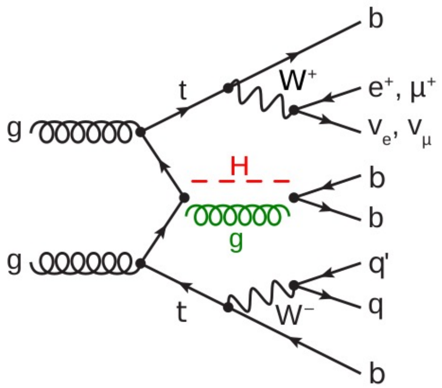
Quantum Classifiers for High Energy Physics

Searching for New Physics at the Quantum Technology Frontier
20-21 January 2022, ETH Zurich

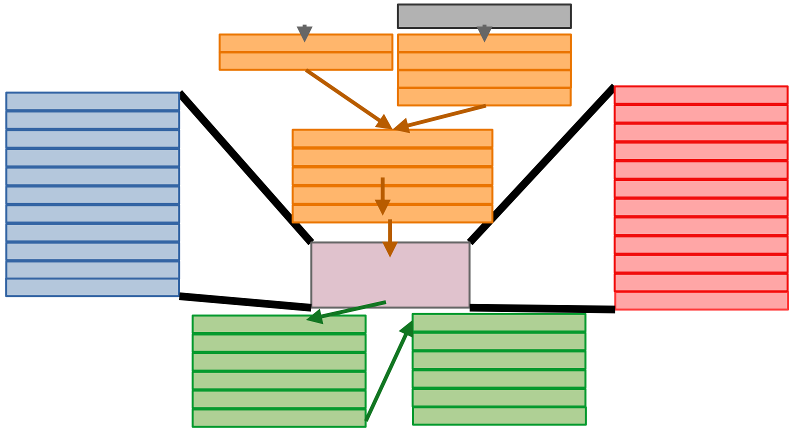
Vasilis Belis
Collaborators: P. Odagiu, S. Gonzalez, C. Reissel, S. Vallecorsa, E. Combarro, F. Reiter, G. Dissertori

Based on: *Higgs analysis with quantum classifiers,* EPJ Web Conf., 251 (2021) 03070, https://doi.org/10.1051/epjconf/202125103070, pre-print: arXiv:2104.07692.

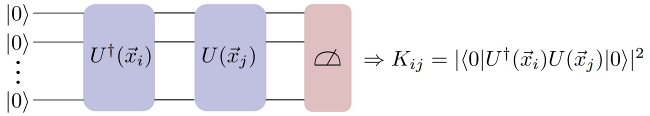
OVERVIEW



The Classification Task

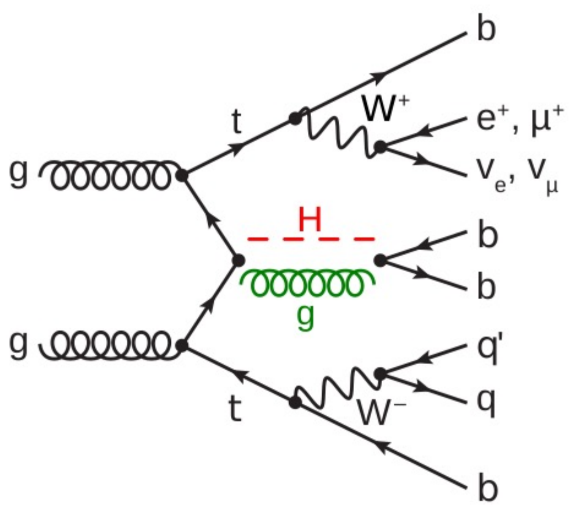


Autoencoders



Quantum Classification

The Studied Process

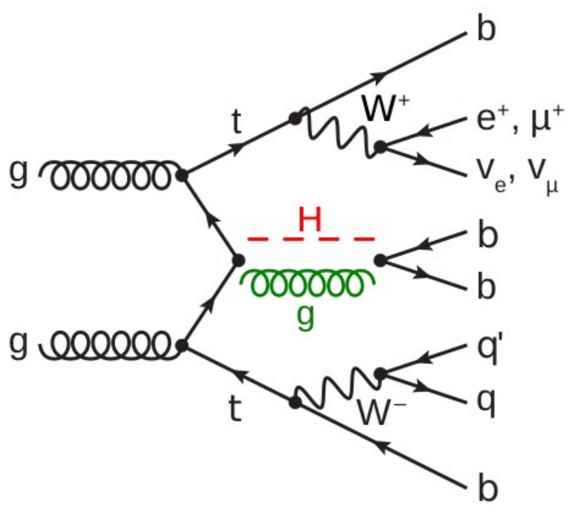


The studied $t\bar{t}H(b\bar{b})$ processes at leading order, including both *signal* and *background.*
This channel is called semi-leptonic since only one of the *W* bosons decays into leptons.

Why study the ttH(bb) process at the LHC?

ttH Yukawa coupling carries information about the scale of new physics (BS15) in a purely fermionic process.

The Studied Process



The studied $t\bar{t}H(b\bar{b})$ processes at leading order, including both *signal* and *background.*
This channel is called semi-leptonic since only one of the *W* bosons decays into leptons.

The physical observables

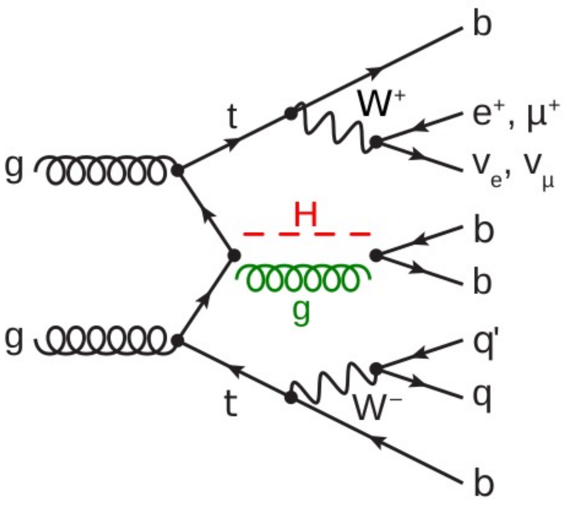
Jets

- Quark production signature: *jet production* (QCD).
- Extra observable: *b-tag*, the probability that a jet comes from a b-quark.

Why study the ttH(bb) process at the LHC?

ttH Yukawa coupling carries information about the scale of new physics (BS15) in a purely fermionic process.

The Studied Process



The studied $\mathrm{t\bar{t}H(b\bar{b})}$ processes at leading order, including both *signal* and *background*.
This channel is called semi-leptonic since only one of the *W* bosons decays into leptons.

The physical observables

Jets

- Quark production signature: *jet production* (QCD).
- Extra observable: *b-tag*, the probability that a jet comes from a b-quark.

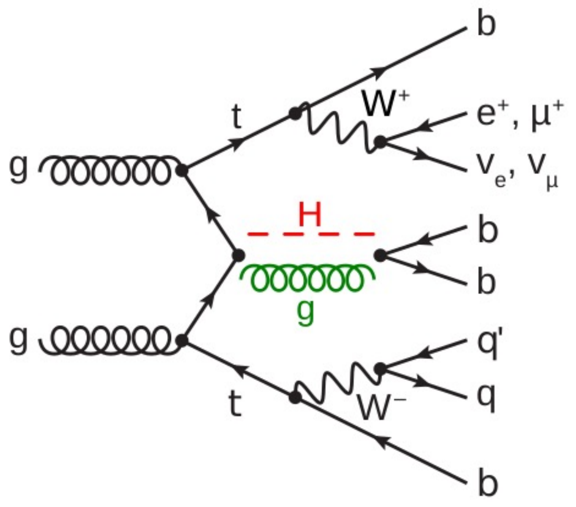
Leptons

- Lepton four-momentum for electrons and muons.
- Neutrinos cannot be detected at the LHC: missing transverse energy and momentum.

Why study the ttH(bb) process at the LHC?

ttH Yukawa coupling carries information about the scale of new physics (BS15) in a purely fermionic process.

The Studied Process



The studied $\mathrm{t\bar{t}H(b\bar{b})}$ processes at leading order, including both *signal* and *background*.
This channel is called semi-leptonic since only one of the *W* bosons decays into leptons.

Why study the ttH(bb) process at the LHC?

ttH Yukawa coupling carries information about the scale of new physics [BS15] in a purely fermionic process.

The physical observables

Jets

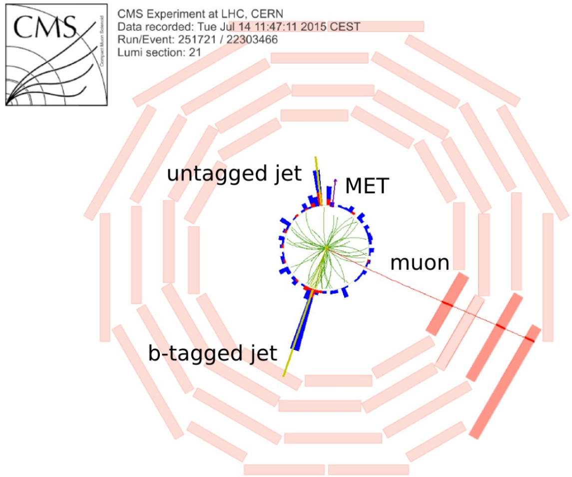
- Quark production signature: *jet production* (QCD).
- Extra observable: *b-tag*, the probability that a jet comes from a b-quark.

Leptons

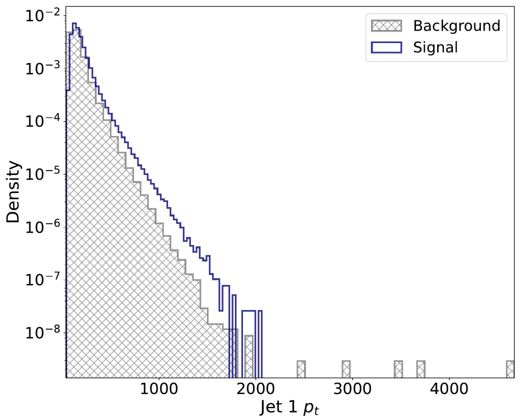
- Lepton four-momentum for electrons and muons.
- Neutrinos cannot be detected at the LHC: missing transverse energy and momentum.

$$n^{\text{features}} = \underbrace{7 \times 8}_{\text{jets}} + \underbrace{1 \times 7}_{\text{lepton}} + \underbrace{1 \times 4}_{\text{MET}} = 67$$

Workflow



compute prob.
distributions



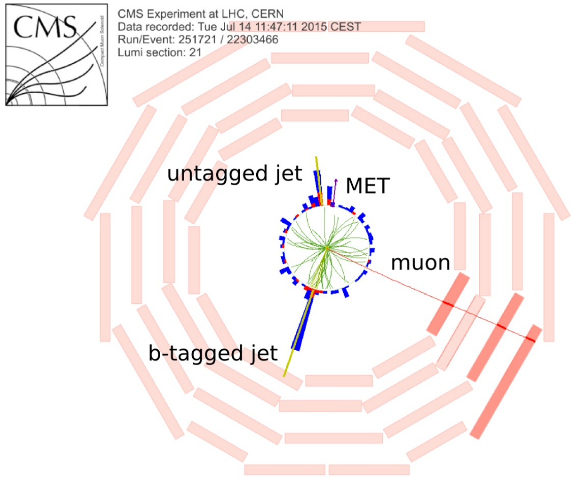
...

MC Simulation

Pure semi-leptonic channel for the ttbb and ttHbb processes.

Workflow

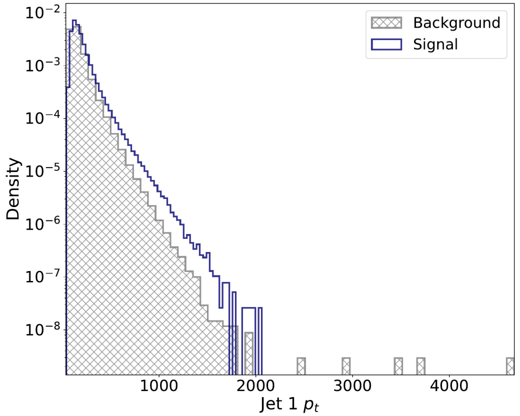
Systematic survey for several normalisation schemes: minmax determined best normalisation for our study.



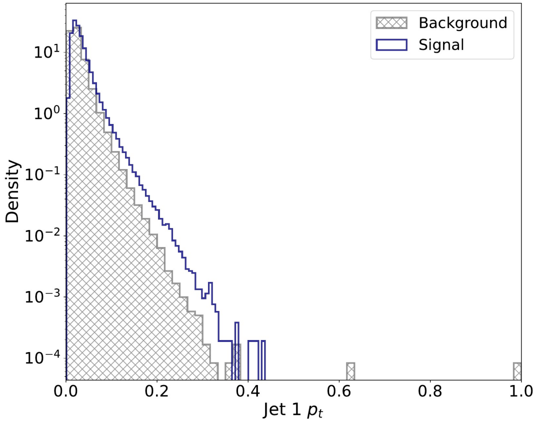
MC Simulation

Pure semi-leptonic channel for the ttbb and ttHbb processes.

compute prob. distributions



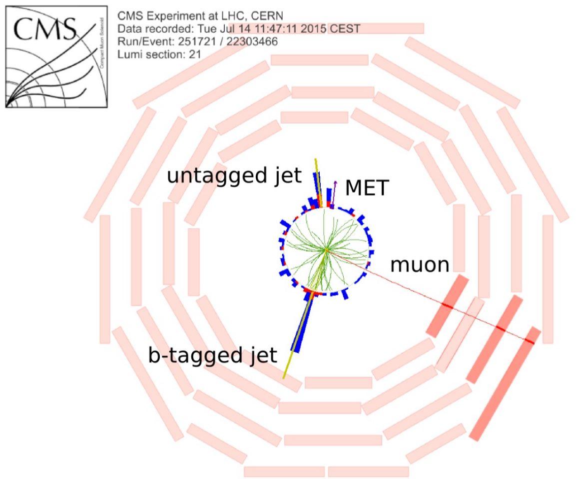
apply cuts and normalise



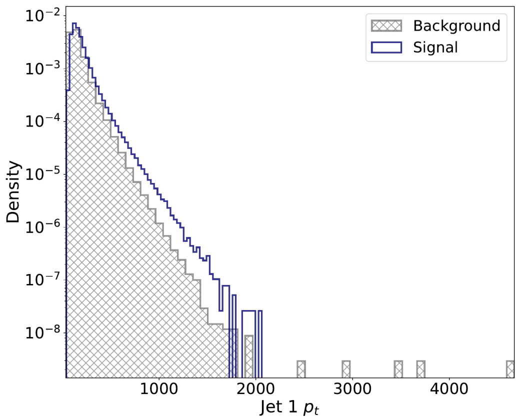
Input for conventional classifiers and our AEs

Workflow

Systematic survey for several normalisation schemes: minmax determined best normalisation for our study.

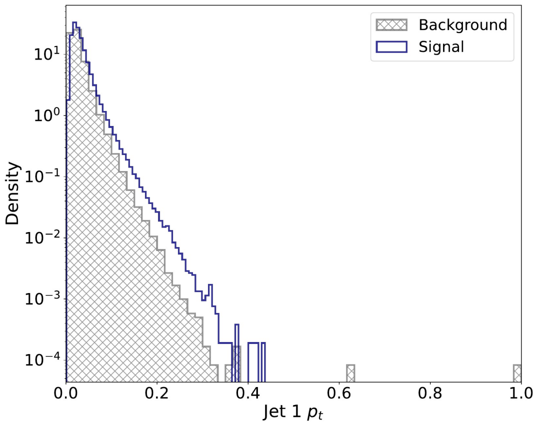


compute prob. distributions →



•••

apply cuts and normalise →



•••

MC Simulation

Pure semi-leptonic channel for the ttbb and ttHbb processes.

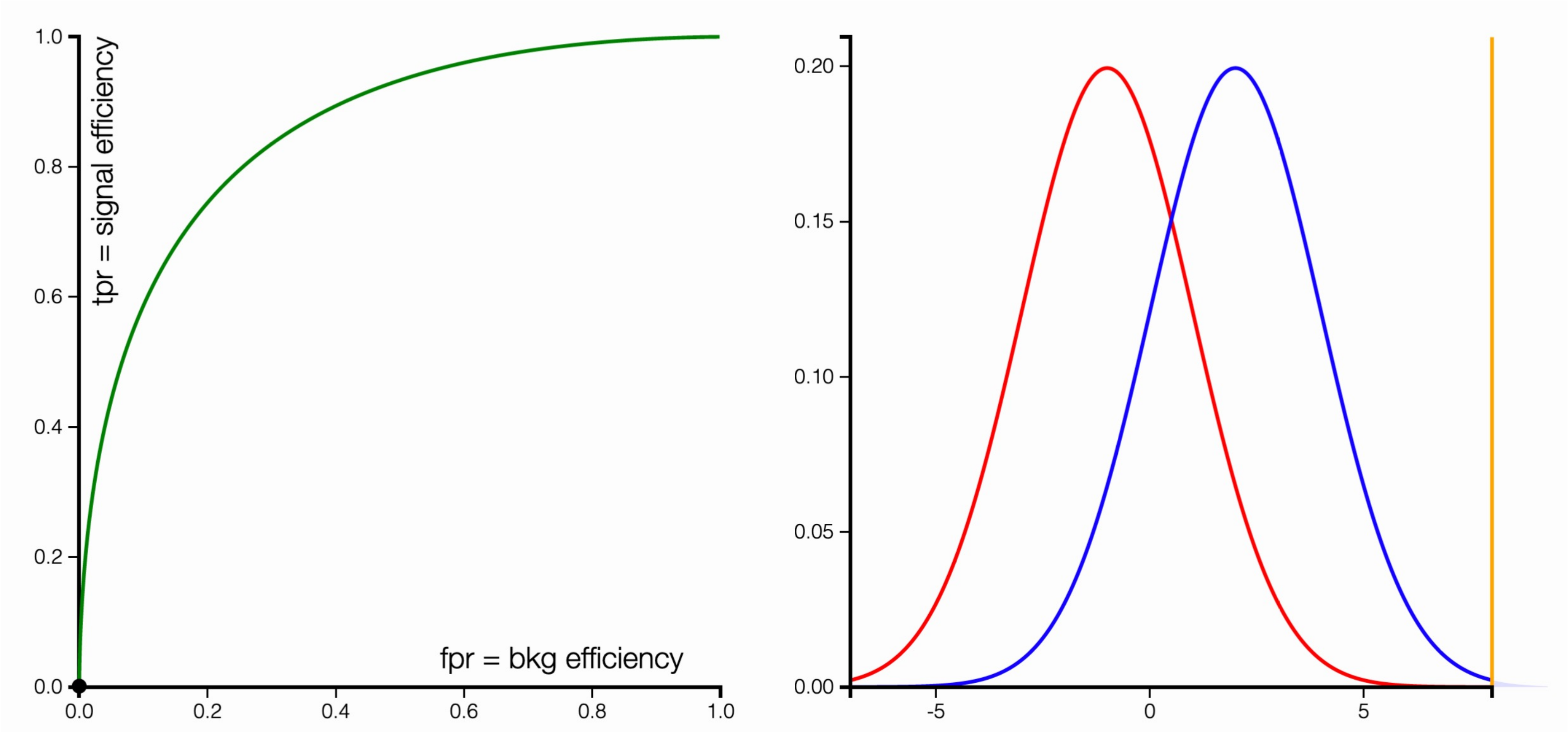
Input for conventional classifiers and our AEs

Classifier models use the normalised data to produce a *test statistic*:

- Conventional ML models: Boosted Decision Trees (BDTs), Deep Neural Networks (NNs) exploiting all input feature correlations [ATL20, CMS19] | due to NISQ device limitaitons we only use 16 out of the 67 variables.
- State-of-the-art approaches for ttH(bb): graph and attention networks, etc. [C.Reissel@ML4Jets] : 0.74 -0.76 AUC.

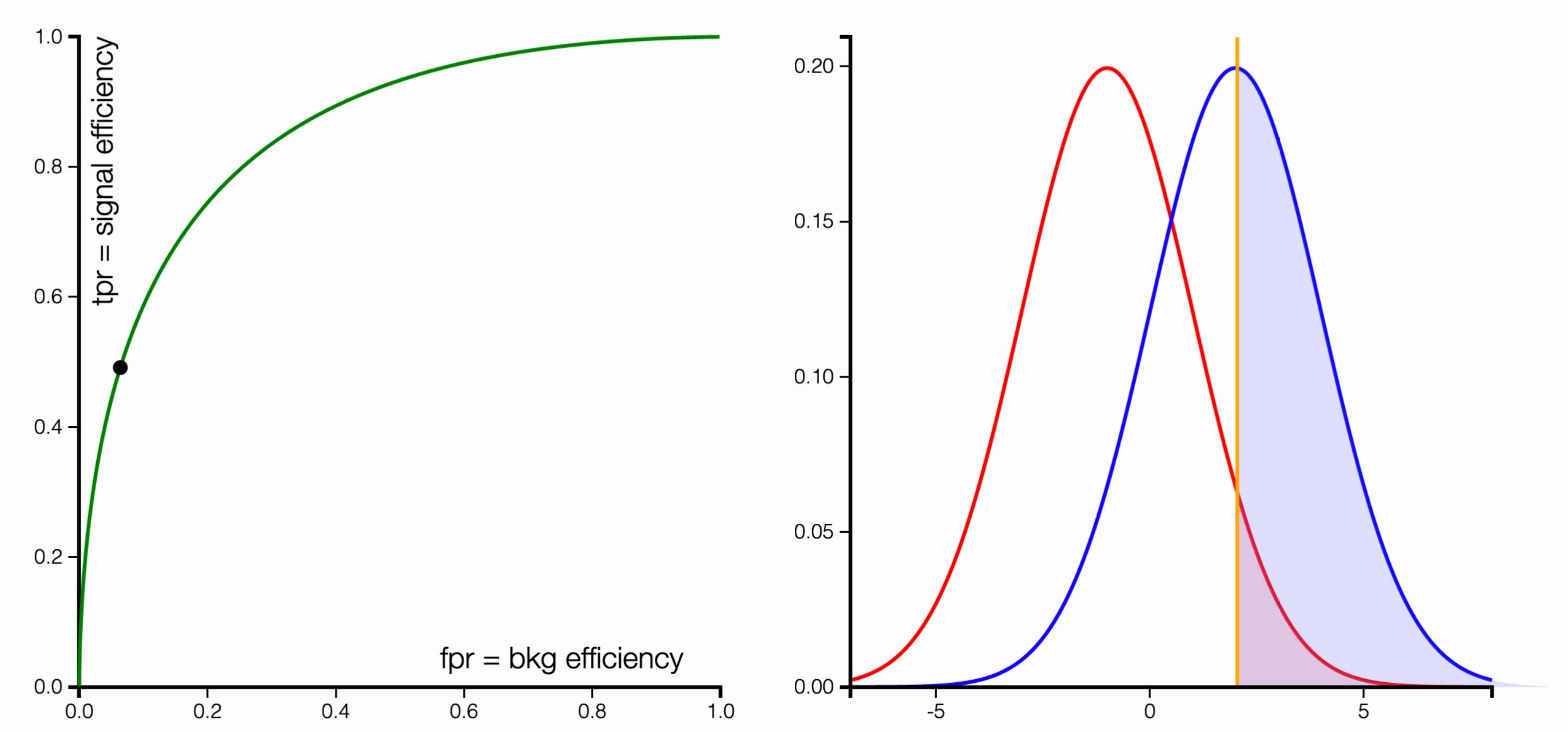
Performance Metric

- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.



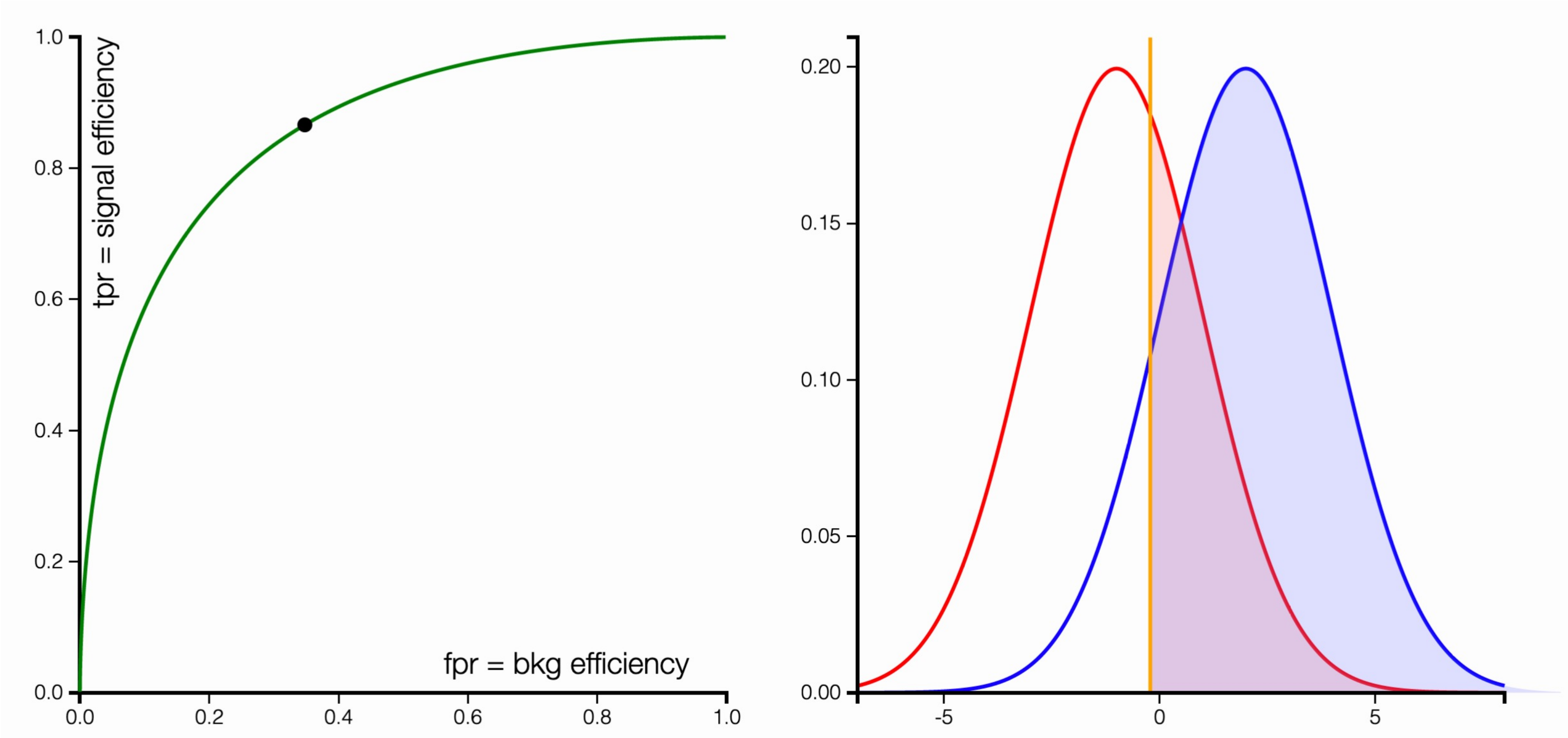
Performance Metric

- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.



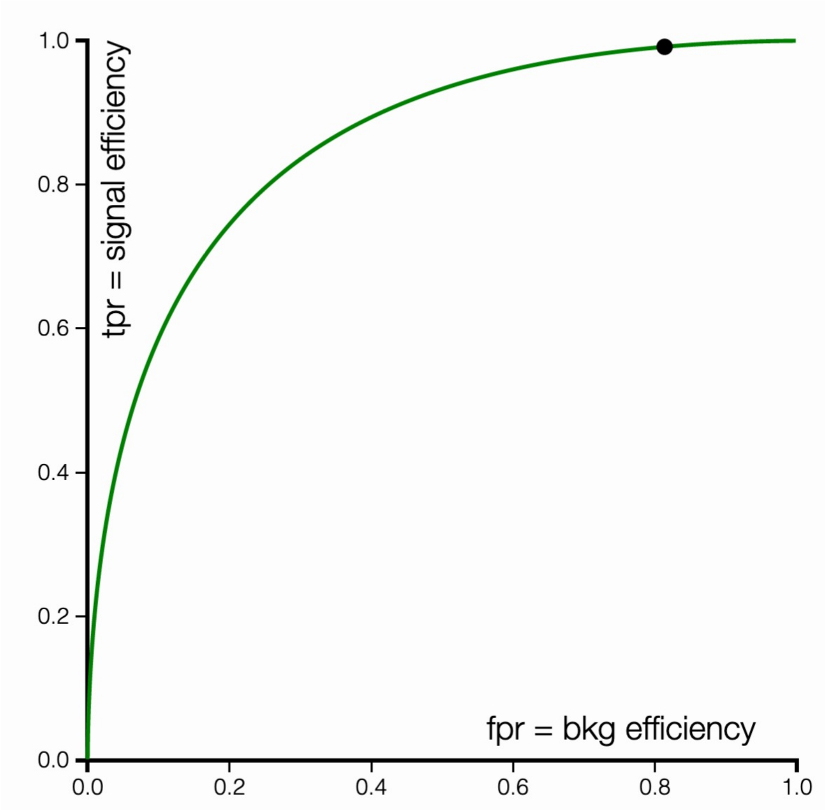
Performance Metric

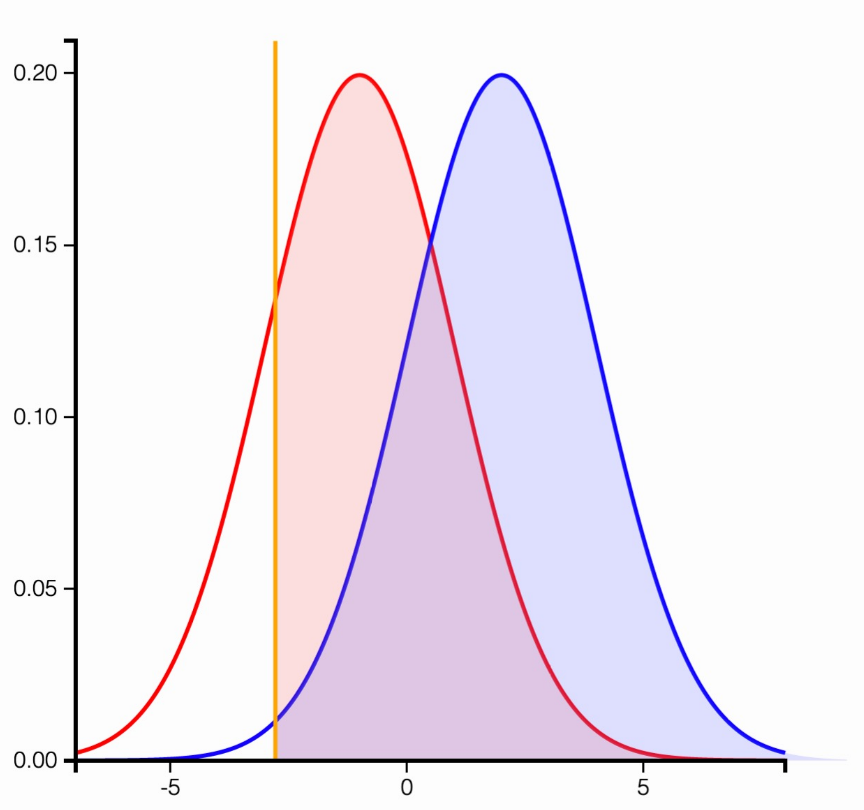
- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.



Performance Metric

- The normalised data samples are split into training, validation, and testing data sets.
- Classification power metric: Receiver Operating Characteristic (ROC) curve.
- More compact metric: Area Under Curve (AUC) of the ROC curve.





Motivation

Why quantum machine learning for HEP?

- Heuristic answer: investigate the new set of ML techniques and methods available and assess advantages.

Motivation

Why quantum machine learning for HEP?

- Heuristic answer: investigate the new set of ML techniques and methods available and assess advantages.

- Fundamental motivation: can quantum models utilise the quantum correlations inherent in HEP data leading to performance advantages?

  - Goal in "Statistics/ML jargon" [KBS21]: Find inductive bias based on prior knowledge on the data generation (quantum process for HEP data).

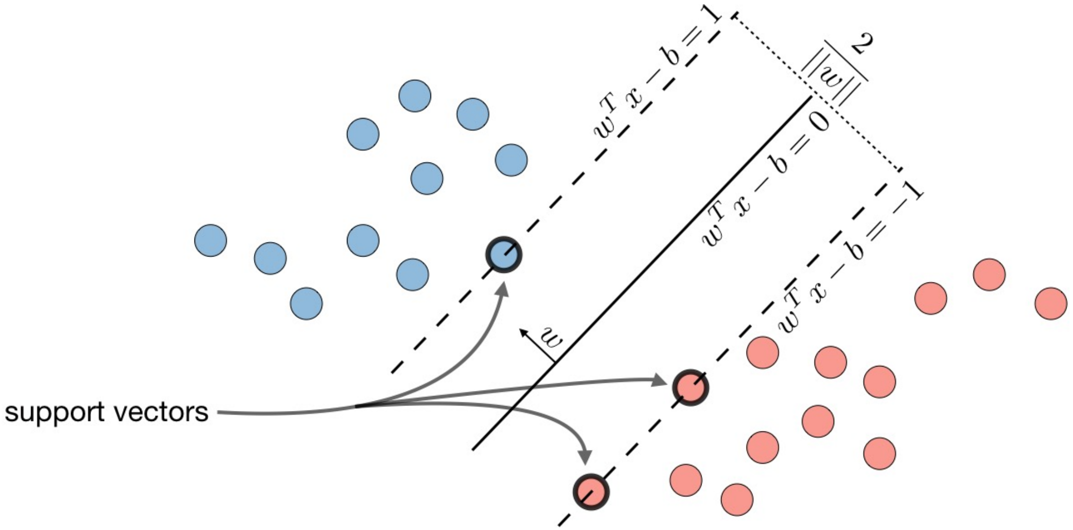
  - If the bias can be constructed and is classically difficult to simulate → quantum advantage.

Motivation

Why quantum machine learning for HEP?

- Heuristic answer: investigate the new set of ML techniques and methods available and assess advantages.
- Fundamental motivation: can quantum models utilise the quantum correlations inherent in HEP data leading to performance advantages?
  - Goal in “Statistics/ML jargon” [KBS21]: Find inductive bias based on prior knowledge on the data generation (quantum process for HEP data).
  - If the bias can be constructed and is classically difficult to simulate → quantum advantage.
- Example: quantum algorithm for HEP event shower simulation, produces accurate results [NPdJB21]. Can simulate naturally the interference diagram.

Support Vector Machines



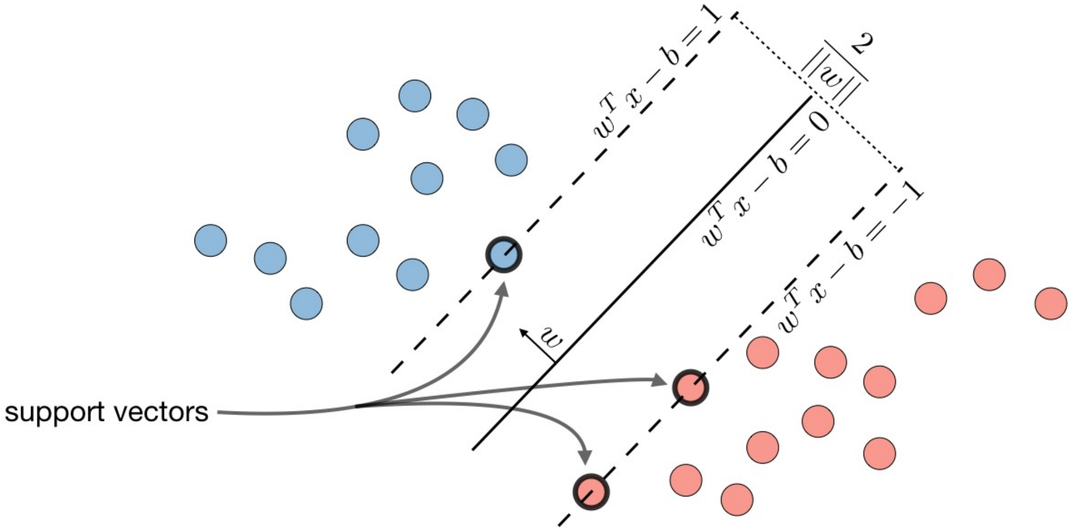
SVM objective function is equivalent to (dual Lagrangian):

maximize $$L(c_1 \dots c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i c_i (\vec{x}_i \cdot \vec{x}_j) y_j c_j$$

subject to $$\sum_{i=1}^{n} c_i y_i = 0, \text{ and } 0 \leq c_i \leq C \text{ for all } i.$$

Kernel trick: $$(\vec{x}_i \cdot \vec{x}_j) \mapsto k(\vec{x}_i, \vec{x}_j) := \phi(\vec{x}_i) \cdot \phi(\vec{x}_j).$$

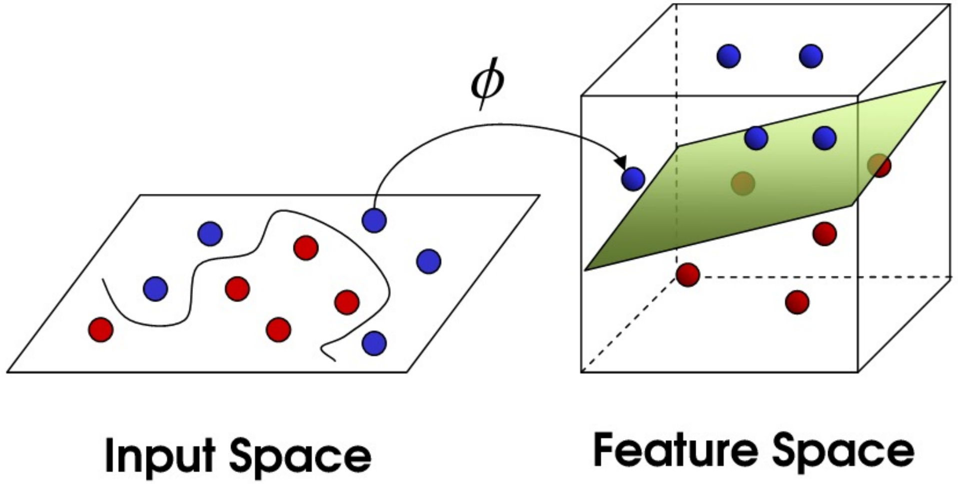
Support Vector Machines



SVM objective function is equivalent to (dual Lagrangian):

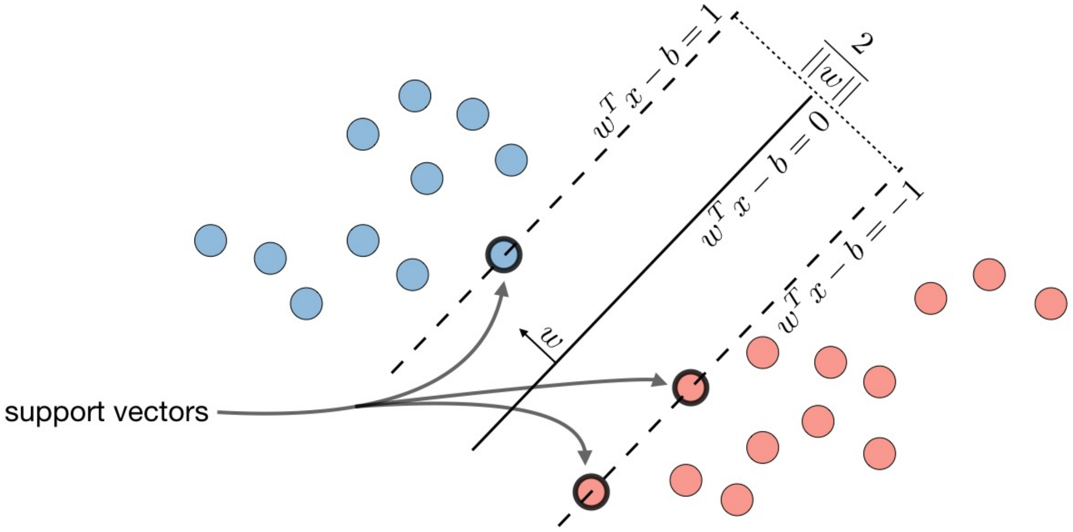
maximize $$L(c_1 \dots c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i (\vec{x}_i \cdot \vec{x}_j) y_j c_j$$

subject to $$\sum_{i=1}^{n} c_i y_i = 0, \text{ and } 0 \leq c_i \leq C \text{ for all } i.$$



Kernel trick: $$(\vec{x}_i \cdot \vec{x}_j) \mapsto k(\vec{x}_i, \vec{x}_j) := \phi(\vec{x}_i) \cdot \phi(\vec{x}_j).$$

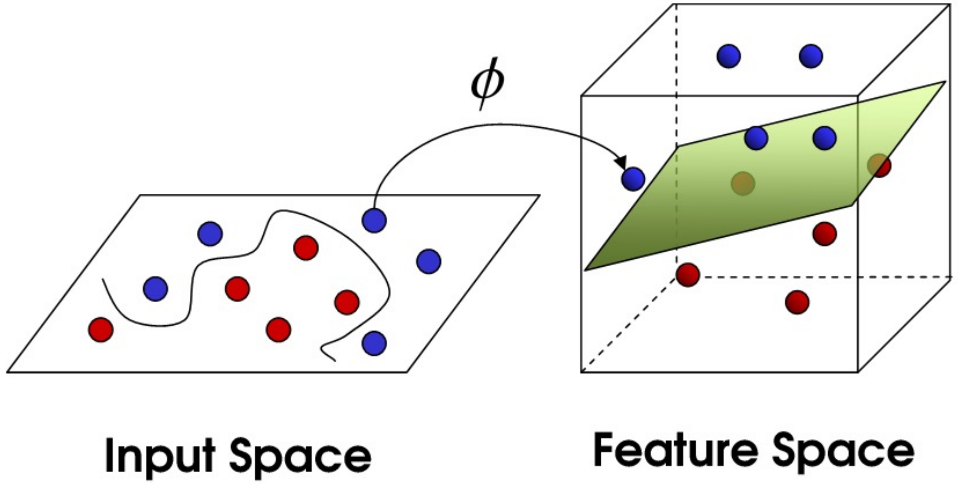
Support Vector Machines



SVM objective function is equivalent to (dual Lagrangian):

maximize $$L(c_1 \dots c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i c_i (\vec{x}_i \cdot \vec{x}_j) y_j c_j$$

subject to $$\sum_{i=1}^{n} c_i y_i = 0, \text{ and } 0 \le c_i \le C \text{ for all } i.$$



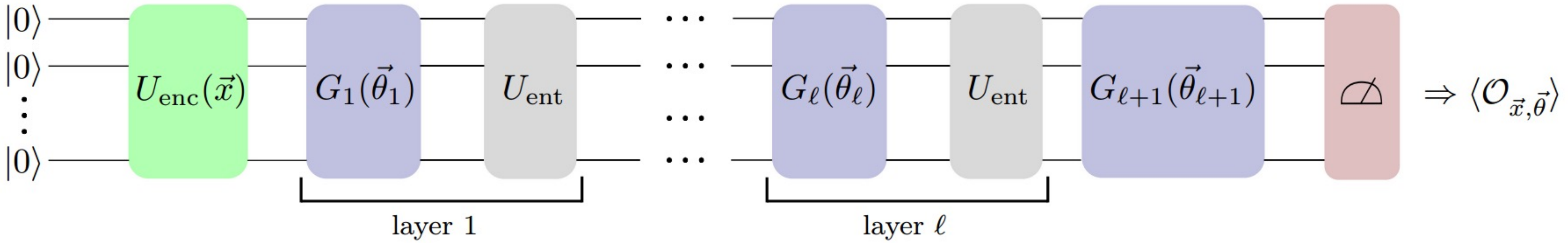
Kernel trick: $(\vec{x}_i \cdot \vec{x}_j) \mapsto k(\vec{x}_i, \vec{x}_j) := \phi(\vec{x}_i) \cdot \phi(\vec{x}_j).$

Make the kernel *quantum*:

$|0\rangle$, $|0\rangle$, $\vdots$, $|0\rangle$ — $U^\dagger(\vec{x}_i)$ — $U(\vec{x}_j)$ — measurement

$$\Rightarrow K_{ij} = |\langle 0|U^\dagger(\vec{x}_i)U(\vec{x}_j)|0\rangle|^2$$

Variational Quantum Circuits



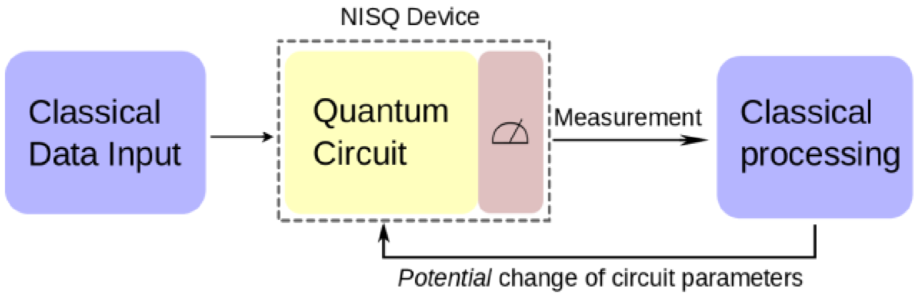
- Data embedding circuit (feature map) here is fixed.
- Layers of parametrised quantum gates → trainable parameters.
- Output of the model → expectation value of an observable on the prepared state $|\psi(\vec{x},\vec{\theta})\rangle$ e.g. measure the first qubit on the computational basis

$$\mathcal{O} = \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \cdots \otimes \mathbb{1},$$

$$f(\vec{x},\vec{\theta}) = \langle \psi(\vec{x},\vec{\theta}) | \, \mathcal{O} \, | \psi(\vec{x},\vec{\theta}) \rangle \equiv \langle \psi(\vec{x}) | \, G^\dagger(\vec{\theta}) \mathcal{O} G(\vec{\theta}) \, | \psi(\vec{x}) \rangle \equiv \langle \mathcal{O} \rangle_{\vec{x},\vec{\theta}}.$$

- Classification: if $\langle \mathcal{O} \rangle_{\vec{x},\vec{\theta}} > 0$ → signal, otherwise background.

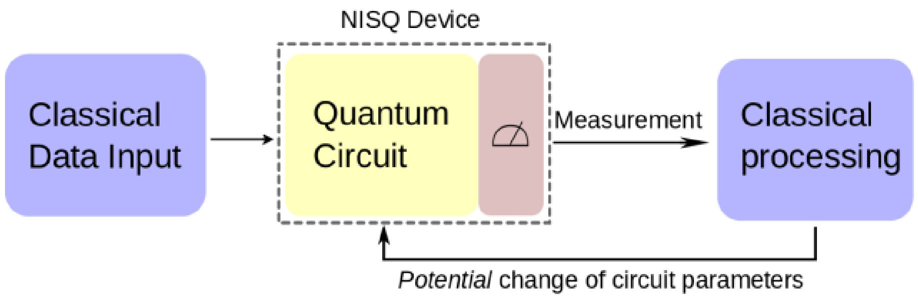
Hybrid Quantum-Classical ML Classifiers



Noisy Intermediate Scale Quantum (NISQ) devices:

- *Circuit width:* limited number of qubits (superconducting qubits at IBM up to 127).

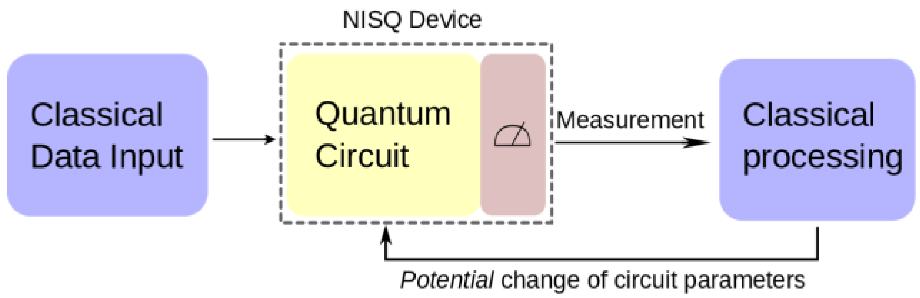
Hybrid Quantum-Classical ML Classifiers



<u>Noisy Intermediate Scale Quantum (NISQ) devices:</u>

- *Circuit width:* limited number of qubits (superconducting qubits at IBM up to 127).
- *Circuit depth:* limited number of operations per qubit (small decoherence times).

Hybrid Quantum-Classical ML Classifiers



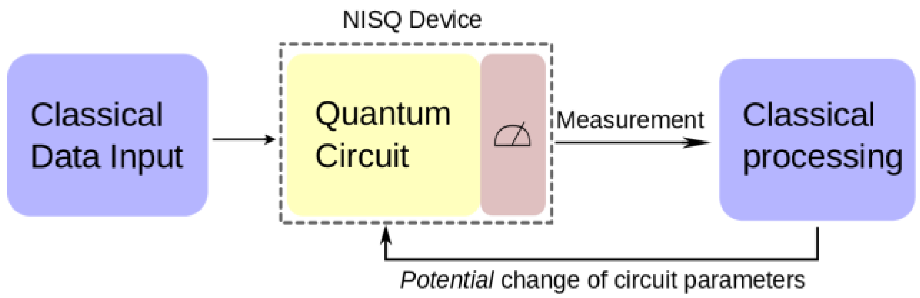
Noisy Intermediate Scale Quantum (NISQ) devices:

- *Circuit width:* limited number of qubits (superconducting qubits at IBM up to 127).
- *Circuit depth:* limited number of operations per qubit (small decoherence times).

QML models for classification:

- Kernel methods: Quantum Support Vector Machine (QSVM).
- Quantum "Neural Networks": Variational/Parametrized Quantum Circuits (VQC/PQC).

Hybrid Quantum-Classical ML Classifiers



Noisy Intermediate Scale Quantum (NISQ) devices:

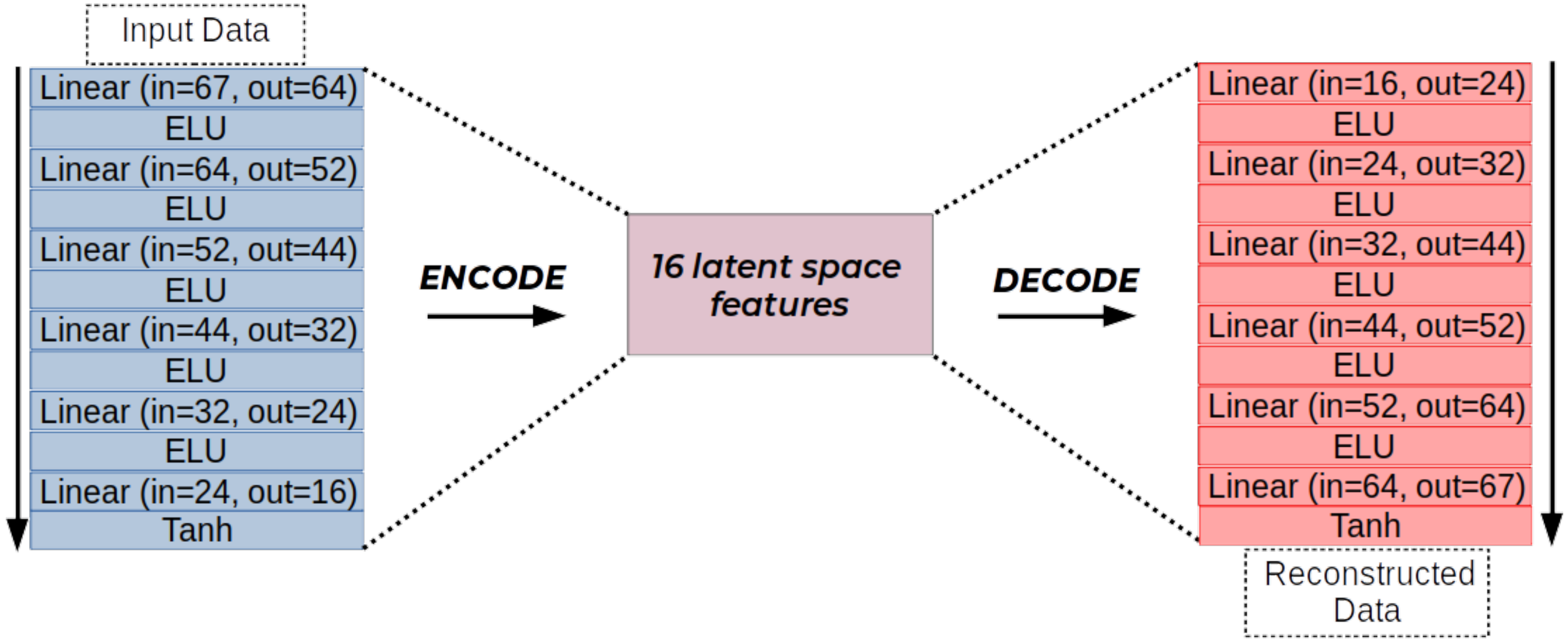
- *Circuit width:* limited number of qubits (superconducting qubits at IBM up to 127).
- *Circuit depth:* limited number of operations per qubit (small decoherence times).

QML models for classification:

- Kernel methods: Quantum Support Vector Machine (QSVM).
- Quantum "Neural Networks": Variational/Parametrized Quantum Circuits (VQC/PQC).

To accommodate for NISQ limitations, feature reduction is needed.

The Vanilla AE

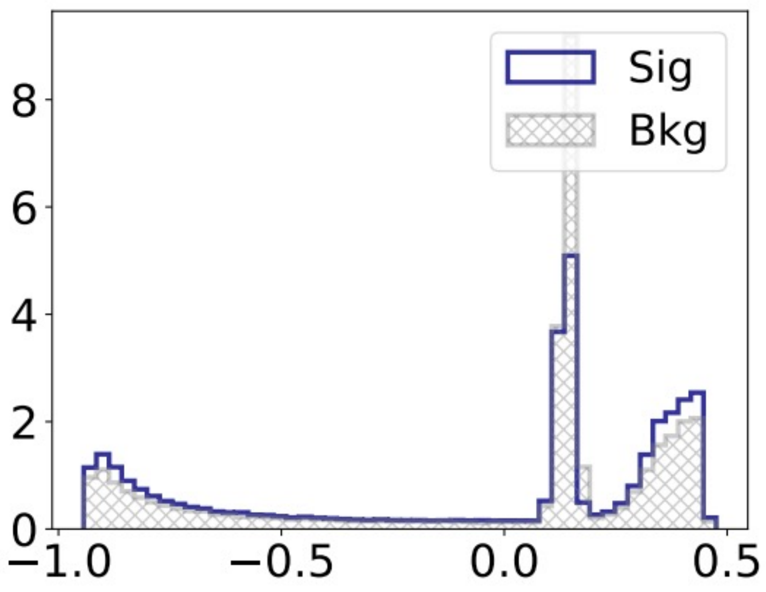


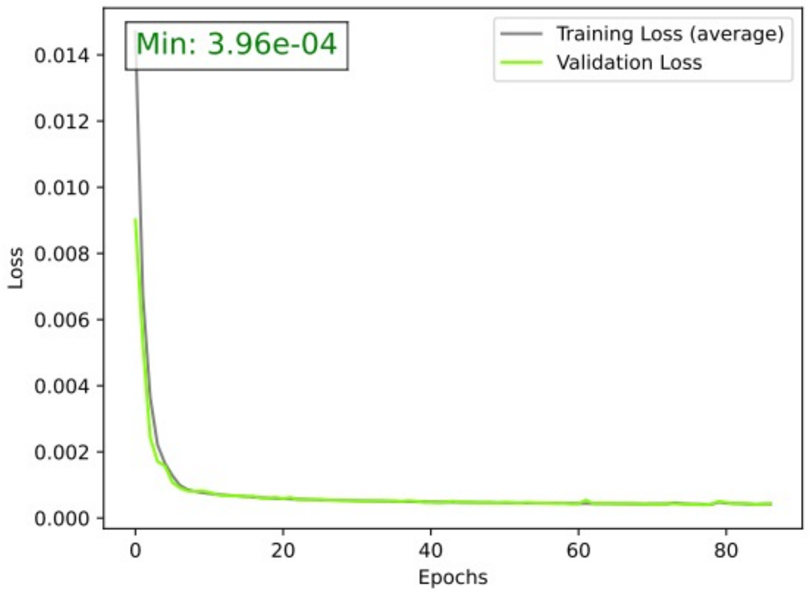
Loss function: Mean Squared Error (MSE) between the input data and the reconstructed data.

$$L_{\mathrm{MSE}} = (y - f(x, \theta))^2$$

The learning rate and the batch size were optimised for minimum MSE loss, yielding 0.0012 for the learning rate with 128 events per batch.

The Vanilla AE



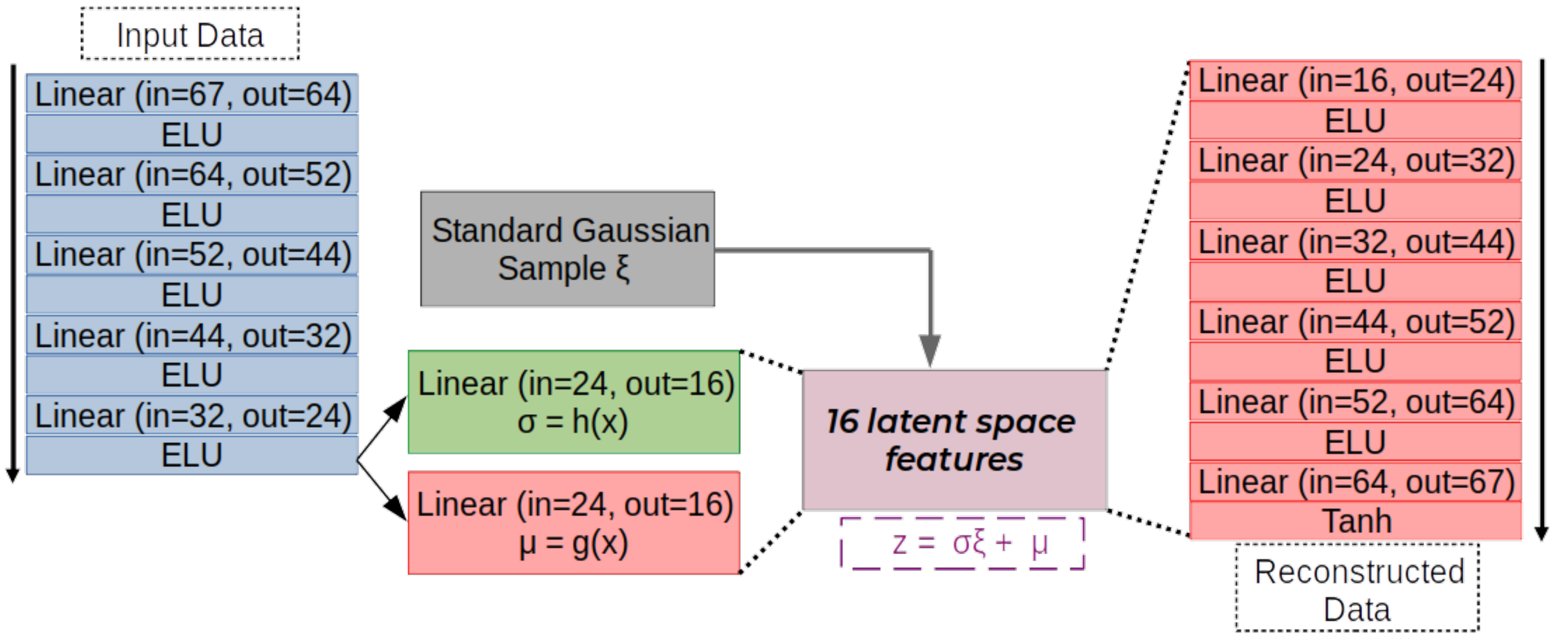


Notable properties: *irregular latent space, tendency to overtrain*.

The loss obtained in this model shows a two fold improvement compared to the standard AE used in the QSVM study at arXiv:2104.07692 with a loss of

$$\mathrm{L}_{\mathrm{MSE}} = 4.77 \pm 10^{-4}$$

The Variational AE

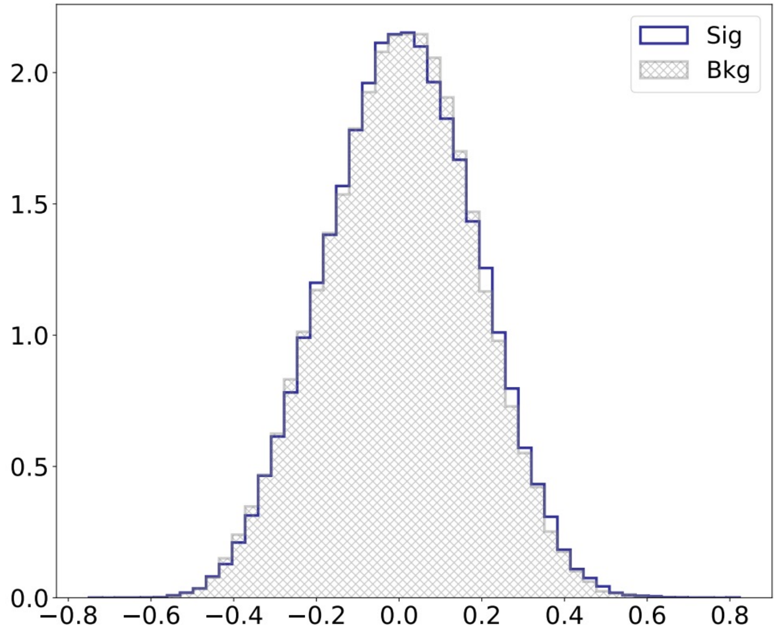


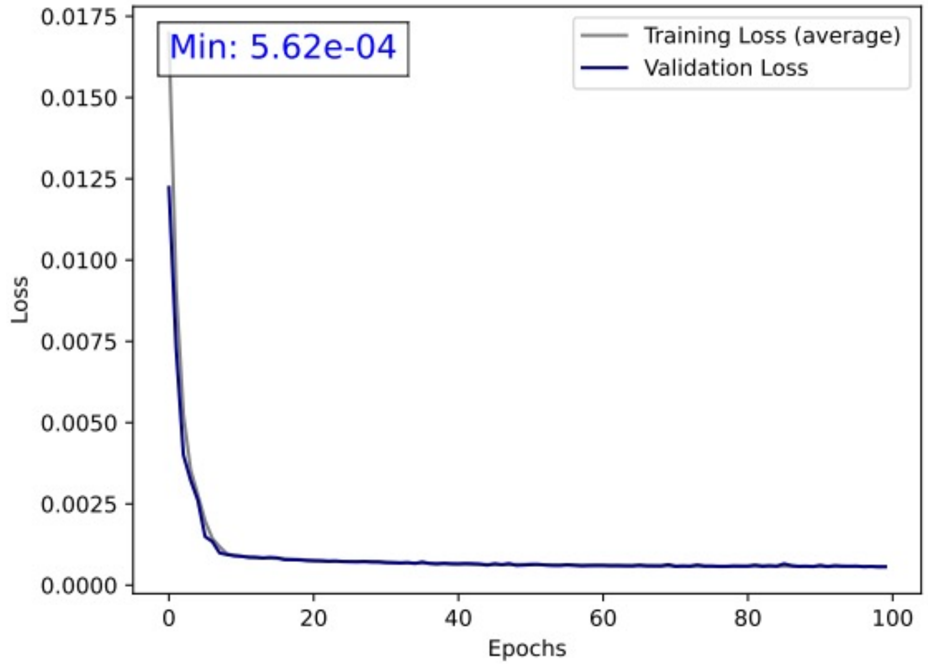
Loss function: Mean Squared Error (MSE) and the *KL Divergence*.

$$L_{VAE} = (1-\alpha)L_{MSE} + \alpha\mathcal{D}_{KL}\left(\mathcal{N}(\mu,\sigma),\, \mathcal{N}(0,I)\right) \qquad \mathcal{D}_{KL} = q(x)[\log\left(q(x)\right) - \log\left(p(x)\right)]$$

The learning rate and the batch size were optimised for minimum overall loss, yielding 0.001 for the learning rate with 128 events per batch, while α=0.5.

The Variational AE

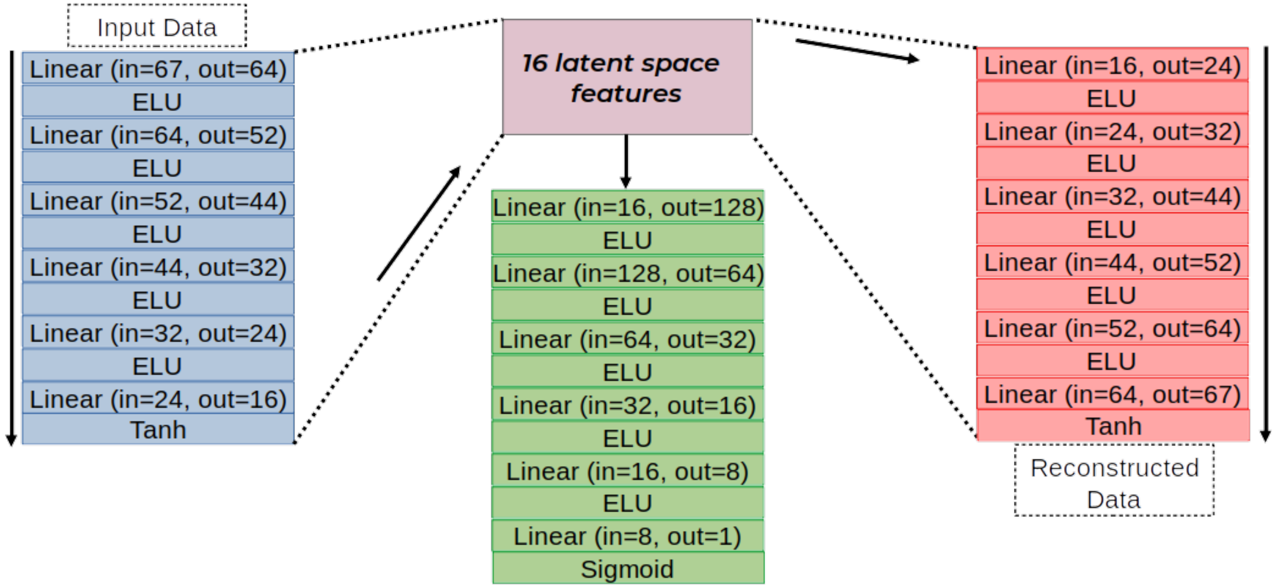




Further, the value of the weight was fine-tuned as well for the best raw reconstruction loss (MSE), giving **α=**0.0005.

$$L_{\mathrm{MSE}} = 4.49 \times 10^{-4}$$

The Classifier AE

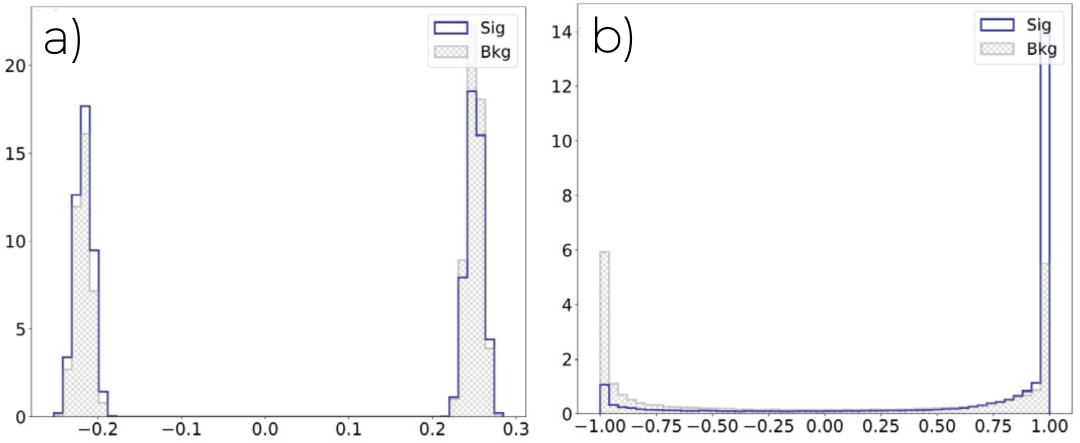


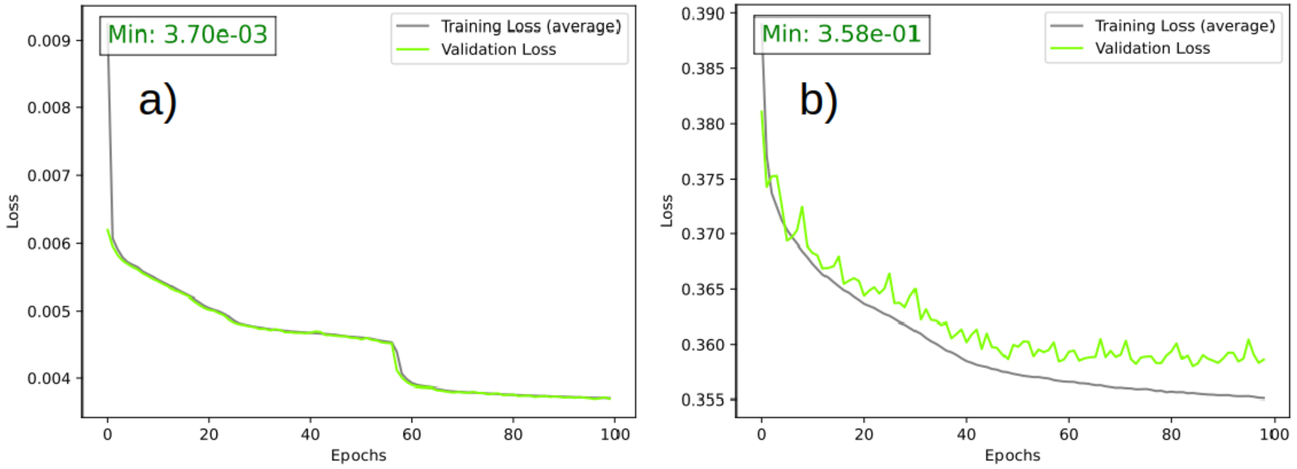
Loss function: Mean Squared Error (MSE) and the *Binary Cross Entropy (BCE)*.

$$L_{\mathrm{CAE}} = (1-\alpha)L_{\mathrm{MSE}} + \alpha L_{\mathrm{BCE}} \qquad L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\left(p(y_i)\right) - (1-y_i)\log\left(1-p(y_i)\right)$$

The learning rate and the batch size were optimised for minimum overall loss, yielding 0.001 for the learning rate with 128 events per batch, while α=0.5.

The Classifier AE





Again, the latent space is irregular and prone to overtraining.

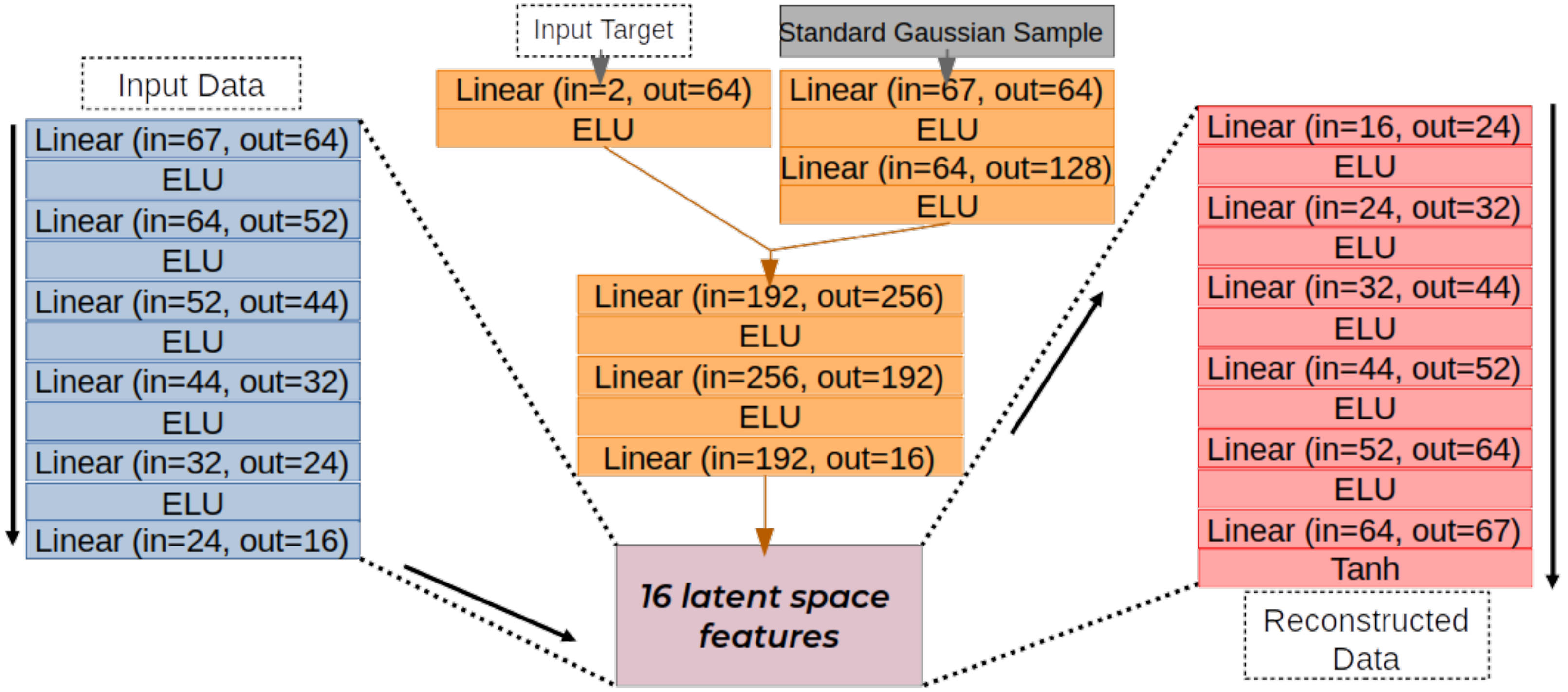
Further, α was fine-tuned as well for:

- the best raw reconstruction loss (MSE): a) $\alpha=3\text{x}10^{-5}$.
- the best unweighted classification loss (BCE): b) $\alpha=0.6$.

a) $L_{MSE} = 5.47\times10^{-4}$ $L_{BCE} = 0.63$

b) $L_{MSE} = 62.97\times10^{-4}$ $L_{BCE} = 0.61$

The Sinkhorn AE



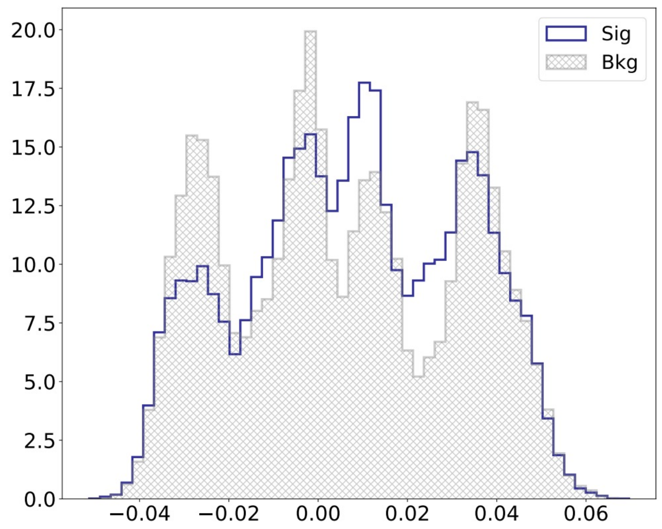
Loss function: Mean Squared Error (MSE) and the *Sinkhorn Loss*.

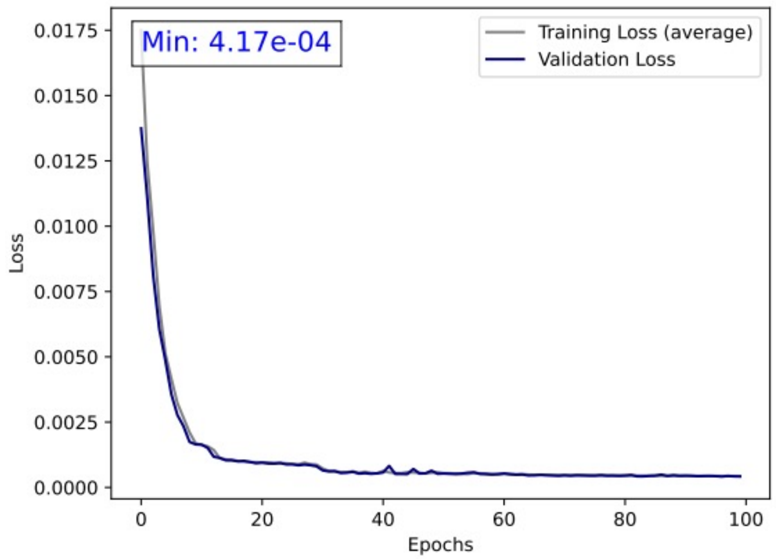
$$L_{SAE} = (1-\alpha)L_{MSE} + \alpha L_{SH}$$

$$W_c(q,p) = \inf_{\Gamma \in \Pi(q,p)} \int\int c(x,y)\Gamma(x,y)\,dx\,dy$$

The hyperparameters are optimised (α=0.5) yielding: learning rate of 0.001 with batch size 128.

The Sinkhorn AE



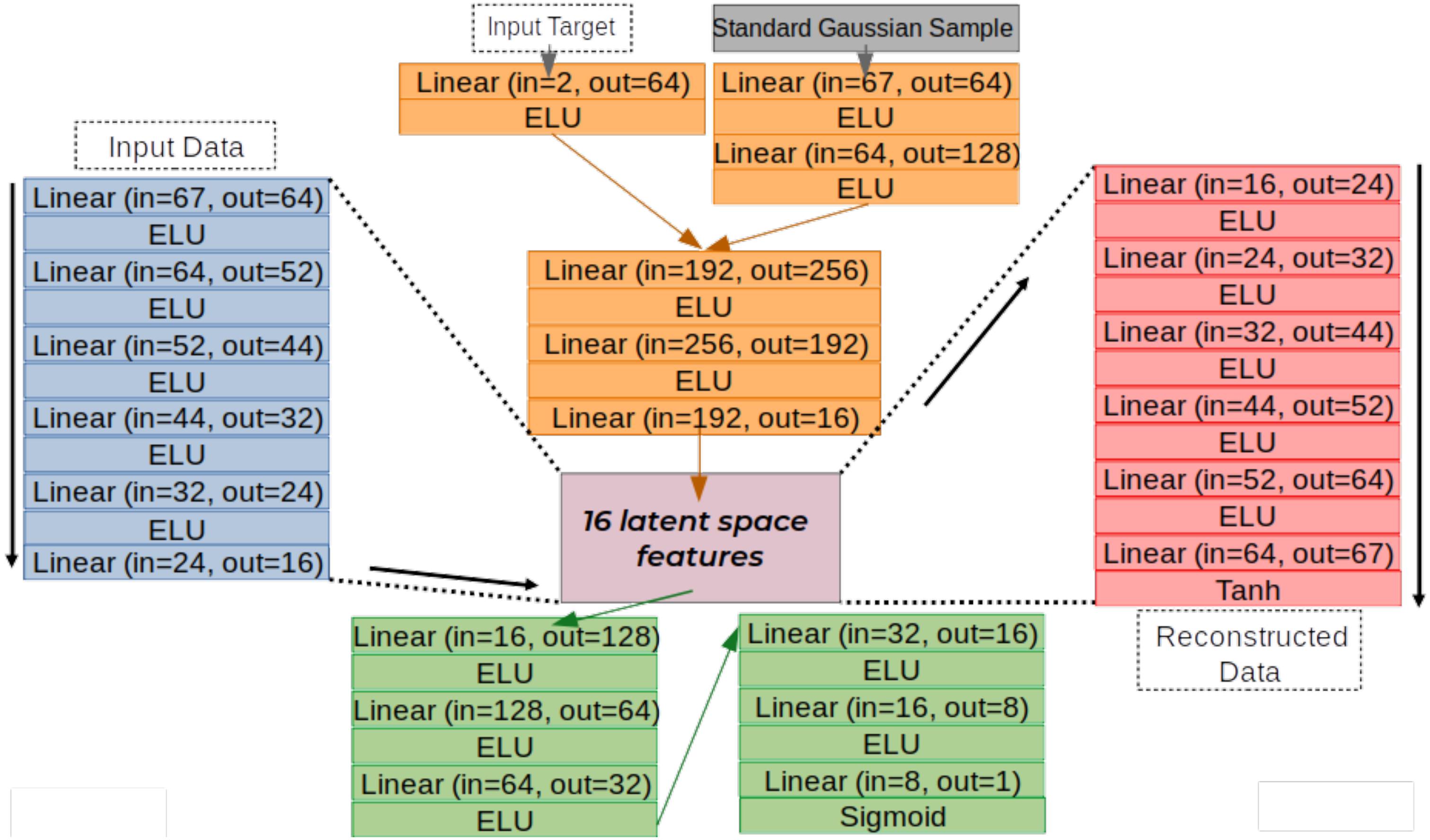


The latent space is *regularised* but allows for divergences from a strict standard normal distribution. The Sinkhorn regularisation is less strict than the variational one.

The α weight was fined tuned as well for lowest MSE, giving α=0.06 and a loss of

$$L_{MSE} = 9.65 \times 10^{-4}$$

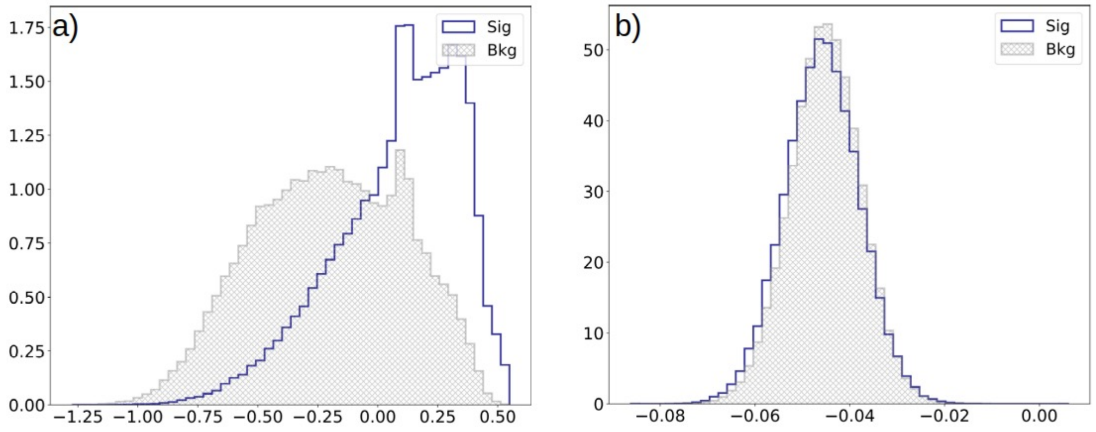
The Sinkclass AE

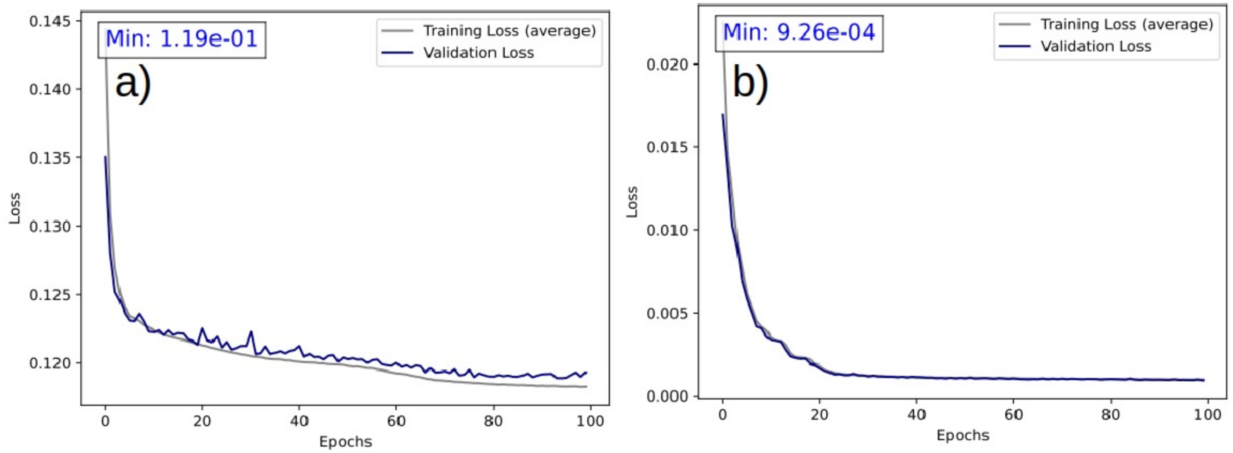


Loss function: all of them!

$$L_{SCAE} = \alpha L_{SH} + \beta L_{BCE} + L_{MSE}$$

The Sinkclass AE





The hyperparameters of the Sinkclass AE were optimised by first setting **α**=1 and β=1, yielding a learning rate of 0.001 and batch size of 128.

Then, the loss weights were optimised for
a) the lowest unweighted BCE.
b) the lowest unweighted MSE.

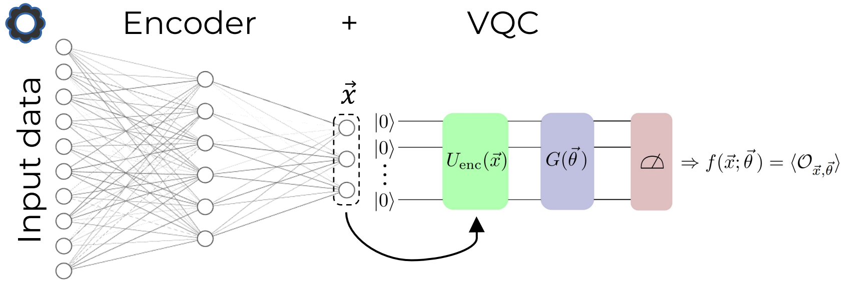
The obtained values are a) **α**=0.02 and **β**=0.2 and b) **α**=0.9 and **β**=0.0008.

a) $L_{MSE} = 26.41 \times 10^{-4}$ $\quad L_{BCE} = 0.65$

b) $L_{MSE} = 24.69 \times 10^{-4}$ $\quad L_{BCE} = 0.61$

AE, QSVM, and *Hybrid VQC

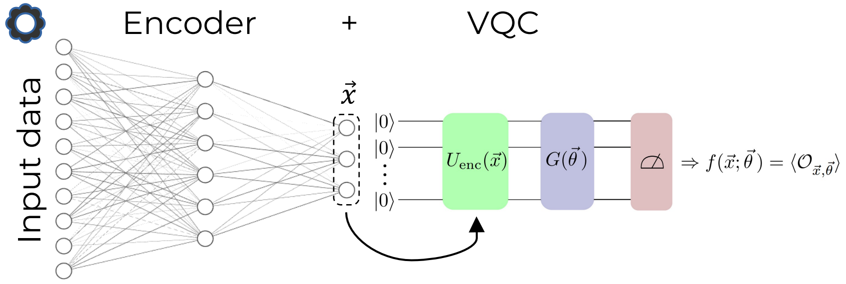
Autoencoder	HP Optimisation	MSE Loss $\times 10^{-4}$	BCE Loss	Classifier AUC	QSVM AUC
Vanilla	-	4.77	-	-	0.56 ± 0.01
Variational	MSE	4.49	-	-	0.56 ± 0.02
Classifier	MSE	5.47	0.63	0.700 ± 0.001	0.56 ± 0.02
	BCE	62.97	0.61	0.734 ± 0.002	0.72 ± 0.01
Sinkhorn	MSE	9.65	-	-	0.51 ± 0.01
Sinkclass	MSE	26.41	0.65	0.642 ± 0.003	0.50 ± 0.01
	BCE	24.69	0.61	0.734 ± 0.002	0.74 ± 0.01



Model	BCE Loss	AUC
Encoder + VQC	0.61	0.702 ± 0.004

AE, QSVM, and *Hybrid VQC

Autoencoder	HP Optimisation	MSE Loss $\times 10^{-4}$	BCE Loss	Classifier AUC	QSVM AUC
Vanilla	-	4.77	-	-	0.56 ± 0.01
Variational	MSE	4.49	-	-	0.56 ± 0.02
Classifier	MSE	5.47	0.63	0.700 ± 0.001	0.56 ± 0.02
	BCE	62.97	0.61	0.734 ± 0.002	0.72 ± 0.01
Sinkhorn	MSE	9.65	-	-	0.51 ± 0.01
Sinkclass	MSE	26.41	0.65	0.642 ± 0.003	0.50 ± 0.01
	BCE	24.69	0.61	0.734 ± 0.002	0.74 ± 0.01



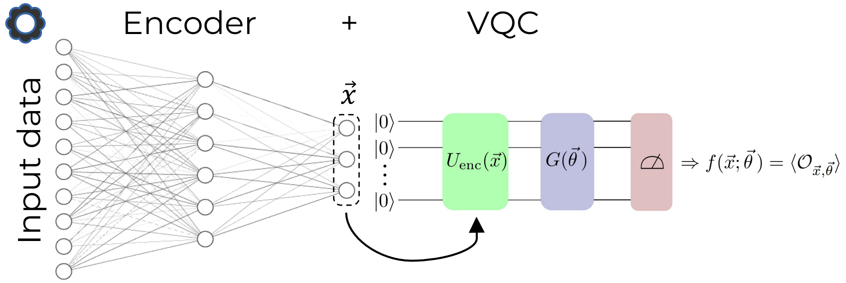
Model	BCE Loss	AUC
Encoder + VQC	0.61	0.702 ± 0.004

Sinkclass AE shows best performance when considering both *reconstruction power* and *classification power*.

It even matches the classical *state-of-the-art* result!

AE, QSVM, and *Hybrid VQC

Autoencoder	HP Optimisation	MSE Loss $\times 10^{-4}$	BCE Loss	Classifier AUC	QSVM AUC
Vanilla	-	4.77	-	-	0.56± 0.01
Variational	MSE	4.49	-	-	0.56± 0.02
Classifier	MSE	5.47	0.63	0.700± 0.001	0.56± 0.02
	BCE	62.97	0.61	0.734± 0.002	0.72± 0.01
Sinkhorn	MSE	9.65	-	-	0.51± 0.01
Sinkclass	MSE	26.41	0.65	0.642± 0.003	0.50± 0.01
	BCE	24.69	0.61	0.734± 0.002	0.74± 0.01

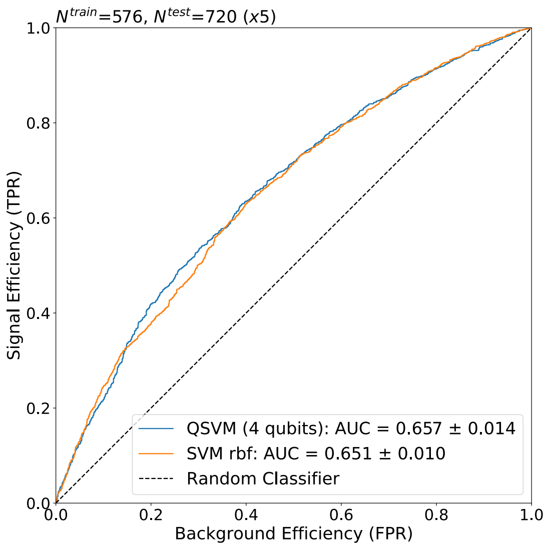


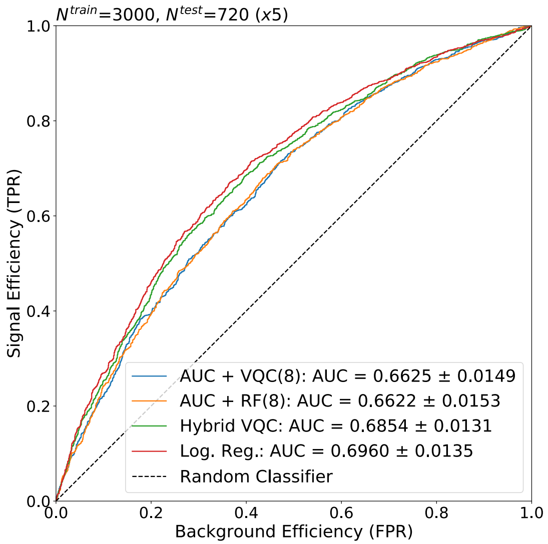
Model	BCE Loss	AUC
Encoder + VQC	0.61	0.702± 0.004

Sinkclass AE shows best performance when considering both *reconstruction power* and *classification power.*

It even matches the classical *state-of-the-art* result!

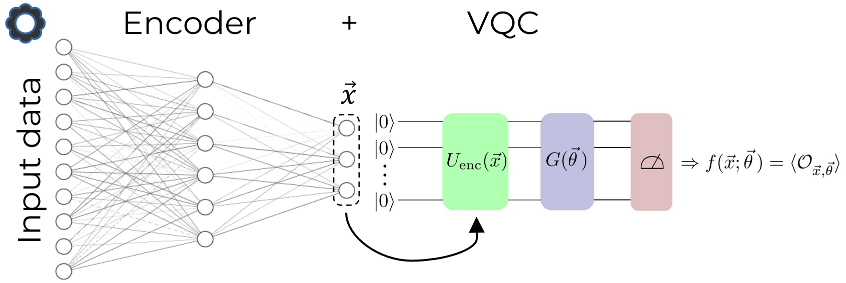
Results from: https://doi.org/10.1051/epjconf/202125103070





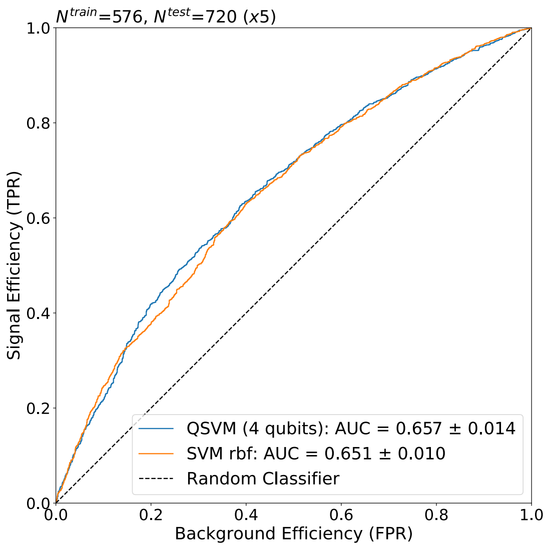
AE, QSVM, and *Hybrid VQC

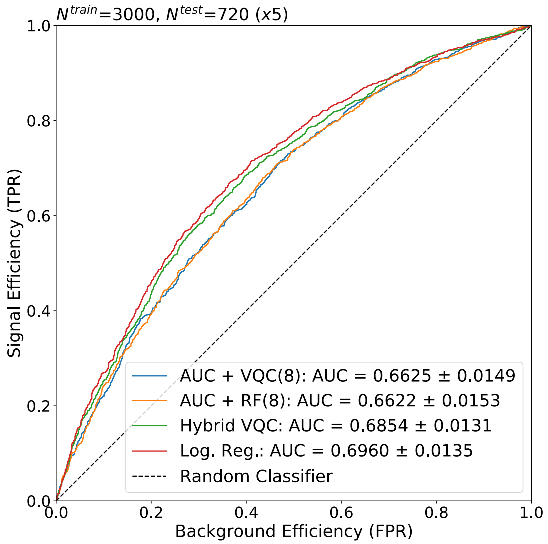
Autoencoder	HP Optimisation	MSE Loss $\times 10^{-4}$	BCE Loss	Classifier AUC	QSVM AUC
Vanilla	-	4.77	-	-	0.56± 0.01
Variational	MSE	4.49	-	-	0.56± 0.02
Classifier	MSE	5.47	0.63	0.700± 0.001	0.56± 0.02
	BCE	62.97	0.61	0.734± 0.002	0.72± 0.01
Sinkhorn	MSE	9.65	-	-	0.51± 0.01
Sinkclass	MSE	26.41	0.65	0.642± 0.003	0.50± 0.01
	BCE	24.69	0.61	0.734± 0.002	0.74± 0.01



Model	BCE Loss	AUC
Encoder + VQC	0.61	0.702± 0.004

Results from: https://doi.org/10.1051/epjconf/202125103070





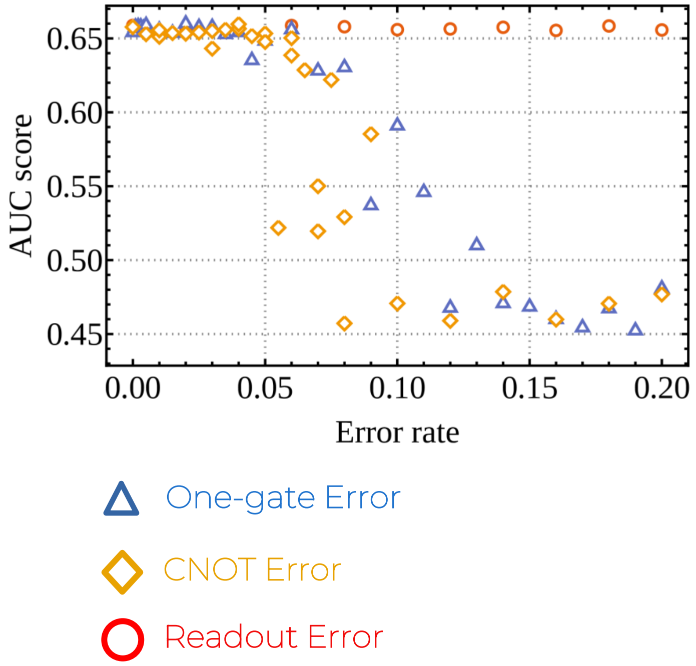
Sinkclass AE shows best performance when considering both *reconstruction power* and *classification power.*

It even matches the classical *state-of-the-art* result!

Conclusions and Outlook

Summary:

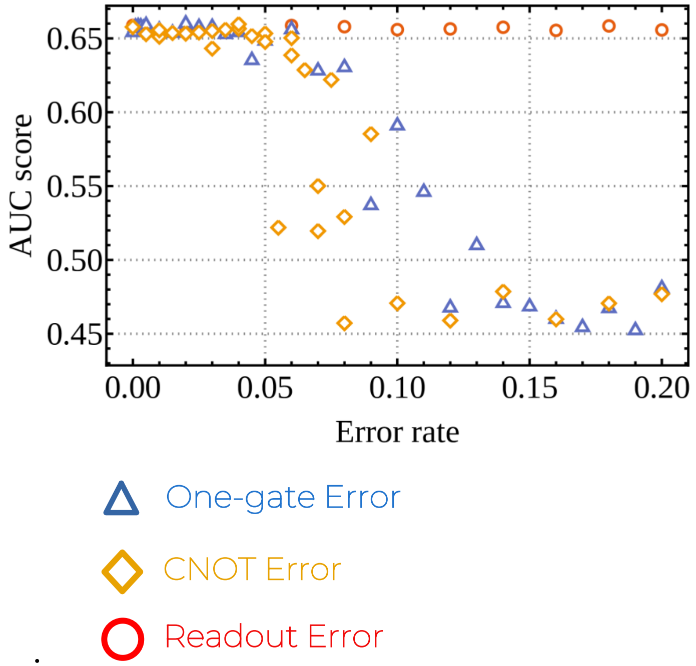
- State-of-the-art performance of the developed hybrid data compression models.
- Feature reduction is crucial, training classical + quantum at the same time yields better results (hybrid VQC) than step-wise training.



Conclusions and Outlook

Summary:

- State-of-the-art performance of the developed hybrid data compression models.
- Feature reduction is crucial, training classical + quantum at the same time yields better results (hybrid VQC) than step-wise training.



Ongoing work:

- Simulations including **hardware noise** and **real hardware** runs.
- **Anomaly detection** (AD) for model independent searches of new physics:
    - Kernel based models.

In preparation: Hybrid VQC for Higgs identification (presented in ACAT 2021)

Future work:

- Quantum branches (QSVM+VQC) on developed networks for feature reduction and AD.

Thank You!

Backup slides.

Event Selection

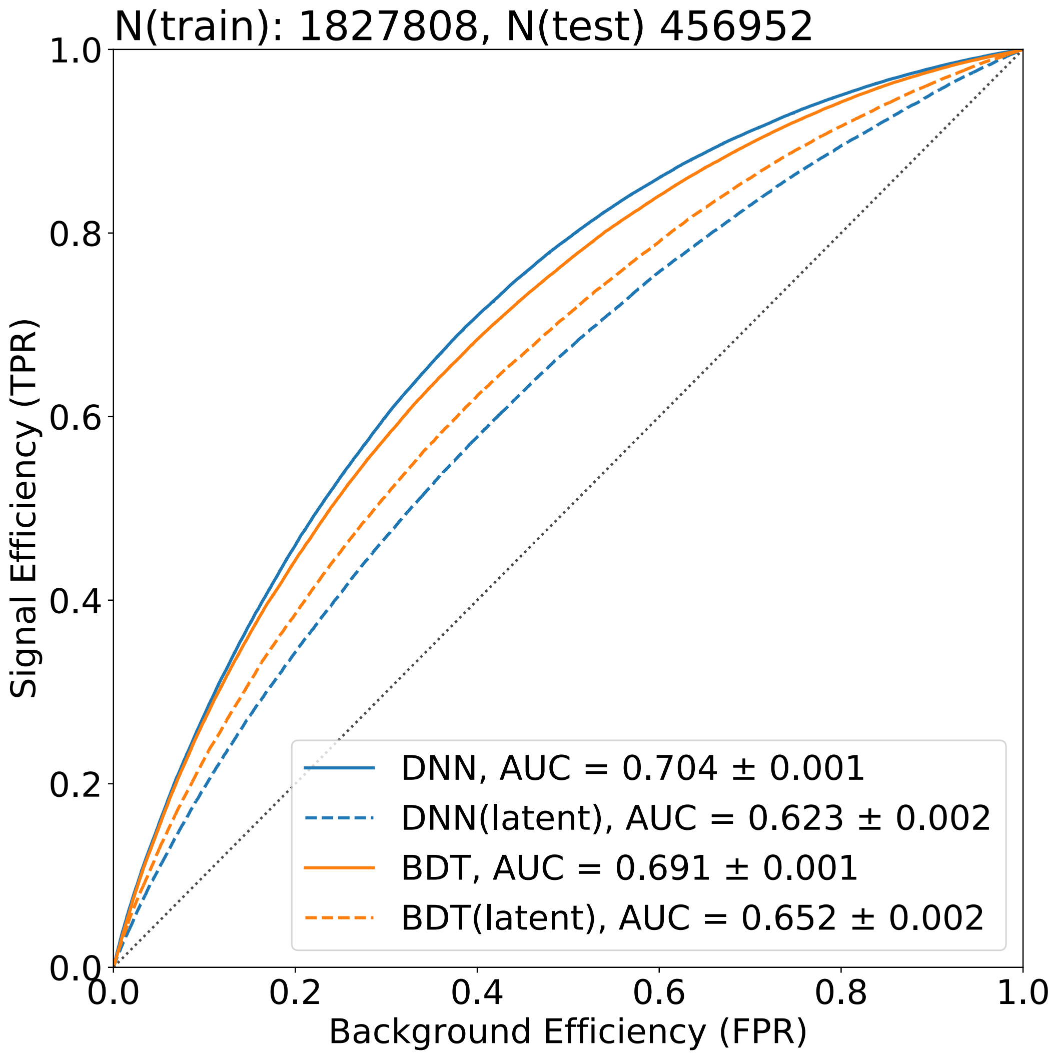
A set of selection cuts were applied to the simulated data to reduce additional backgrounds:

1. Electrons: $p_T > 30\,\text{GeV}$, $|\eta| < 2.1$
2. Muons: $p_T > 26\,\text{GeV}$, $|\eta| < 2.1$
3. Jets: $p_T > 30\,\text{GeV}$, $|\eta| < 2.4$

Furthermore, each event must contain at least 4 jets, 2 b-tagged jets, and exactly 1 lepton. Finally, the 7 most energetic jets are selected, allowing for an extra jet to account for possible final state radiation. The variables in the analysed data set are as follows:

1. Jet related features: (p_T, η, ϕ, E, b-tag, p_x, p_y, p_z)
2. Leptonic features: (p_T, η, ϕ, E, p_x, p_y, p_z)
3. Missing energy related features: (ϕ, p_x, p_y, p_T)

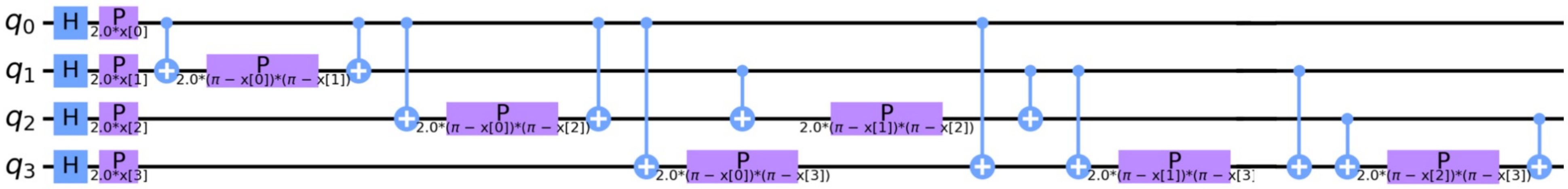
Classification with conventional methods



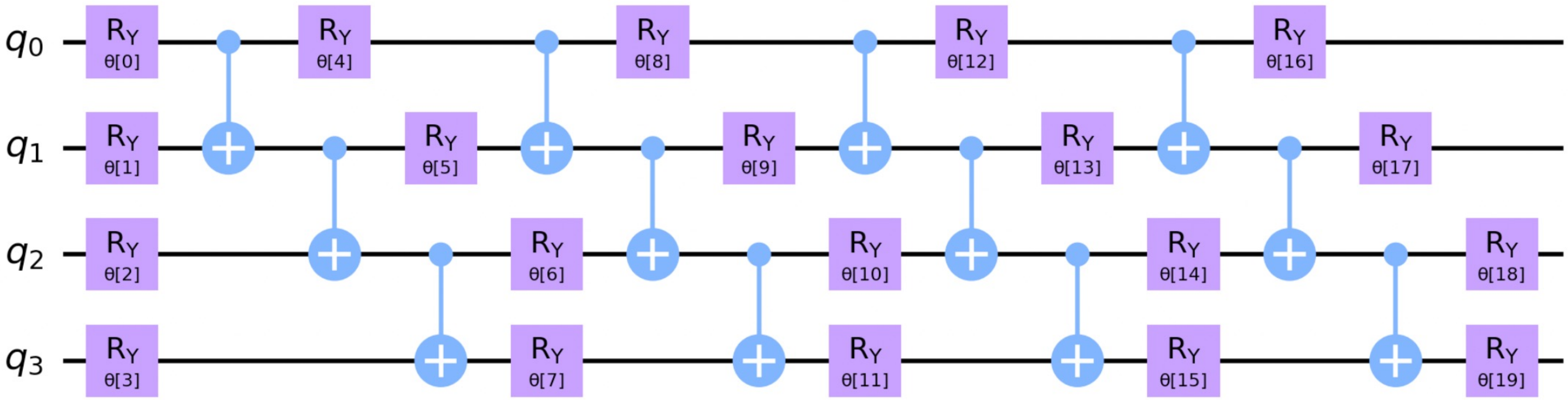
- Assess performance of realistic HEP approaches on generated our data set.
- Full CMS simulation yields higher classifier performance.
- Models trained on full set of input features (67) and a reduced set (16) → benchmark.
- Measure of information loss (discriminating power reduction).

VQC circuits

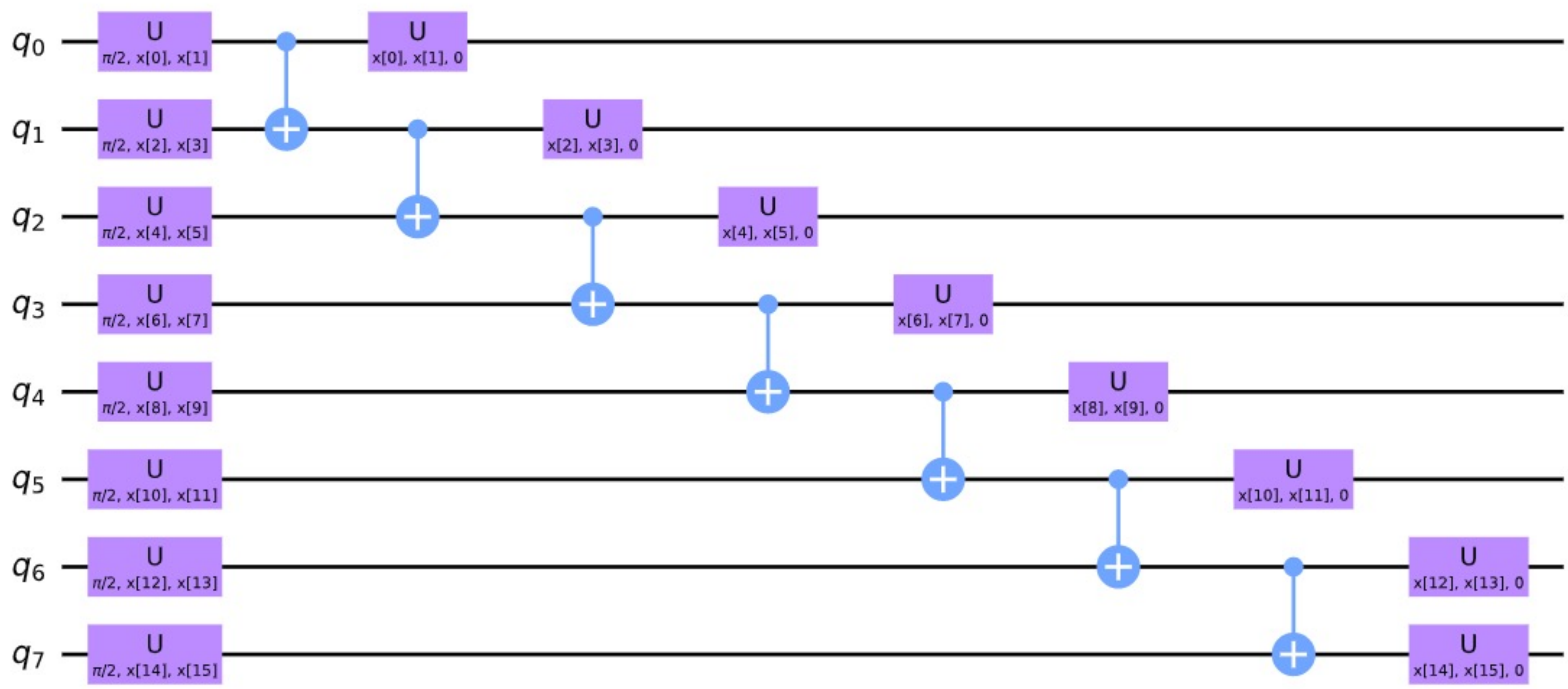
Data encoding for the VQC model [HCTea19]:



Parametrised quantum circuit (QNN):



QSVM Circuit



Conventional FR

Model	AUC	C	Feature Extraction Type
Bernoulli Restricted Boltzmann Machine	0.651 ± 0.016	0.01	Neural Network
Locally Linear Embedding	0.533 ± 0.014	0.01	Manifold Learning
Spectral Embedding	0.526 ± 0.013	0.1	Manifold Learning
Independent Component Analysis	0.528 ± 0.006	0.01	Linear
Non-negative Matrix Factorisation	0.599 ± 0.013	0.001	Linear
Principal Component Analysis	0.541 ± 0.015	10	Linear

Best results obtained with Bernoulli Restricted Boltzmann Machine:
This method is very close to an autoencoder.

Quantum machine learning classifier models

Kernel-based models (Quantum Support Vector Machines):

- Convex optimization tasks.
- Typically required circuits are deeper.
- $\mathcal{O}(n2)$ complexity construction of the kernel matrix elements.

Quantum Neural Networks (Variational Quantum Circuits):

- Non-convex optimization.
- Vanishing gradient problem (Barren plateaus).
- $\mathcal{O}(n)$ complexity.

Encoding (embedding) the classical data in a quantum circuit [SP18]:

Amplitude encoding: exponentially decrease the needed number of qubits *but* have deep circuits.

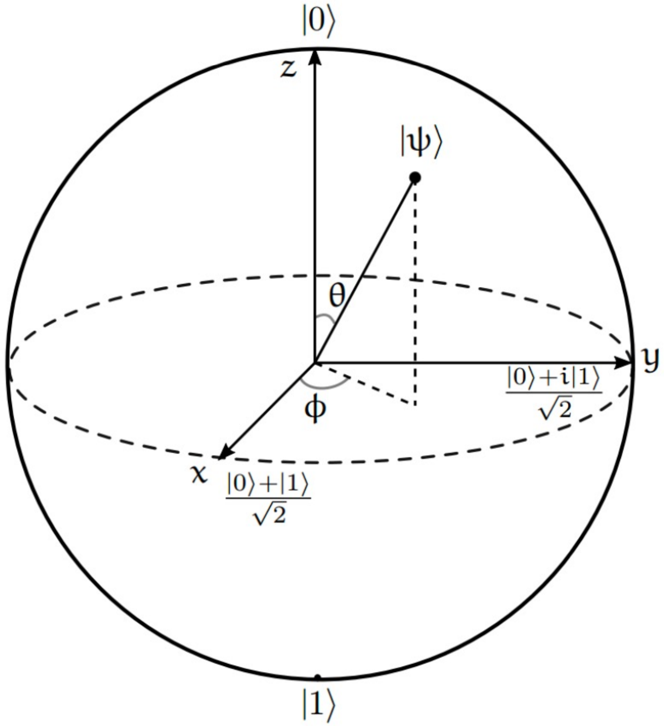
Angle (direct) encoding: map each feature to a separate qubit shallow but wider circuits.

Data re-uploading [PSCLGFL20]: repeat any data embedding circuit.

Basics of quantum information processing

The qubit:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \equiv \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

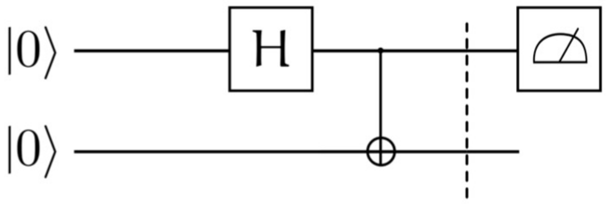


Generic qubit operations (quantum gates) $U = e^{-i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}} \in \mathrm{SU}(2)$:

$$U(\theta,\phi,\lambda) = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) & -e^{i\lambda}\sin\left(\frac{\theta}{2}\right) \\ e^{i\phi}\sin\left(\frac{\theta}{2}\right) & e^{i(\phi+\lambda)}\cos\left(\frac{\theta}{2}\right) \end{pmatrix}$$

Construct all possible gates from $U(\theta,\phi,\lambda)$

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \equiv U\left(\frac{\pi}{2}, 0, \pi\right)$$



$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Quantum gates and universality

Single qubit gates:

- A generic quantum gate can be decomposed in a series of R_y and R_z [BBC+95]

$$U(\theta, \phi, \lambda) = R_z(\lambda) R_y(\theta) R_z(\phi)$$

- For hardware implementation: more convenient to decompose to gates that have a direct physical operation analogue on the device.

Multi-qubit gates:

- 2-qubit SWAP and CNOT (Control-X) gates and the 3-qubit Toffolli gate

$$CX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

- Any control-U gate can be written as a combination of CX, R_y and R_z gates.

Quantum Gate Universality [DiV95]: The above “building blocks” can construct any quantum circuit acting on n qubits, i.e. SU(2^n), operating on at most *two-qubits* at a time.

Hardware Preliminary Results

IBMQ noise model	Run 1	Run 2	Run 3	Run 4	Run 5
belem	0.6598 ± 0.0181	0.6508 ± 0.0183	0.6571 ± 0.0209	0.6582 ± 0.0186	0.6561 ± 0.0192
bogota	0.6590 ± 0.0181	0.6598 ± 0.0191	0.6608 ± 0.0205	cluster error	0.6576 ± 0.0169
lima	0.6574 ± 0.0179	0.6577 ± 0.0187	0.6582 ± 0.0194	0.6578 ± 0.0189	0.6551 ± 0.0175
manila	0.6592 ± 0.0198	0.6576 ± 0.0209	0.6515 ± 0.0188	0.6585 ± 0.0190	0.6586 ± 0.0197
quito	0.6558 ± 0.0218	0.6579 ± 0.0196	0.6567 ± 0.0178	0.6586 ± 0.0197	0.6567 ± 0.0208
santiago	0.6562 ± 0.0197	0.6580 ± 0.0188	0.6603 ± 0.0204	0.6602 ± 0.0181	0.6577 ± 0.0184

References i

- *Measurement of the Higgs boson decaying to b-quarks produced in association with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Tech. Report ATLAS-CONF-2020-058, CERN, Geneva, Nov 2020.
- Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter, *Elementary gates for quantum computation*, Phys. Rev. A **52** (1995), 3457–3467.
- Belis Vasilis, González-Castillo Samuel, Reissel Christina, Vallecorsa Sofia, Combarro Elías F., Dissertori Günther, and Reite Florentin, *Higgs analysis with quantum classifiers*, EPJ Web Conf. **251** (2021), 03070.

References ii

F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark yukawa coupling?*, Journal of Experimental and Theoretical Physics **120** (2015), no. 3, 335–343.

Andrew Blance and Michael Spannowsky, *Quantum machine learning for particle physics using a variational quantum classifier*, arXiv preprint arXiv:2010.07335 (2020).

Measurement of $\mathrm{t\bar{t}H}$ *production in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel in* $41.5\,\mathrm{fb}^{-1}$ *of proton-proton collision data at* $\sqrt{s} = 13\,\mathrm{TeV}$, Tech. Report CMS-PAS-HIG-18-030, CERN, Geneva, 2019.

David P. DiVincenzo, *Two-bit gates are universal for quantum computation*, Phys. Rev. A **51** (1995), 1015–1022.

References iii

- V. Havlíček, A.D. Córcoles, K. Temme, and et al, *Supervised learning with quantum-enhanced feature spaces*, Nature **567** (2019), 209–212.
- Jonas M. Kübler, Simon Buchholz, and Bernhard Schölkopf, *The inductive bias of quantum kernels*, 2021.
- Alex Mott, Joshua Job, Jean-Roch Vlimant, Daniel Lidar, and Maria Spiropulu, *Solving a higgs optimization problem with quantum annealing for machine learning*, Nature **550** (2017), no. 7676, 375–379.
- Benjamin Nachman, Davide Provasoli, Wibe A. de Jong, and Christian W. Bauer, *Quantum algorithm for high energy physics simulations*, Phys. Rev. Lett. **126** (2021), 062001.
- Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I Latorre, *Data re-uploading for a universal quantum classifier*, Quantum **4** (2020), 226.

References iv

- Maria Schuld and Francesco Petruccione, *Supervised learning with quantum computers*, Springer International Publishing, 2018.
- Koji Terashi, Michiru Kaneda, Tomoe Kishimoto, Masahiko Saito, Ryu Sawada, and Junichi Tanaka, *Event classification with quantum machine learning in high-energy physics*, Computing and Software for Big Science **5** (2021), no. 1, 1–11.
- Sau Lan Wu, Jay Chan, Wen Guan, Shaojun Sun, Alex Wang, Chen Zhou, Miron Livny, Federico Carminati, Alberto Di Meglio, Andy CY Li, et al., *Application of quantum machine learning using the quantum variational classifier method to high energy physics analysis at the lhc on ibm quantum computer simulator and hardware with 10 qubits*, arXiv preprint arXiv:2012.11560 (2020).