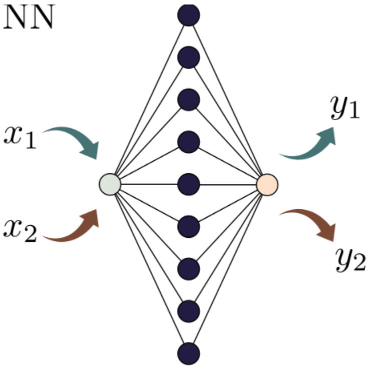
Theorem (Central limit theorem)

Let $\{X_1, X_2, \ldots, X_S\}$ be i.i.d. random variables. The sum $X_1 + X_2 + \cdots + X_S$ converges to a Gaussian distribution when $S \to \infty$.

The input of the network is $x \in \mathbb{R}$, and the output is given by

$$f(x) = b + \sum_{l=1}^{N_h} v_l h_l(x)$$

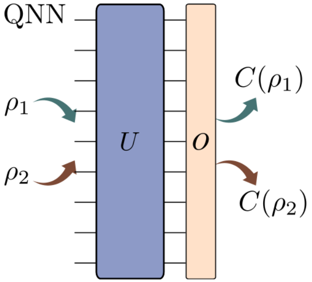
where $h_l(x) = \phi(a_l + u_l x)$ models the action of each neuron in the hidden layer.



GPs in QML

GPs in QML

- Quantum neural network (i.e. variational circuit) U
- Set of initial pure states $\{\rho_i\}_i$
- Traceless observable from a set $\{O_j\}_j$ such that $O_j^2 = \mathbb{1} \quad \forall j$
- Cost function of the form $C_j(\rho_i) = \mathrm{Tr}[U\rho_i U^\dagger O_j]$



$$\mathscr{C} = (C_j(\rho_i), \dots, C_j(\rho_{i'}), \dots) \xrightarrow{d\to\infty} \mathrm{GP}$$

GPs in QML

We cannot leverage the central limit theorem (or its variants) because of the **unitarity constraint**.

$$C_j(\rho_i) = \sum_{k,k',r,r'=1}^{d} u_{kk'}\rho_{k'r}u^*_{r'r}o_{r'k}$$

The entries of Haar random unitaries over the unitary group $\mathbb{U}(d)$ are **i.d.**, but **correlated** as $\frac{-1}{d-1}$ (same row) or $\frac{1}{(d-1)^2}$ (different row).

Strategy

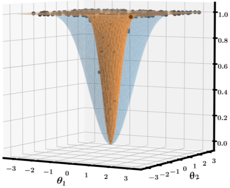
i) Computing all the moments over the Haar measure of certain groups (unitary and orthogonal).

ii) Showing they match those of a multivariate Gaussian.

GPs in QML

First moments

$$\mathbb{E}_{\mathbb{U}(d)}[C_j(\rho_i)] = \mathbb{E}_{\mathbb{O}(d)}[C_j(\rho_i)] = 0$$



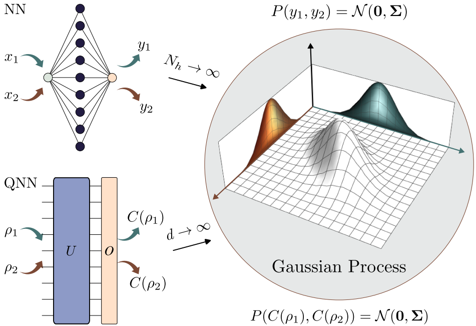
Second moments

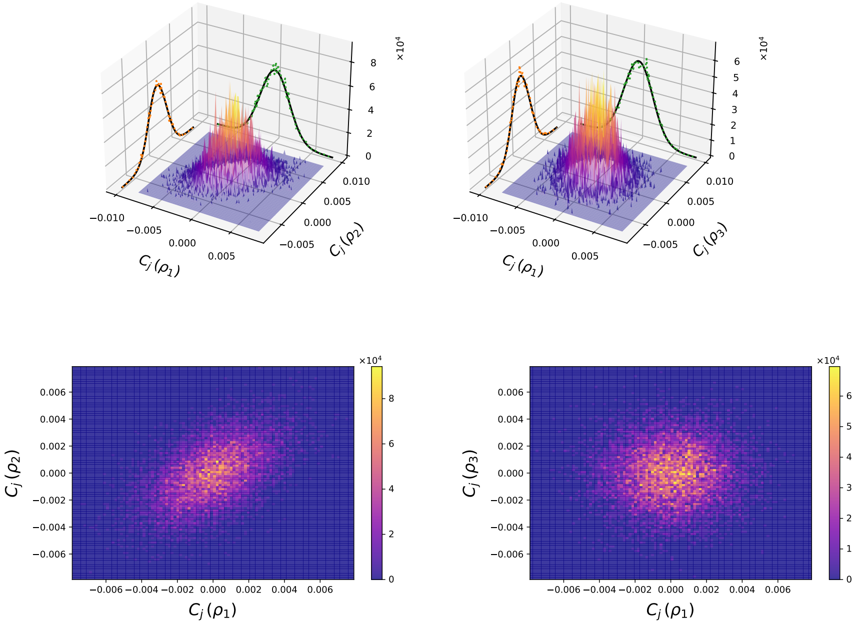
$$\boldsymbol{\Sigma}^{\mathbb{U}}_{i,i'} = \frac{d}{d^2-1}\left(\mathrm{Tr}[\rho_i\rho_{i'}] - \frac{1}{d}\right)$$

$$\boldsymbol{\Sigma}^{\mathbb{O}}_{i,i'} = \frac{2(d+1)}{(d+2)(d-1)}\left(\mathrm{Tr}[\rho_i\rho_{i'}]\left(1-\frac{1}{d+1}\right) - \frac{1}{d+1}\right)$$

GPs in QML

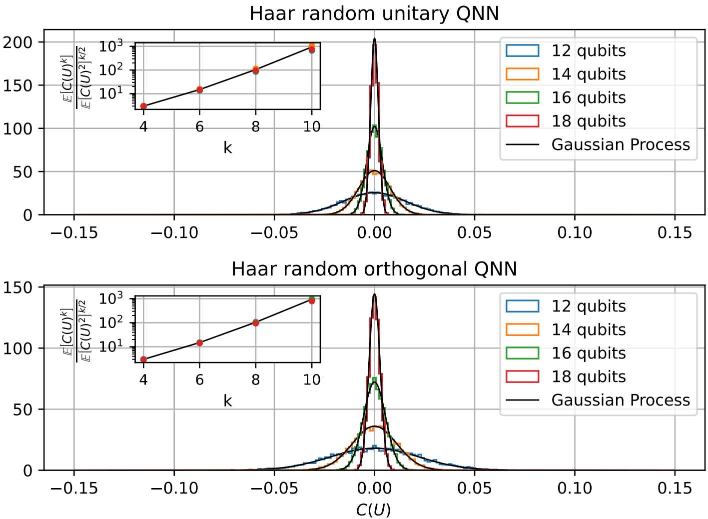
Dataset. For all $\rho_i, \rho_{i'} \in \mathscr{D}$:	GP	Correlation	Statement
$\mathrm{Tr}[\rho_i \rho_{i'}] \in \Omega\left(\frac{1}{\mathrm{poly}(\log(d))}\right)$	Yes	Positive	Theorem 1
$\mathrm{Tr}[\rho_i \rho_{i'}] = \frac{1}{d}$	Yes	Null	Theorem 2
$\mathrm{Tr}[\rho_i \rho_{i'}] = 0$	Yes	Negative	Theorem 3




$C_j(\rho_1)$
$C_j(\rho_2)$
$C_j(\rho_3)$
×10^4
−0.010
−0.005
0.000
0.005
0.010
−0.006 −0.004 −0.002 0.000 0.002 0.004 0.006

Corollary

For any ρ_i *and* O_j*, we have* $P(C_j(\rho_i)) = \mathcal{N}(0, \sigma^2)$*, where* $\sigma^2 = \frac{1}{d}, \frac{2}{d}$ *when* U *is Haar random over* $\mathbb{U}(d)$ *and* $\mathbb{O}(d)$*, resp.*



GPs in QML

Corollary

Assuming that there exists a parametrized gate in U of the form $e^{-i\theta H}$ for some Pauli operator H, then

$$P(|C_j(\rho_i)| \geq c),\ P(|\partial_\theta C_j(\rho_i)| \geq c) \in \mathcal{O}\left(\frac{1}{ce^{dc^2}\sqrt{d}}\right).$$

Concentration of measure is doubly-exponential in the number of qudits!

Corollary

Let U be drawn from a t-design. Then, under the same conditions for which Theorems 1, 2 and 3 hold, the vector $\mathscr{C}$ matches the first t moments of a GP.

An extension of Chebyshev's inequality to higher order moments leads to

$$P(|C_j(\rho_i)| \geq c),\ P(|\partial_\theta C_j(\rho_i)| \geq c) \in \mathcal{O}\left(\frac{(2\left\lfloor\frac{t}{2}\right\rfloor)!}{2^{\left\lfloor\frac{t}{2}\right\rfloor}\left(\left\lfloor\frac{t}{2}\right\rfloor\right)!(dc^2)^{\left\lfloor\frac{t}{2}\right\rfloor}}\right)$$

New and tighter bounds for t-designs.

GPs in QML

Theorem

Given the set of observations $y(\rho_1), \ldots, y(\rho_m)$ obtained from $N \in \mathcal{O}(\mathrm{poly}(\log d))$ measurements, then the predictive distribution of the GP is trivial:

$$P(C_j(\rho_{m+1}) | C_j(\rho_1), \ldots, C_j(\rho_m)) = \mathcal{N}(0, \sigma^2)$$

We cannot use the GPs to efficiently predict new outputs of the QNN using Bayesian statistics.

Proof technique

Proof technique

We need to compute quantities of the form (for arbitrary k)

$$\mathbb{E}_G\left[\mathrm{Tr}\left[\mathrm{U}^{\otimes \mathrm{k}}\Lambda\left(\mathrm{U}^{\dagger}\right)^{\otimes \mathrm{k}}\mathrm{O}^{\otimes \mathrm{k}}\right]\right] = \mathrm{Tr}\left[\mathbb{E}_{\mathrm{G}}\left[\mathrm{U}^{\otimes \mathrm{k}}\Lambda\left(\mathrm{U}^{\dagger}\right)^{\otimes \mathrm{k}}\right]\mathrm{O}^{\otimes \mathrm{k}}\right],$$

where $\Lambda = \rho_{i_1} \otimes \cdots \otimes \rho_{i_k}$ and $G = \mathbb{U}(d), \mathbb{O}(d)$.

$\mathcal{T}_G^{(k)}[\Lambda] \equiv \int d\mu(U) U^{\otimes k}\Lambda\left(U^{\dagger}\right)^{\otimes k}$ is called the twirl over G.

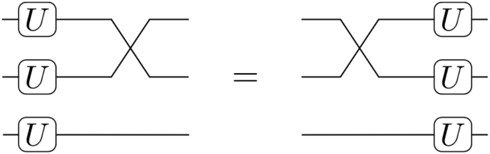
Recall that

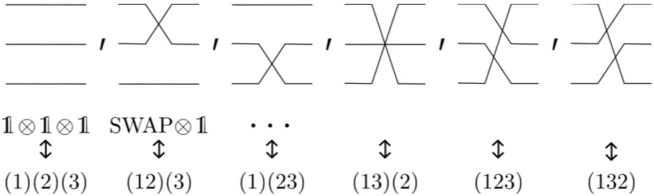
- the moments of $\mathscr{C} = (C_j(\rho_i), \ldots, C_j(\rho_{i'}), \ldots)$ are of the form $\mathbb{E}_{\mathbb{U}(d)}\left[\prod_{\lambda=1}^k C_j(\rho_{i_\lambda})\right] = \mathbb{E}_{\mathbb{U}(d)}\left[\prod_{\lambda=1}^k \mathrm{Tr}[U\rho_{i_\lambda}U^{\dagger}O_j]\right]$
- $\mathrm{Tr}[A \otimes B] = \mathrm{Tr}[A]\,\mathrm{Tr}[B]$

Proof technique

The twirl belongs to the k-th order commutant.

The k-th order commutant of the tensor representation of $U(d)$ is the set of permutations, i.e. the symmetric group S_k (Schur-Weyl duality)





Proof technique

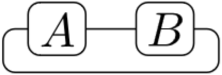
We can expand the twirl as

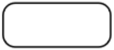
$$\mathcal{T}_G^{(k)}[\Lambda] = \sum_{\mu=1}^{D} c_\mu(\Lambda) P_\mu \,, \quad \text{with} \quad P_\mu \text{ a basis of } \mathcal{C}^{(k)}(G)$$

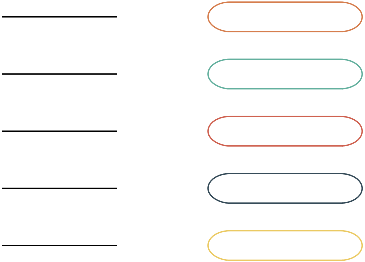
The coefficients $c_\mu(\Lambda)$ can be calculated using **Weingarten calculus**, i.e.

$$\vec{c}(\Lambda) = A^{-1} \cdot \vec{b}(\Lambda) \,,$$

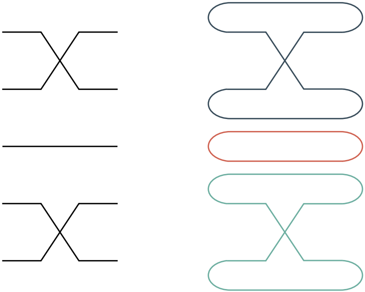
where $\vec{b}(\Lambda) = (\mathrm{Tr}[\Lambda P_1], \ldots, \mathrm{Tr}[\Lambda P_D])$ and $(A)_{\nu\mu} = \mathrm{Tr}[P_\nu P_\mu]$ is the Gram matrix.

$\mathrm{Tr}[AB] =$ 

$\mathrm{Tr}[\mathbb{1}] =$  $= d$

a) 

$e = (1)(2)(3)(4)(5)$ $\quad \chi(e) = d^5$

b) 

$\sigma = (12)(3)(45)$ $\quad \chi(\sigma) = d^3$

Proof technique

To compute higher moments, we need **asymptotic** Weingarten calculus (the A matrix has size $k! \times k!$).

$$A = d^k(\widetilde{A} + \tfrac{1}{d}B)$$

$$\widetilde{A} = \begin{pmatrix} \underbrace{\boxed{\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{matrix}}}_{I_k \times I_k} & & & \mathbf{0} \\ & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} & & \\ & & \ddots & \\ \mathbf{0} & & & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \end{pmatrix}$$

Proof technique

Let the representation of the symmetric group S_k be

$$P_d(\sigma) = \sum_{i_1,\ldots,i_k=0}^{d-1} |i_{\sigma^{-1}(1)}, \ldots, i_{\sigma^{-1}(k)}\rangle\langle i_1, \ldots, i_k|$$

We can prove that

$$\begin{aligned}\mathbb{E}_{\mathbb{U}(d)}[U^{\otimes k}\Lambda(U^\dagger)^{\otimes k}] =& \frac{1}{d^k}\sum_{\sigma\in S_k} \mathrm{Tr}[\Lambda P_d(\sigma)]P_d(\sigma^{-1}) \\ &+ \frac{1}{d^k}\sum_{\sigma,\pi\in S_k} c_{\sigma,\pi}\mathrm{Tr}[\Lambda P_d(\sigma)]P_d(\pi)\,,\end{aligned}$$

with $c_{\sigma,\pi} \in \mathcal{O}(\frac{1}{d})$.

Proof technique

The previous implies

$$\mathbb{E}_{\mathbb{U}(d)}\left[\mathrm{Tr}\left[U^{\otimes k}\Lambda\,(U^{\dagger})^{\otimes k}O^{\otimes k}\right]\right] = \frac{1}{d^k}\sum_{\sigma\in S_k}\mathrm{Tr}[\Lambda P_d(\sigma)]\mathrm{Tr}[P_d(\sigma^{-1})O^{\otimes k}]$$
$$+\frac{1}{d^k}\sum_{\sigma,\pi\in S_k} c_{\sigma,\pi}\mathrm{Tr}[\Lambda P_d(\sigma)]\mathrm{Tr}[P_d(\pi)O^{\otimes k}]$$

- $\mathrm{Tr}[P_d(\sigma)O^{\otimes k}] = 0$ if σ contains a cycle of odd length.
- $\mathrm{Tr}[P_d(\sigma)O^{\otimes k}] = d^r$ if k is even and σ is a product of r disjoint cycles of even length.
- In particular, $\mathrm{Tr}[P_d(\sigma)O^{\otimes k}] = d^{k/2}$ if k is even and σ is the product of $k/2$ disjoint transpositions.

Proof technique

We find that the moments of $\mathscr{C}$ are

$$\mathbb{E}_{\mathbb{U}(d)}\left[\mathrm{Tr}\left[U^{\otimes k}\Lambda(U^{\dagger})^{\otimes k}O^{\otimes k}\right]\right]=\frac{1}{d^{k/2}}\sum_{\sigma\in T_k}\prod_{\{t,t'\}\in\sigma}\mathrm{Tr}[\rho_t\rho_{t'}] \tag{1}$$

According to Wick's theorem, the moments of a multivariate Gaussian are

$$\mathbb{E}[C_j(\rho_i),\ldots,C_j(\rho_{i'})\ldots]=\sum_{\sigma\in T_k}\prod_{t,t'\in\sigma}\mathrm{Cov}[C_j(\rho_t)C_j(\rho_{t'})]$$

(Recall that asymptotically $\mathbf{\Sigma}^{\mathbb{U}}_{i,i'}=\frac{\mathrm{Tr}[\rho_i\rho_{i'}]}{d}$)

Conclusions

Conclusions

- Under certain conditions, the output distribution of deep Haar random QNNs converges to a Gaussian process in the limit of large Hilbert space dimension.
- The situation is more nuanced that in classical NNs. We need to make assumptions on the states processed by the QNN, as well as on the measurement operator.
- Our results may be useful in more general settings where Haar random unitaries / t-designs are considered, such as quantum information scramblers and black holes, or many-body physics.

DGM, Martin Larocca and M. Cerezo (2023)
Deep quantum neural networks form Gaussian processes
arXiv preprint: arXiv.2305.09957