Unsupervised classification of entanglement using geometry and symmetry fixing

Emil Hössjer & Vahid Azimi Mousolou

Department of Physics and Astronomy, Uppsala University

E-mail: emil.hoessjer@gmail.com

E-mail: vahid.azimi-mousolou@physics.uu.se

9 September 2024

Abstract. We model the boundary of the set of separable states using a distance function from the maximally mixed state, which we call the radius of the separable set. By sampling separable states as convex combinations of product states, we attempt a neural model to learn the radius as a function of direction. We furthermore show how the entanglement symmetry of local unitary invariance can be used to simplify data driven approaches to entanglement problems. We restrict our implementation to the two-qubit case, but the method can be generalized to any multipartite system. Our method improves on the accuracy of the previous unsupervised learning approach.

Contents

1. Introduction

Quantum entanglement is the feature of nature that the state of two or more particles can not be described by a local theory. More than a puzzling phenomenon, it is a resource that can be consumed to perform tasks such as quantum- teleportation, cryptography and error correction [1, 3, 7]. The problem of determining whether a given (mixed) state is *entangled* or *separable* is classic in quantum information theory. For the simplest bipartite systems of dimensions 2×2 and 2×3, the classification problem has an analytical solution in terms of the Peres-Horodecki (PPT) criterion, but in general the problem is NP-hard [4].

Data-driven methods are ubiquitous in the current era. The straightforward data-driven approach to entanglement classification would be to train a model to find the decision boundary between two sets of separable and entangled quantum states. Such a supervised machine learning approach requires however a large number of labelled training data, which is very computationally expensive to obtain. In this work we therefore consider an unsupervised approach to the classification problem, which circumvents the need of labelled data. In an unsupervised approach, separable states are sampled as convex combinations of product states and a model is trained to delimit the separable domain based on the data. In a previous paper [2] a generative model using a generative adversarial net (GAN) was considered for this. This model works by learning the distribution of the sampled data, and any new data that is deemed unlikely to come from the same distribution is predicted to be entangled. However, this attempt did not make use of any domain specific knowledge about the set of separable states. Rather it was a generic approach that can be applied to any set from which one can sample. In the current work we instead take a geometric approach to learning the separable domain, exploiting convexity to model the set by its radius from the center. This amounts to learning the boundary of the separable set, and everything outside the learned boundary is predicted to be entangled. We furthermore show how the local unitary symmetry of entanglement can be fixed, which simplifies the set of density matrices and produces a simpler machine learning problem.

Our presentation of the method is general, but we restrict the implementation to the two-qubit case. As explained, this is not a case where a learning based classification model is of interest, but it serves well as an initial testing ground since evaluating the model is simple using the PPT criterion. Furthermore the two-qubit case does not yet have a satisfactory solution within an unsupervised machine learning framework. Our method can be implemented for any multipartite discrete system.

To sample a test set we use one of the principal candidates for measures on the set of density matrices and we balance the obtained set to get an equal number of separable and entangled states. Our model improves the classification accuracy of the GAN model from ca 76% to 86%. This number is further increased to 94% if we use symmetry. We note however that our model performs imperfectly on the given data - it misses out on a lot of information. We point out a solution that we think would improve the accuracy

considerably.

Code is available on `https://github.com/emho7468/...`

2. Density matrices

For an N-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^N$, the pure states are rays $|\phi\rangle \in \mathbb{C}P^{N-1}$. Density matrices are obtained as convex combinations of pure states

$$\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|,$$

where $p_i > 0$ and $\sum_i p_i = 1$. In a given basis, ρ can be represented by a matrix which is non-negative definite and has unit trace. We will denote the set of density matrices by $\mathcal{D}$. The space $\mathcal{D}$ is convex, and regarded as real it has dimension $N^2 - 1$. Entanglement arises in bipartite (or multipartite) systems $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Separable (non-entangled) mixed states are those ρ that can be obtained as convex combinations of product states

$$\rho = \sum p_i \rho_i^A \otimes \rho_i^B, \tag{1}$$

where $\rho_i^{A/B}$ are density matrices on $\mathcal{H}_{A/B}$ and the p_i are positive weights summing to 1. We will denote the set of separable matrices by $\mathcal{S}$. Clearly $\mathcal{S}$ is convex as well.

There is a unique and natural measure on $\mathbb{C}P^{N-1}$ which is unitarilay invariant, given by the induced Haar measure of $U(N)$. There is however no natural measure on the set of mixed states, but all of the principal candidates preserve the unitary invariance under $\rho \to U\rho U^\dagger$ [10]. For any such measure ν on $\mathcal{D}$, the center of mass is the maximally mixed state $\frac{1}{N} I_N$. This can be seen by applying a unitary transformation U to

$$\mathbb{E}[\rho] = \frac{1}{\mathrm{Vol}(\mathcal{D})} \int_{\mathcal{D}} \rho \; d\nu(\rho),$$

which gives

$$\begin{aligned} U\mathbb{E}[\rho]U^\dagger &= \frac{1}{\mathrm{Vol}(\mathcal{D})} \int_{\mathcal{D}} U\rho U^\dagger \; d\nu(\rho) \\ &= \frac{1}{\mathrm{Vol}(\mathcal{D})} \int_{\mathcal{D}} \rho \; d\nu(U^\dagger \rho U) \\ &= \mathbb{E}[\rho]. \end{aligned} \tag{2}$$

But for this to hold for all unitary U we must have $\mathbb{E}[\rho] = \lambda I_N$, where unit trace implies $\lambda = \frac{1}{N}$.

Let $||\rho|| := \sqrt{\mathrm{Tr}(\rho^2)}$ denote the Frobenius (Hilbert-Schmidt) norm, and define a radius function R as the distance from the center

$$R(\rho) = ||\rho - \frac{1}{N} I_N||.$$

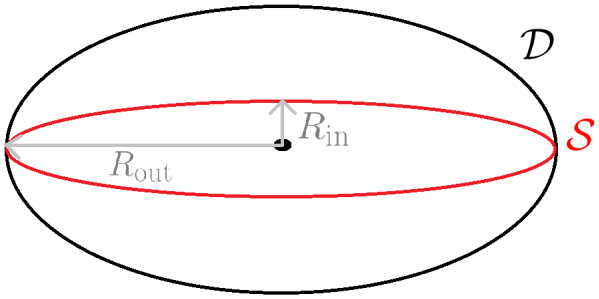

Figure 1. Simple two dimensional depiction of $\mathcal{S} \subset \mathcal{D}$.

For any pure state $|\phi\rangle\langle\phi|$ we have

$$R\big(|\phi\rangle\langle\phi|\big) = \sqrt{1 - \frac{1}{N}} =: R_{\text{out}}, \tag{3}$$

and since the pure states are the extreme points of $\mathcal{D}$ it follows that the set of density matrices is contained inside a ball of radius $\sqrt{1 - \frac{1}{N}}$ centered at $\frac{1}{N} I_N$. This is the smallest ball that contains $\mathcal{D}$ and $\mathcal{S}$, as in particular equation (3) holds for pure product states. For bipartite systems, the largest ball centered at $\frac{1}{N} I_N$ that is contained in $\mathcal{S}$ was shown in [5] to have radius

$$R_{\text{in}} = \sqrt{\frac{1}{N(N-1)}}. \tag{4}$$

For multipartite systems, a lower bound is given in [6]. In terms of convexity theory, the radii in (3) and (4) are known as the *outradius* and *inradius* of the set $\mathcal{S}$, see figure 1. In particular, the set of separable states is determined by a radius function R that is given by the distance from $\frac{1}{N} I_n$ to $\partial\mathcal{S}$ in different directions, and which is bounded by the in- and outradius. In this paper we shall try to learn this function using data from $\mathcal{S}$, but we will use a parametrized description which preserves the above considerations. To do this, we expand $\rho \in \mathcal{D}$ in a traceless Hermitian ON-basis h_i, $i = 1, \ldots, N^2 - 1$ as

$$\rho = \frac{1}{N} I_N + \sum_i x_i h_i$$

with coordinates $x_i \in \mathbb{R}$. In terms of these coordinates the radius function becomes

$$R\big(\rho(x)\big) = ||\sum_i x_i h_i|| = \sqrt{\sum_{i,j} x_i x_j \operatorname{Tr}(h_i h_j)} = \sqrt{\sum_i x_i^2} = |x|,$$

since $\operatorname{Tr}(h_i h_j) = \delta_{ij}$. Thus, using a traceless Hermitian ON-basis we get coordinates on the set of density matrices in terms of which $\mathcal{D}$ is a convex set centered at origo. Furthermore, $\mathcal{S}$ is contained inside a Euclidean ball of radius R_{out} and contains a ball of radius R_{in}, both centered at origo.

Our specific implementation concerns one-qubit states and two-qubit states. The one-qubit case is parametrized by

$$\rho = \frac{1}{2} I_2 + \frac{1}{\sqrt{2}} r \cdot \sigma,$$

where $r \in \mathbb{R}^3$ and σ is the vector of Pauli matrices. The parametrization is valid for all $|r| \leq R_{\text{out}} = \sqrt{\frac{1}{2}}$, yielding the Bloch ball. A two-qubit state is parametrized by

$$\rho = \frac{1}{4} I_4 + \frac{1}{2}\big(r \cdot \sigma \otimes I_2 + I_2 \otimes s \cdot \sigma + \sum_{i,j} t_{ij} \sigma_i \otimes \sigma_j\big) \tag{5}$$

where $(r_i, s_i, t_{ij}) \in \mathbb{R}^{15}$. Any coordinate can be retrieved by an operation similar to

$$r_x = \text{Tr}\,\big(\frac{\sigma_x \otimes I_2}{2} \rho\big).$$

The in- and outradii are $\frac{1}{2\sqrt{3}}$ and $\frac{\sqrt{3}}{2}$ respectively.

3. Neural model

We want to train a neural model to learn the radius of the set $\mathcal{S}$ as a function of direction, based on a training set of interior points. Using coordinates as in section 2 we convert density matrices $x \in \mathbb{R}^{N^2-1}$ to machine learning features as

$$\text{(input) } N(x) = \frac{x}{|x|} \in S^{N^2-2} \tag{6}$$

$$\text{(output) } R(x) = |x| \in [0, R_{\text{out}}]. \tag{7}$$

A test point will be predicted to be outside our learned domain (entangled) if $R_{\text{test}} > R_{\text{pred}}$. In accordance with classification nomenclature we define

$$\begin{aligned} \text{FP} &= \text{ number of separable points predicted to be entangled} \\ \text{FN} &= \text{ number of entangled points predicted to be separable} \\ \text{FPR} &= \frac{FP}{N_{\text{separable}}}, \quad \text{FNR} = \frac{FN}{N_{\text{entangled}}}. \end{aligned}$$

As loss function between a training radius R_{train} and a predicted radius R_{pred} we take the skewed loss

$$\mathcal{L}(R_{\text{train}}, R_{\text{pred}}) = \begin{cases} R_{\text{pred}} - R_{\text{train}} & \text{if } R_{\text{pred}} > R_{\text{train}} \\ k(R_{\text{train}} - R_{\text{pred}}) & \text{otherwise} \end{cases}$$

where $k > 0$ is a hyperparameter. A large k implies

- we penalize undervaluing the radius function
- low classification error on training data

- high FNR

whereas the opposite holds for small k. Our prescription in the choice of k then is that for a given neural model and training length (choice of epoch number, batch size and learning rate), k should be chosen so that at the end of training we have FPR $\approx$ FNR. Since we use a perfectly balanced test set this means that FP $\approx$ FN, or roughly half of the incorrectly classified test points are predicted to be outside the separable domain and half are predicted to be inside. We believe that this is a reasonable objective to achieve the highest overall accuracy on the test set.

As neural model we found that a fully connected architecture with ReLU activation and hidden layer nodes $(64, 32, 32, 16, 16, 8)$ performs well. For the symmetry reduced data in section 5, a smaller network with hidden nodes $(32, 32, 16, 16, 8)$ is used.

4. Sampling

Sampled density matrices are converted to coordinates $x \in \mathbb{R}^{N^2-1}$ following section 2 and subsequently to input- and output features $N(x)$ and $R(x)$ following section 3. In this section we discuss the sampling of density matrices for training- and test data respectively. Our discussion of the training data partly focuses on the two-qubit system, which is the case that we implement in this paper.

4.1. Training set

We sample separable density matrices according to

$$\rho_{\text{train}} = \sum_{i=1}^{n_{\text{mix}}} p_i \rho_i^A \otimes \rho_i^B \tag{8}$$

where the $p_i > 0$ are random weights summing to 1 and the $\rho_i^{A/B}$ are $2 \times 2-$density matrices. To decide on how to sample, we make the following observations:

- We want to sample as close to the boundary of separable states $\partial\mathcal{S}$ as possible, as it is the boundary that we are trying to learn.
- Along a given direction $N(x)$ of a point x, the further the point is from the boundary the more mixing it has and vice versa. Therefore we want to keep mixing low.
- Mixing increases with n_{mix}.
- Mixing increases with the mixing of the factors $\rho_i^{A/B}$.
- Any separable two-qubit state has a decomposition (8) in terms of pure factors of at most 4 terms. [8]

In light of these observations we sample pure factors $\rho_i^{A/B}$ and we use $n_{\text{mix}} = 4$. Furthermore we shall impose the requirement that the radius of any training point exceeds the minimum radius in equation (4). Data with lower radii than the one known to be the minimum do not offer any new information about the separable domain.

To sample pure states we use the natural measure induced by the Haar measure from the unitary group. A sampling scheme for the N-dimensional case is prescribed in [10]. However, for the case $N = 2$ we can just use the Bloch representation. Unitary invariance translates into rotational invariance for $\hat{r}$ in $\rho = \frac{1}{2}I_2 + \frac{1}{\sqrt{2}}\hat{r}\cdot\sigma$, so we can use any rotationally invariant sampling scheme of the sphere S^2 to sample pure qubit states $\rho_i^{A/B}$.

To sample the p_i we let $\tilde{p}_i \sim U(0,1)$ and set $p_i = \tilde{p}_i / \sum_i \tilde{p}_i$.

For a general N-dimensional quantum system there is no result of how many terms are minimally required in a decomposition of a separable state into pure product states as in (8). However, the Carathéodory theorem provides an upper bound of N^2. In an implementation for a larger quantum system we would suggest to use a range of values below this bound.

4.2. Test set

There are multiple measures on the space of density matrices that one can use to sample a test set [10]. In this paper we use the map

$$T(M) = \frac{MM^\dagger}{\mathrm{Tr}(MM^\dagger)}, \tag{9}$$

which has range all of $\mathcal{D}$. We let M follow a complex Ginibre ensemble, i.e. $M = Z_1 + iZ_2$ where the entries of $Z_1, Z_2 \in \mathbb{R}^{N\times N}$ are iid random variables distributed according to the normal distribution $\mathcal{N}(0,1)$. The resulting measure is equivalent to the one induced by the Frobenius norm [10]. We employ the PPT-criterion to label the data as entangled or separable. For the two-qubit case ca 34% of states sampled under this measure are separable. For larger systems this imbalance is expected to be worse [9]. To avoid having a naive classifier achieve well over 50% on the test set, we discard the surplus of entangled states and use a perfectly balanced set of 40000 separable and 40000 entangled states.

5. Symmetry fixing

Entanglement is invariant under local unitary operations. In this section we propose to fix this symmetry, which essentially reduces the set of density matrices to equivalence classes under transformations by the group of locally unitary operators. We provide an explicit construction in the two-qubit case which straightforwardly generalizes to any n-qubit system.

A locally unitary operation on a two-qubit state ρ is a transformation

$$\rho \to u_a \otimes u_b \; \rho \; (u_a \otimes u_b)^\dagger,$$

where $u_a, u_b \in U(2)$. Fixing this symmetry reduces $\mathcal{D}$ to $\mathcal{D}/(U(2)\otimes U(2))$. As a first step we want a unitary $2\times 2-$matrix M such that

$$Mr\cdot\sigma M^\dagger = |r|\sigma_z.$$

This corresponds to a coordinate frame in which the Bloch vector of a single qubit points in the positive z direction. For

$$M = \begin{pmatrix} a & -e^{-i\varphi}b \\ b^* & e^{i\varphi}a^* \end{pmatrix}$$

where we may assume wlog that $a \geq 0$ is real, we get

$$\begin{aligned} a &= \sqrt{\frac{1}{2}\Big(1 + \frac{r_z}{|r|}\Big)} \\ |b| &= \sqrt{\frac{1}{2}\Big(1 + \frac{r_z}{|r|}\Big)} \\ \arg(b) &= \operatorname{sign}(r_y) \arccos\big(-\frac{r_x}{2a|b||r|}\big) \\ \varphi &: \text{arbitrary.} \end{aligned}$$

The fact that φ is arbitrary means that there is a residual symmetry after we fix r. This corresponds to a rotation in the xy-plane. Let M_r denote the above transformation. We have for

$$\rho = \frac{1}{4}I_4 + \frac{1}{2}\big(r \cdot \sigma \otimes I_2 + I_2 \otimes s \cdot \sigma + \sum_{i,j} t_{ij}\sigma_i \otimes \sigma_j\big)$$

that

$$M_r \otimes M_s \rho (M_r \otimes M_s)^\dagger = \frac{1}{4}I_4 + \frac{1}{2}\big(|r|\sigma_z \otimes I_2 + I_2 \otimes |s|\sigma_z + \sum_{i,j} \tilde{t}_{ij}\sigma_i \otimes \sigma_j\big)$$

for some new coordinates $\tilde{t}_{ij}$. For the remaining symmetry, let $N = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{pmatrix}$ and consider the action of N on $t \cdot \sigma$. The z-coordinate is invariant but for $\varphi = -\arg(t_x + it_y)$ we have

$$Nt \cdot \sigma N^\dagger = \sqrt{t_x^2 + t_y^2}\sigma_x + t_z\sigma_z.$$

Letting N_{t_x,t_y} denote this N we have that

$$\begin{aligned} &N_{\tilde{t}_{xz},\tilde{t}_{yz}} \otimes N_{\tilde{t}_{zx},\tilde{t}_{zy}} M_r \otimes M_s \rho (M_r \otimes M_s)^\dagger (N_{\tilde{t}_{xz},\tilde{t}_{yz}} \otimes N_{\tilde{t}_{zx},\tilde{t}_{zy}})^\dagger \\ =&\frac{1}{4}I_4 + \frac{1}{2}\Big[|r|\sigma_z \otimes I_2 + I_2 \otimes |s|\sigma_z + \sum_{i,j\in\{x,y\}} \hat{t}_{ij}\sigma_i \otimes \sigma_j \\ &+ \sqrt{\tilde{t}_{xz}^2 + \tilde{t}_{yz}^2}\sigma_x \otimes \sigma_z + \sqrt{\tilde{t}_{zx}^2 + \tilde{t}_{zy}^2}\sigma_z \otimes \sigma_x + \tilde{t}_{zz}\sigma_z \otimes \sigma_z\Big], \end{aligned}$$

for some new coordinates $\hat{t}_{ij}$. This symmetry fixing reduces a 15-dimensional space to a 9-dimensional space and furthermore 4 of the remaining coordinates are non-negative. Previously we had to learn a set contained in a 15-dimensional ball, but now we only have to learn a set confined to a fraction $\left(\frac{1}{2}\right)^4$ of a 9-dimensional ball.

The generalization to n qubits is straightforward using a similar basis: all single index components are fixated along the positive z-axis and for the n-index component $t_{i_1\cdots i_n}$, all pairs of the form $t_{z\cdots x\cdots z}, t_{z\cdots y\cdots z}$ are fixated along the positive x-axis. Such symmetry fixing reduces the dimension from 4^n-1 to $d_n = 4^n - 3n - 1$ and the resulting space is confined to a fraction $\left(\frac{1}{2}\right)^{2n}$ of a d_n-dimensional ball.

6. Results

All accuracies are reported on a test set of 40000 separable and 40000 entangled states, sampled according to section 4.2. Example runs with and without symmetry fixing are shown in figure 2 (a) and (b). In figure 2 (c) we consider the role of the number of training data for performance. In this case we train for a long time - 400 epochs - to counteract the effect of more batch updates when using more data. After 400 epochs the accuracy has approximately stagnated regardless of the number of training data. In figure 2 (d) we show the average sampled radius for different choices of n_{mix}. "With inradius" means that we discard data within the trivial inradius. Finally, some numerical values are presented in table 1. Here "Radius-model" denotes the model presented in this paper and "GAN" signifies the generative model used for unsupervised learning in the paper by Chen et al [2]. Our accuracy of the GAN model differs from the one reported in that paper because we use a different test set, see Appendix A for details.

Model	Radius-model ($k = 1.5$)	Radius-model w/ symmetry ($k = 7.0$)	GAN	GAN w/ symmetry
n_{train}	10^5	10^5	$1.6 \cdot 10^5$	$1.6 \cdot 10^5$
n_{epoch}	200	200	100	100
Accuracy	85.8%	93.7%	75.6%	79.7%
Time/epoch [s]	18.8	13.8	48.6	48.5

Table 1. Model comparison. Results shown are averages from three runs. Our accuracy of the GAN model differs from the one reported in [2] because we use a different test set, see Appendix A for details.

7. Discussion

We see from figure 2 (a) that the classification accuracy on the training set is only at around 60%. Since we *know* that this data is part of the separable domain, an accuracy below 100% means that we do not use all of the information contained in the training set. The cause of the low training accuracy is the surprisingly low k-value of $k = 1.5$ (chosen according to section 3 to achieve the highest test accuracy). The situation is better in the symmetry fixed case shown in figure 2 (b). In this case, a higher value of $k = 7$ is preferable and the training accuracy is almost 90%. What we have derived then is a rather unconventional regression task. There is simply too much "bad" data

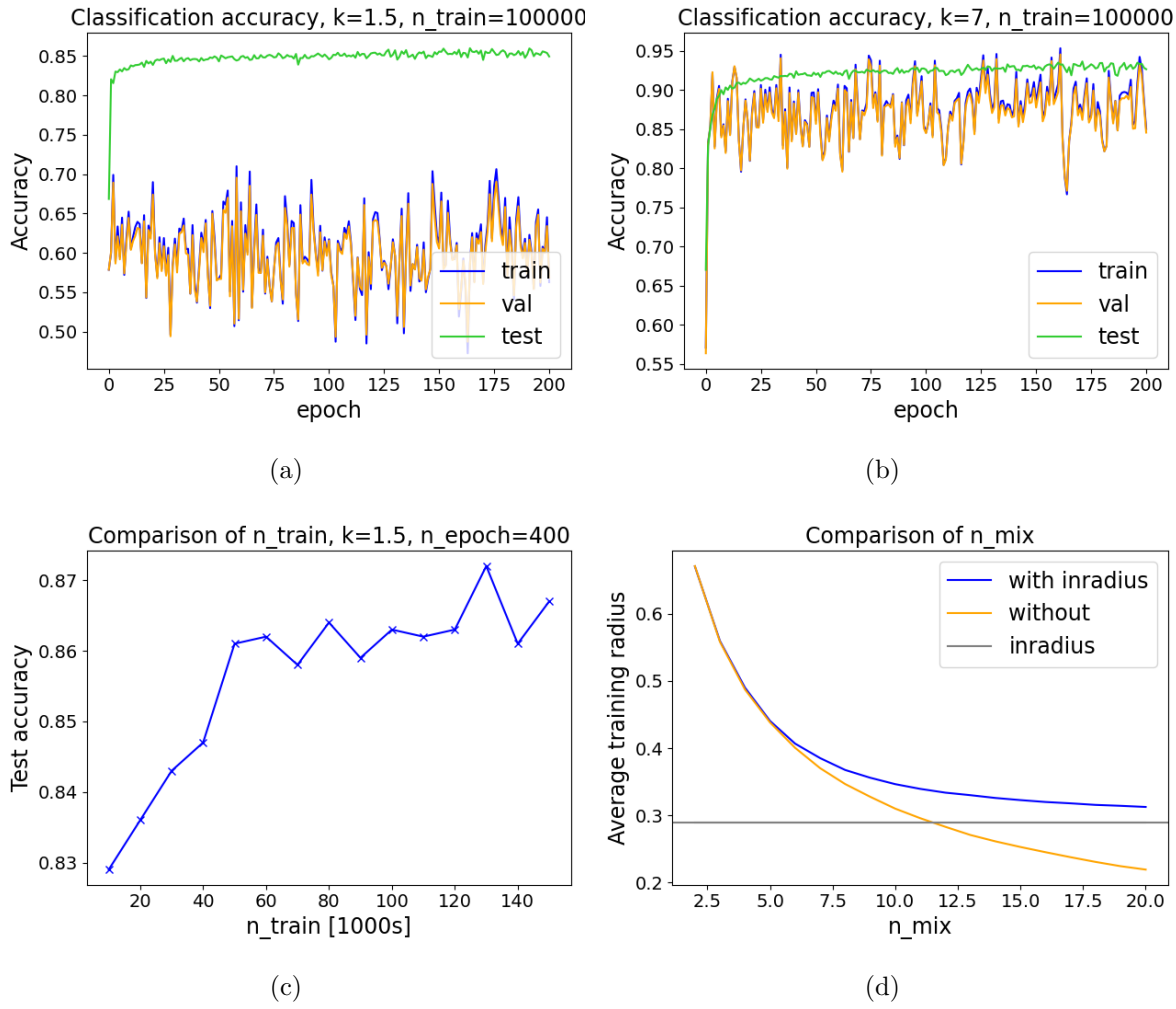

Figure 2. (a) Example run (b) Example run with symmetry fixing (c) Results for different n_{train} (d) Average sampled training radii for different n_{mix}

that does not contribute to the exterior of the sampled domain and which obstructs the regression solution. Just adapting the loss parameter k is not enough to address this. To improve the accuracy further, it seems that we would need to discard this bad data.

From figure 2 (c) we see that the performance of the model does not depend very much on the number of training data. Already for 10000 training data we achieve 83% accuracy and for $n_{\text{train}} \geq 50000$ we get over 86%. It is good that the model performs well with little data, but it is bad that it does not improve more as data increases. We believe that this lack of performance is due to the point previously made: more data means more information about the separable domain, but also more bad data that impedes the regression solution.

In figure 2 (d) we see the rapid decrease in radius with increasing n_{mix}. We also see that for our choice of $n_{\text{mix}} = 4$, a lower constraint given by the inradius does not do much. For larger n_{mix} this bound starts to have a great impact. This would be useful in the consideration of a larger quantum system.

We achieve 94% accuracy on the test set which is very good. However, central to this success is symmetry fixing and from the dimensional consideration in section 5, we expect the improvement from symmetry fixing to be smaller for multipartite systems. The accuracy of 86% without symmetry fixing is better than the 76% of the GAN model, but one would likely want a higher number. We think that this is very achievable through some clever data handling. We can for example imagine a setup where an initial training set is sampled and the model is trained for a short time, followed by the model going through the training data and discarding any data that is well within the radius predicted by the model. The model can then be retrained on the smaller data set which better approximates the separable boundary, and new data can be sampled which is selected for additional training only if it lies outside the boundary already predicted by the model.

8. Conclusion

We consider the problem of unsupervised classification of entanglement in quantum systems. The unsupervised approach is motivated by the fact that entanglement classification is in general computationally intractable, but to sample separable states is very fast. We suggest to exploit the convexity of the separable domain and learn the boundary as a radial distance function from the maximally mixed center. We also show how the symmetry of local unitary invariance related to entanglement can be fixed to simplify the set of density matrices and yield an easier machine learning problem. Evaluated on the two-qubit system, our model improves the accuracy from the previous unsupervised approach from 76% to 86% and further to 94% if symmetry fixing is employed. Our results reveal a serious shortcoming of the method however. The model fails to learn a large proportion of the training data. Our proposed solution to this issue is to discard training data that do not contribute to the exterior of the sampled domain. We believe that with a clever way of handling training data, the model presented in this paper can learn the separable domain of any quantum system to a high accuracy.

Appendix A.

In this section we clarify the difference between the test set that we use and the test set employed in the paper by Chen et al [2] which leads to the discrepancy in reported accuracy of the model in said paper, here referred to as the GAN model.

As specified in section 4.2, we sample a test set using the map

$$T(M) = \frac{MM^\dagger}{\mathrm{Tr}(MM^\dagger)} \tag{A.1}$$

where M follows a complex Ginibre ensemble. Chen et al sample entangled states using this measure, but for separable states they use the same sampling scheme as they do

for their training set. Their training set is sampled as

$$\sum_{i=1}^{n_{\text{mix}}} p_i \rho_i^A \otimes \rho_i^B$$

where $\rho_i^{A/B}$ are mixed $2\times2-$density matrices and they let n_{mix} vary from 1 to 20. Let us denote our test set by "test set 1" and theirs by "test set 2". From a radial perspective it is easy to see that test set 2 is easier to classify correctly than test set 1. Namely, using mixed factors $\rho_i^{A/B}$ and an average mixing parameter of $\text{avg}(n_{\text{mix}}) = 10$ yields highly mixed states on average, which have low radius and are far from the boundary $\partial\mathcal{S}$. It is well known and intuitive that states further from the decision boundary are easier to classify. Indeed, figure A1 shows the distribution of sampled radii in the separable parts of the two test sets as well as in the entangled subset. Clearly, based on radius alone it is easier to distinguish the separable states from the entangled ones in test set 2 than it is in test set 1. For example, the classifier that predicts everything inside the inner radius R_{in} as separable and everything outside as entangled achieves 55% accuracy on test set 1 and 90% on test set 2.

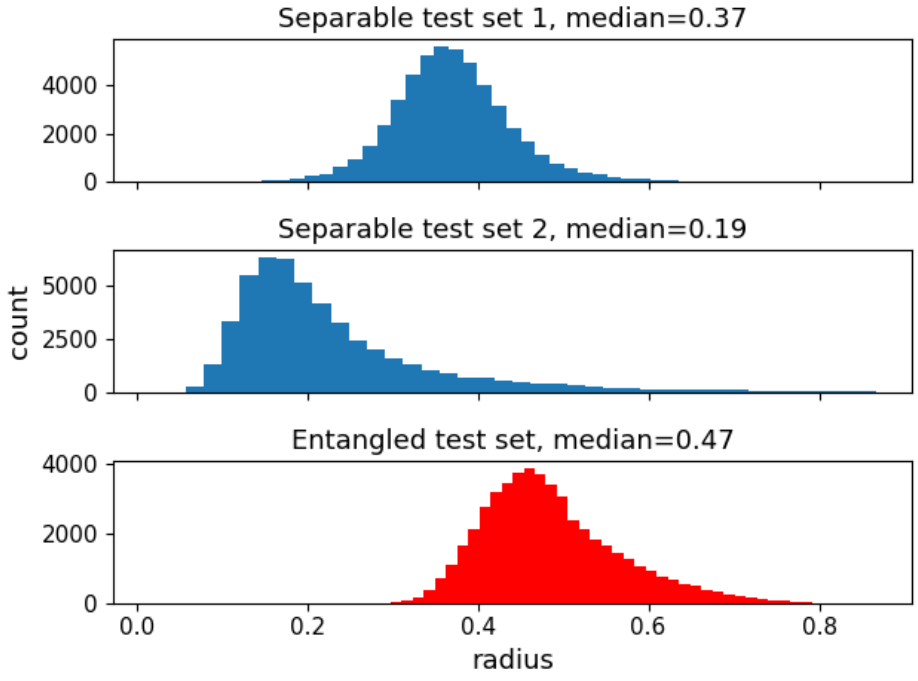

Figure A1. Radius comparison of test sets

References

[1] Dik Bouwmeester et al. "Experimental quantum teleportation". In: *Nature* 390.6660 (1997), pp. 575–579. ISSN: 1476-4687. DOI: 10.1038/37539. URL: http://dx.doi.org/10.1038/37539.

[2] Yiwei Chen et al. "Detecting quantum entanglement with unsupervised learning". In: *Quantum Science and Technology* 7.1 (2021), p. 015005. ISSN: 2058-9565. DOI: 10.1088/2058-9565/ac310f. URL: http://dx.doi.org/10.1088/2058-9565/ac310f.

[3] Artur K. Ekert. "Quantum cryptography based on Bell's theorem". In: *Phys. Rev. Lett.* 67 (6 1991), pp. 661–663. DOI: 10.1103/PhysRevLett.67.661. URL: https://link.aps.org/doi/10.1103/PhysRevLett.67.661.

[4] Leonid Gurvits. *Classical deterministic complexity of Edmonds' problem and Quantum Entanglement*. 2003. arXiv: quant-ph/0303055 [quant-ph]. URL: https://arxiv.org/abs/quant-ph/0303055.

[5] Leonid Gurvits and Howard Barnum. "Largest separable balls around the maximally mixed bipartite quantum state". In: *Physical Review A* 66.6 (2002). ISSN: 1094-1622. DOI: 10.1103/physreva.66.062311. URL: http://dx.doi.org/10.1103/PhysRevA.66.062311.

[6] Leonid Gurvits and Howard Barnum. "Separable balls around the maximally mixed multipartite quantum states". In: *Physical Review A* 68.4 (2003). ISSN: 1094-1622. DOI: 10.1103/physreva.68.042312. URL: http://dx.doi.org/10.1103/PhysRevA.68.042312.

[7] Emanuel Knill and Raymond Laflamme. "Theory of quantum error-correcting codes". In: *Phys. Rev. A* 55 (2 1997), pp. 900–911. DOI: 10.1103/PhysRevA.55.900. URL: https://link.aps.org/doi/10.1103/PhysRevA.55.900.

[8] Anna Sanpera, Rolf Tarrach, and Guifré Vidal. "Local description of quantum inseparability". In: *Physical Review A* 58.2 (1998), pp. 826–830. ISSN: 1094-1622. DOI: 10.1103/physreva.58.826. URL: http://dx.doi.org/10.1103/PhysRevA.58.826.

[9] Karol Życzkowski et al. "Volume of the set of separable states". In: *Phys. Rev. A* 58 (2 1998), pp. 883–892. DOI: 10.1103/PhysRevA.58.883. URL: https://link.aps.org/doi/10.1103/PhysRevA.58.883.

[10] Karol Zyczkowski and Hans-Jürgen Sommers. "Induced measures in the space of mixed quantum states". In: *Journal of Physics A: Mathematical and General* 34.35 (2001), pp. 7111–7125. ISSN: 1361-6447. DOI: 10.1088/0305-4470/34/35/335. URL: http://dx.doi.org/10.1088/0305-4470/34/35/335.