Studies of Quantum Dots using Machine Learning

Even Marius Nordhagen

University of Oslo

evenmn@fys.uio.no

December 18, 2019

Outline

- Motivation
- Quantum Theory
- Machine Learning Theory
- Methods
- Results
- Conclusions

Motivation

Studies of Quantum Dots using Machine Learning

Studies of Quantum Dots using Machine Learning

Quantum Dots

- What are quantum dots?

Quantum Dots

- What are quantum dots?
- Why are quantum dots interesting?
  - Quantum dots are expected to be the next big thing in display technology[1,2]
  - Quantum dots are used in quantum computers
  - Researchers have managed to study two-dimensional quantum dots in the laboratory[3]
  - An array of interesting physical phenomena can be observed in quantum dots

Studies of Quantum Dots using Machine Learning

Studies of Quantum Dots using Machine Learning

Machine Learning

- “*Machine learning is the science of getting computers to act without being explicitly programmed*[4].”

Machine Learning

- “*Machine learning is the science of getting computers to act without being explicitly programmed*[4].”

- Image recognition

Machine Learning

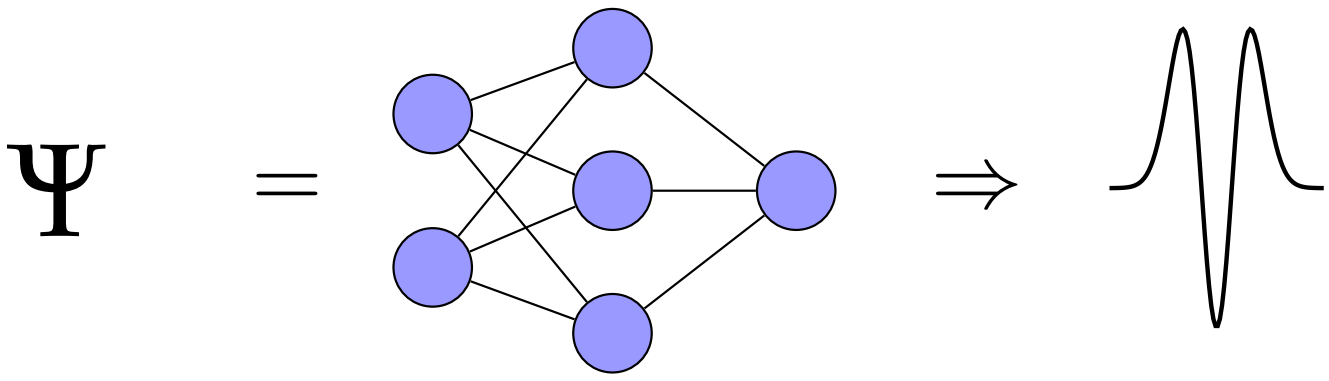
- > *Machine learning is the science of getting computers to act without being explicitly programmed*[4].
- Image recognition
- Natural language processing

Machine Learning + Quantum Mechanics

- Neural networks are eminent function approximators

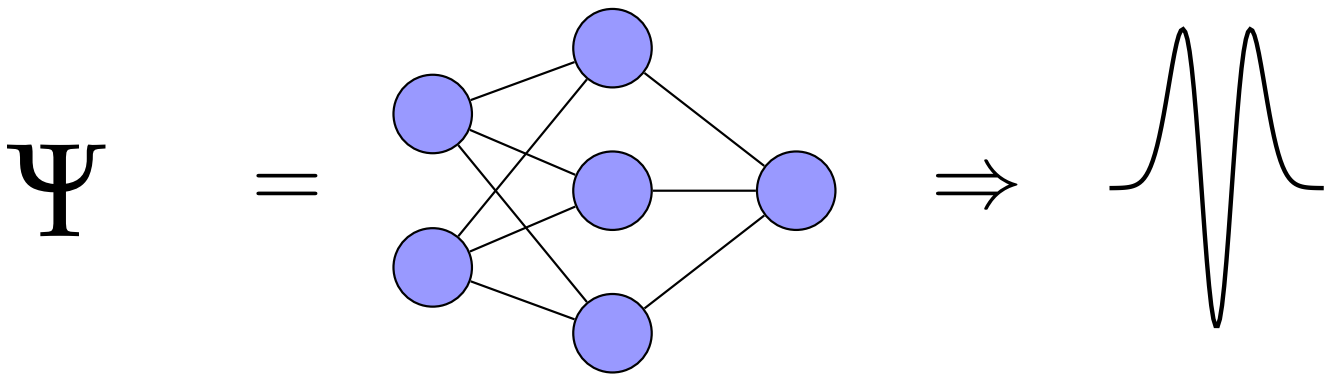
Machine Learning + Quantum Mechanics

- Neural networks are eminent function approximators



Machine Learning + Quantum Mechanics

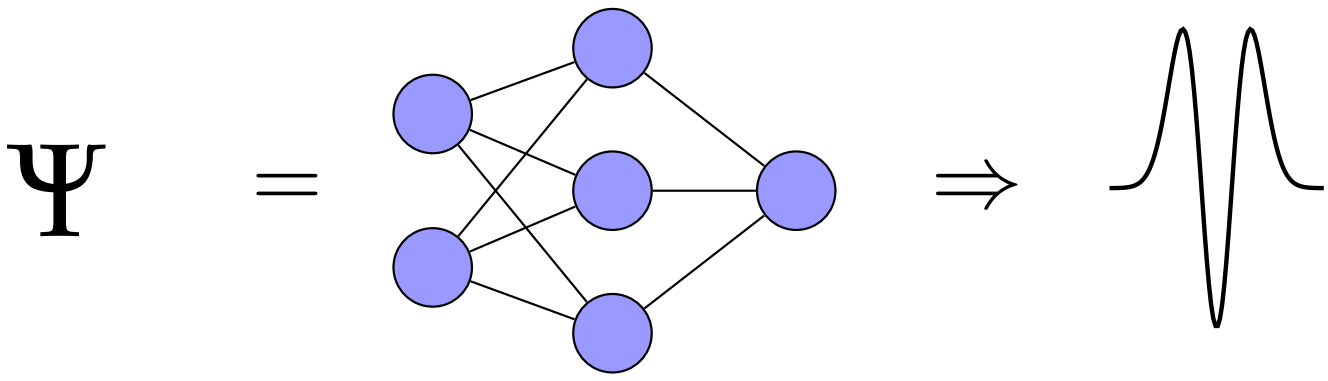
- Neural networks are eminent function approximators



- Existing methods are reminiscent of machine learning algorithms

Machine Learning + Quantum Mechanics

- Neural networks are eminent function approximators



- Existing methods are reminiscent of machine learning algorithms

- Literature study (Carleo & Troyer[5], Flugsrud[6], Pfau *et al.*[7])

Ethics in Science

- Respect for other's work

Ethics in Science

- Respect for other's work
- Reproducibility

Quantum Theory

The Schrödinger Equation

$$\hat{\mathcal{H}}\Psi = E\Psi$$

The Schrödinger Equation

$$\hat{\mathcal{H}}\Psi = E\Psi$$

$$\Downarrow$$

$$E = \frac{\int d\boldsymbol{X}\Psi^*(\boldsymbol{X})\hat{\mathcal{H}}\Psi(\boldsymbol{X})}{\int d\boldsymbol{X}\Psi^*(\boldsymbol{X})\Psi(\boldsymbol{X})}$$

The Variational Principle

The variational principle serves as a way of finding the ground state energy. For an arbitrary trial wave function $\Psi_T(\boldsymbol{X})$, it states that the obtained energy is larger or equal to the ground state,

$$E_0 \leq E = \frac{\int d\boldsymbol{X} \Psi_T^*(\boldsymbol{X}) \hat{\mathcal{H}} \Psi_T(\boldsymbol{X})}{\int d\boldsymbol{X} \Psi_T^*(\boldsymbol{X}) \Psi_T(\boldsymbol{X})}.$$

Thus, by minimizing the obtained energy, E, we can estimate the ground state energy.

Quantum Dots

Circular quantum dots $\rightarrow$ electrons confined in a harmonic oscillator potential:

$$\hat{\mathcal{H}} = \sum_{i=1}^{N} \left[-\frac{1}{2}\nabla_i^2 + \frac{1}{2}\omega^2 |\boldsymbol{r}_i|^2 + \sum_{j>i}^{N} \frac{1}{r_{ij}} \right].$$

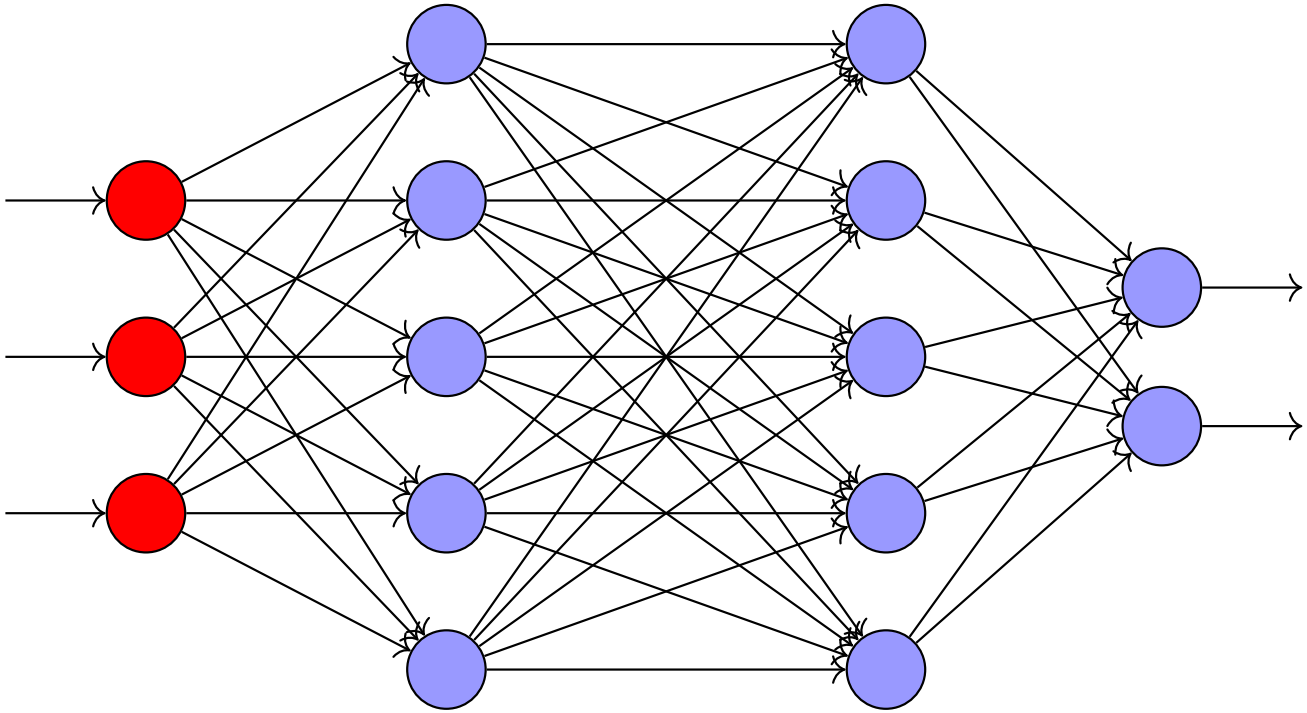
The number of electrons that give full shells are given by

$$N = 2\binom{n+d}{d},$$

which are the magic numbers.

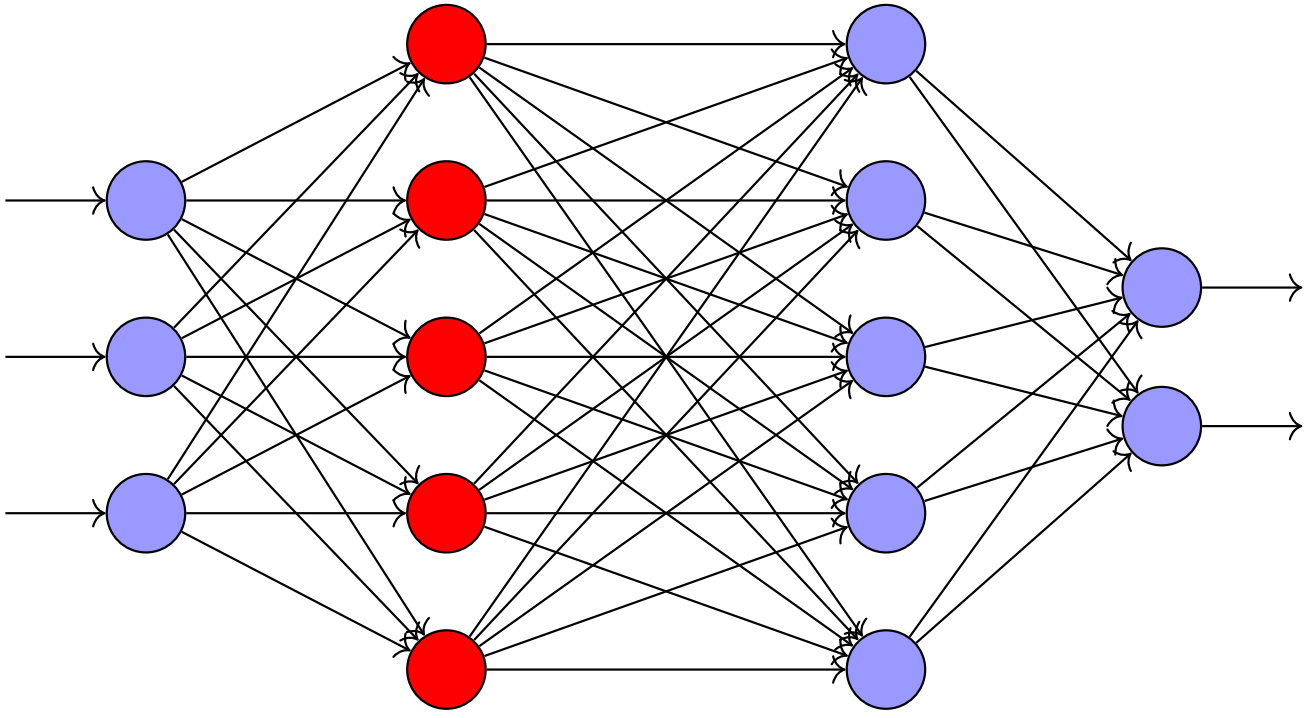
Machine Learning Theory

Feed-forward Neural Network (FNN)



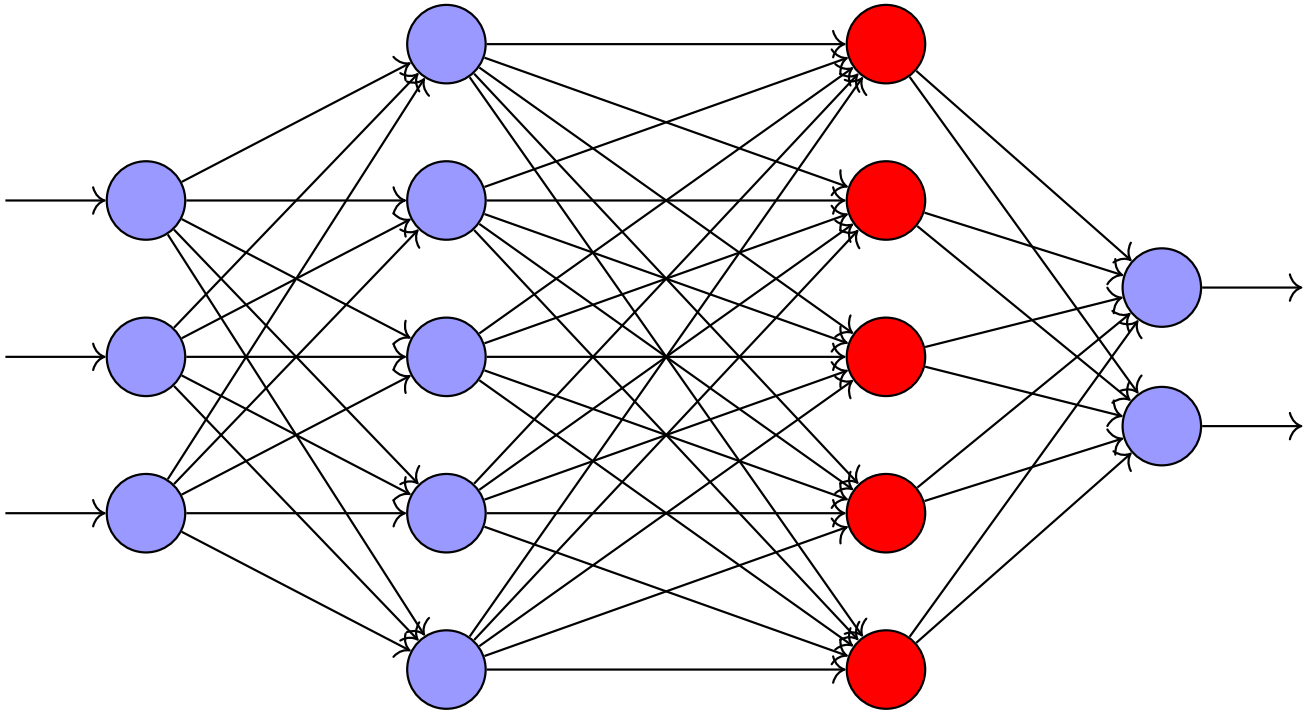
$$a_0 = x$$

Feed-forward Neural Network (FNN)



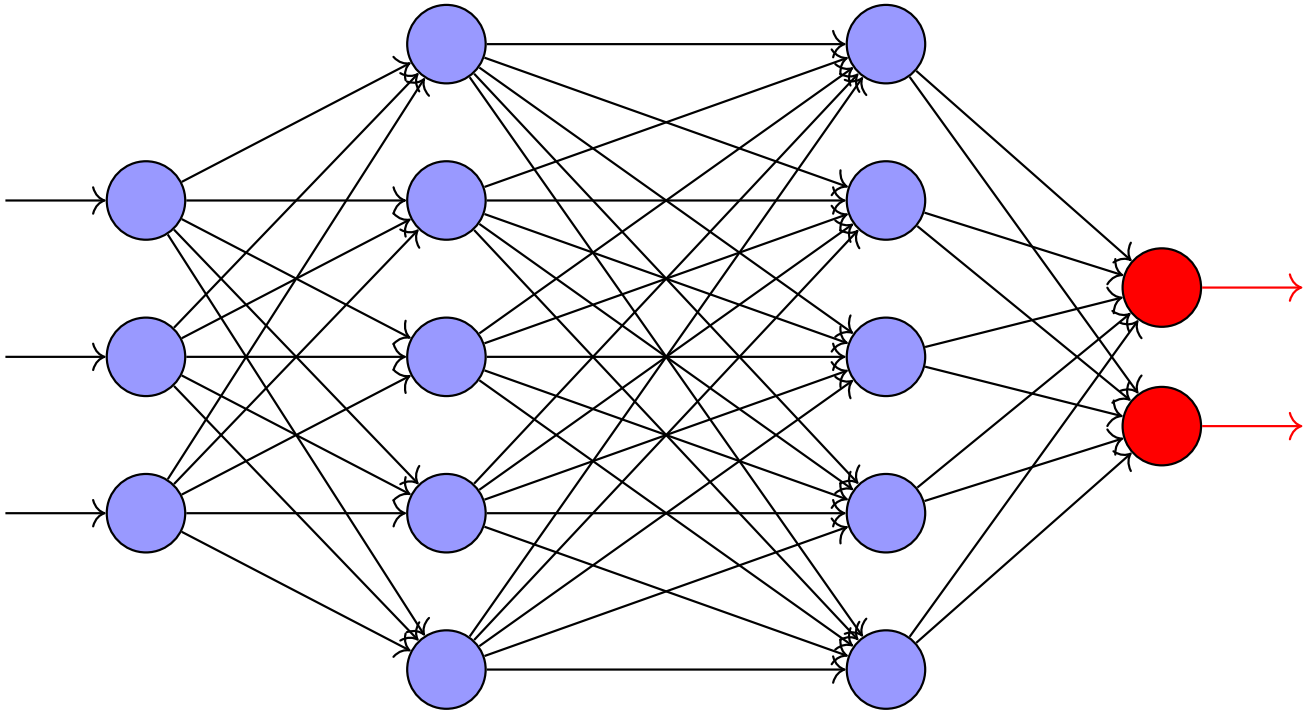
$$\boldsymbol{a}_0 = \boldsymbol{x} \qquad \boldsymbol{a}_1 = f_1(\boldsymbol{a}_0)$$

Feed-forward Neural Network (FNN)



$$\boldsymbol{a}_0 = \boldsymbol{x} \qquad \boldsymbol{a}_1 = f_1(\boldsymbol{a}_0) \qquad \boldsymbol{a}_2 = f_2(\boldsymbol{a}_1)$$

Feed-forward Neural Network (FNN)



$$\boldsymbol{a}_0 = \boldsymbol{x} \qquad \boldsymbol{a}_1 = f_1(\boldsymbol{a}_0) \qquad \boldsymbol{a}_2 = f_2(\boldsymbol{a}_1) \qquad \tilde{\boldsymbol{y}} = f_3(\boldsymbol{a}_2)$$

Cost function

- The cost function defines the error

Cost function

- The cost function defines the error
- Mean square error (MSE):

$$\mathcal{C} = \frac{1}{2}\sum_{i=1}^{n}(y - \tilde{y})^2.$$

Cost function

- The cost function defines the error
- Mean square error (MSE):

$$\mathcal{C} = \frac{1}{2}\sum_{i=1}^{n}(y - \tilde{y})^2.$$

- Attempt to minimize the cost function

Optimization Algorithms

- Minimize the cost function

Optimization Algorithms

- Minimize the cost function

- The gradient descent method:

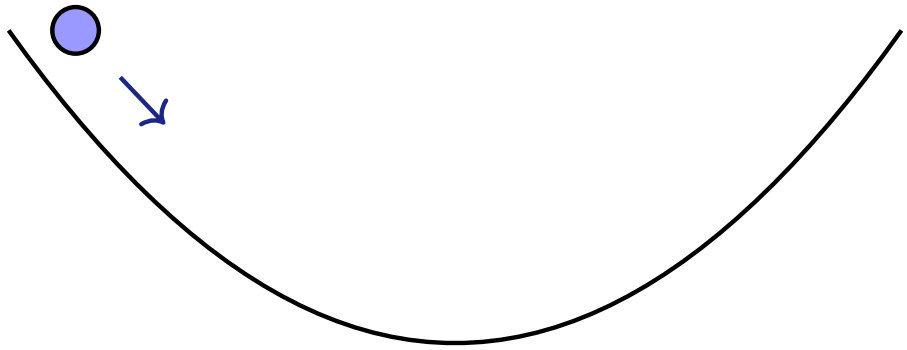
$$\theta^{+} = \theta - \frac{\partial \mathcal{C}}{\partial \theta}.$$

Optimization Algorithms

- Minimize the cost function

- The gradient descent method:

$$\theta^{+} = \theta - \frac{\partial \mathcal{C}}{\partial \theta}.$$

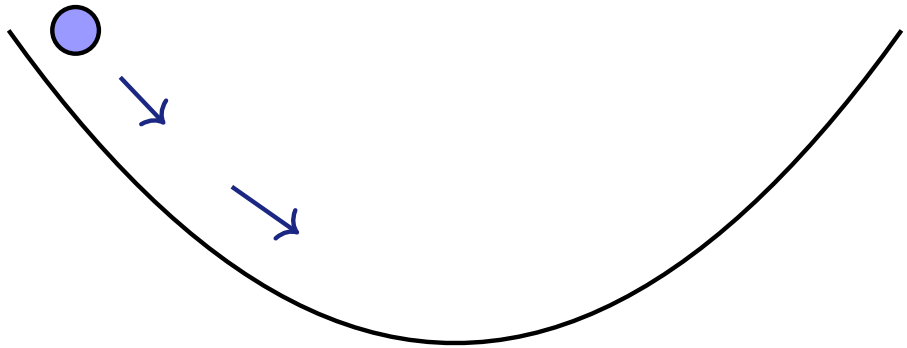


Optimization Algorithms

- Minimize the cost function

- The gradient descent method:

$$\theta^{+} = \theta - \frac{\partial \mathcal{C}}{\partial \theta}.$$

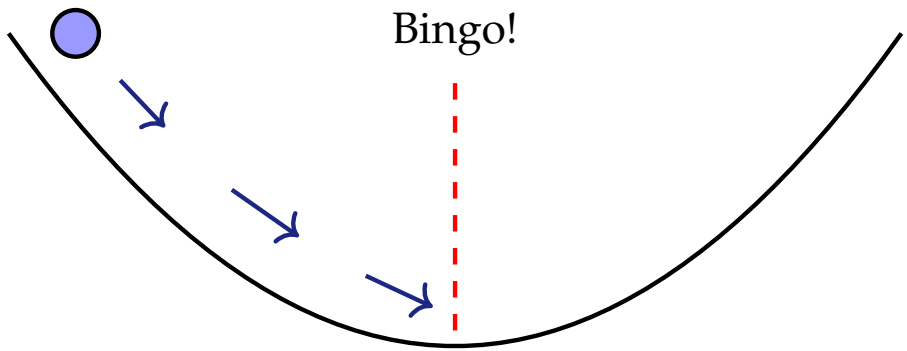


Optimization Algorithms

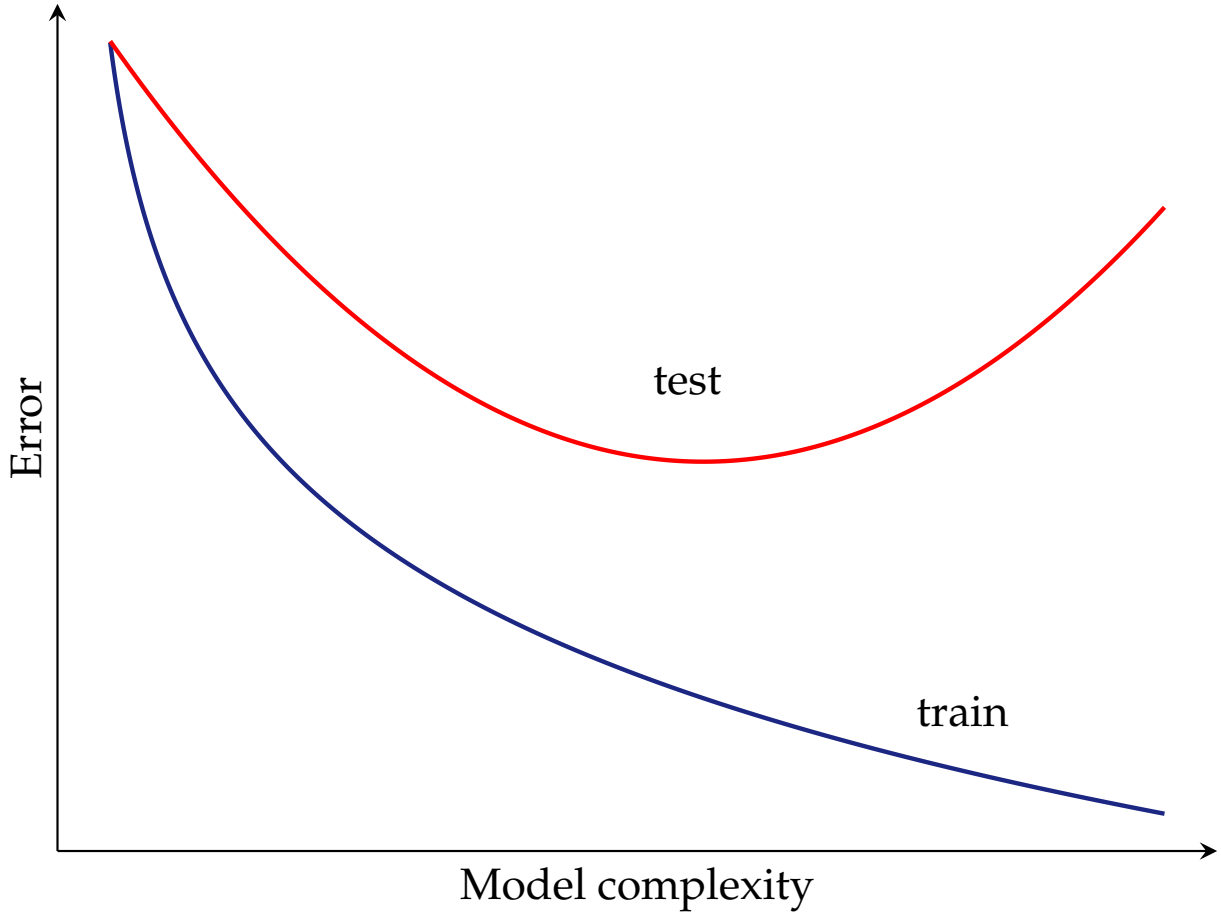
- Minimize the cost function

- The gradient descent method:

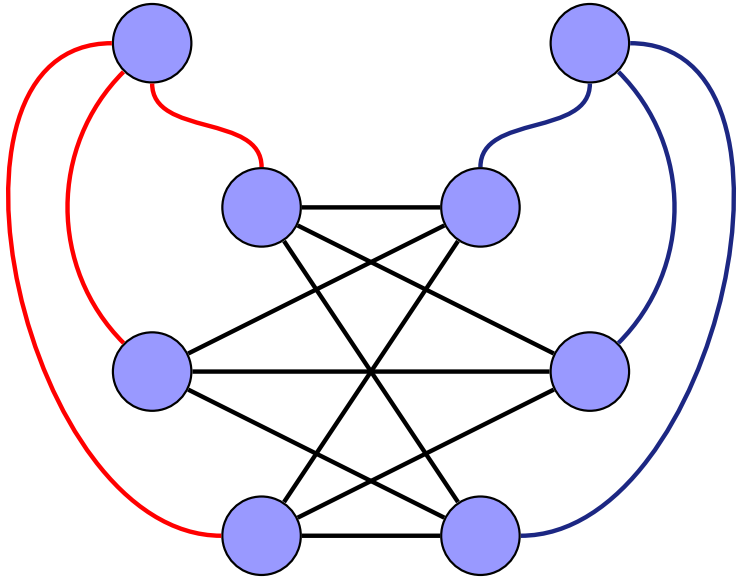
$$\theta^{+} = \theta - \frac{\partial \mathcal{C}}{\partial \theta}.$$



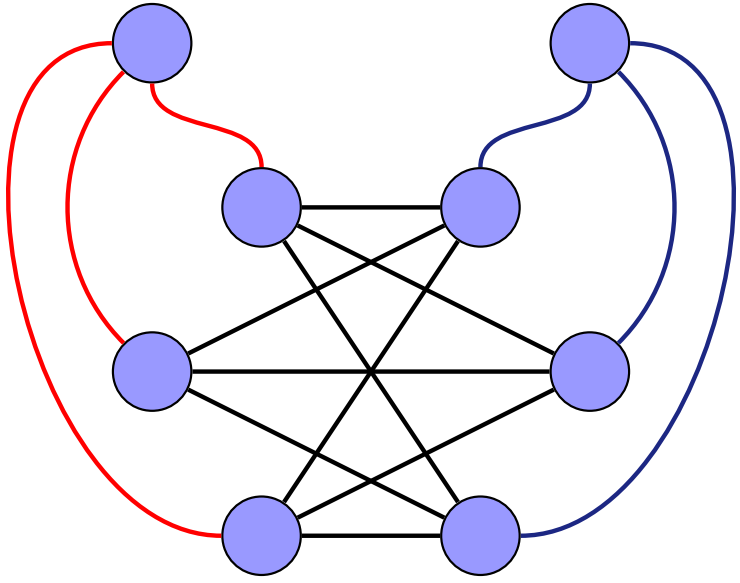
Find Appropriate Complexity



Restricted Boltzmann Machines



Restricted Boltzmann Machines



visible hidden

$$E(\boldsymbol{x},\boldsymbol{h}) = -\sum_{i=1}^{V}\frac{(x_i-a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{H} h_j b_j - \sum_{i=1}^{V}\sum_{j=1}^{H}\frac{x_i w_{ij} h_j}{\sigma_i^2}$$

Probability Distribution

The joint probability distribution is given by the Boltzmann distribution:

$$P(\boldsymbol{x}, \boldsymbol{h}) = \frac{1}{Z} \exp(-E(\boldsymbol{x}, \boldsymbol{h})/kT).$$

The marginal distribution of the visible units is given by

$$P(\boldsymbol{x}) = \sum_{\{\boldsymbol{h}\}} P(\boldsymbol{x}, \boldsymbol{h}).$$

Methods

Variational Monte Carlo (VMC)

Exploit the variational principle in order to obtain the ground state energy

$$E_0 < E_{\mathrm{VMC}} = \frac{\int d\boldsymbol{R}\Psi_T(\boldsymbol{R})^*\hat{\mathcal{H}}\Psi_T(\boldsymbol{R})}{\int d\boldsymbol{R}\Psi_T(\boldsymbol{R})^*\Psi_T(\boldsymbol{R})},$$
$$= \int d\boldsymbol{R}E_L(\boldsymbol{R})P(\boldsymbol{R}),$$

with

$$E_L(\boldsymbol{R}) = \frac{1}{\Psi_T(\boldsymbol{R})}\hat{\mathcal{H}}\Psi_T(\boldsymbol{R}) \quad \wedge \quad P(\boldsymbol{R}) = \frac{\Psi_T(\boldsymbol{R})^*\Psi_T(\boldsymbol{R})}{\int d\boldsymbol{R}\Psi_T(\boldsymbol{R})^*\Psi_T(\boldsymbol{R})}$$

Monte Carlo Integration

We attempt to solve the integral by sampling from the probability density function $P(\boldsymbol{R}) \propto \Psi_T(\boldsymbol{R})^* \Psi_T(\boldsymbol{R})$:

$$E_{\mathrm{VMC}} = \int d\boldsymbol{R} E_L(\boldsymbol{R}) P(\boldsymbol{R}),$$

$$\approx \frac{1}{M} \sum_{i=1}^{M} E_L(\boldsymbol{R}_i).$$

Trial Wave Function Ansatz

The Slater-Jastrow function is the *de facto* standard trial wave function for electronic structure systems,

$$\Psi_T(\boldsymbol{R}) = |\hat{D}(\boldsymbol{R})|J(\boldsymbol{R}),$$

where the Slater matrix,

$$\hat{D}(\boldsymbol{R}) = \begin{pmatrix} \phi_1(\boldsymbol{r}_1) & \phi_2(\boldsymbol{r}_1) & \dots & \phi_N(\boldsymbol{r}_1) \\ \phi_1(\boldsymbol{r}_2) & \phi_2(\boldsymbol{r}_2) & \dots & \phi_N(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\boldsymbol{r}_N) & \phi_2(\boldsymbol{r}_N) & \dots & \phi_N(\boldsymbol{r}_N) \end{pmatrix},$$

contains all the single-particle functions.

Single-particle Functions

The Hermite functions,

$$\phi_n(\boldsymbol{r}) \propto H_n(\sqrt{\omega}\boldsymbol{r}) \exp\left(-\frac{1}{2}\alpha\omega|\boldsymbol{r}|^2\right),$$

are often used as the single-particle functions for quantum dots. The Gaussian can be factorized out from the Slater determinant,

$$|\hat{D}(\boldsymbol{R};\alpha)| \propto \exp\left(-\frac{1}{2}\alpha\omega|\boldsymbol{R}|^2\right) \begin{vmatrix} H_1(\boldsymbol{r}_1) & H_2(\boldsymbol{r}_1) & \dots & H_N(\boldsymbol{r}_1) \\ H_1(\boldsymbol{r}_2) & H_2(\boldsymbol{r}_2) & \dots & H_N(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ H_1(\boldsymbol{r}_N) & H_2(\boldsymbol{r}_N) & \dots & H_N(\boldsymbol{r}_N) \end{vmatrix}.$$

Restricted Boltzmann Machine

As suggested by Carleo & Troyer[5], we use the marginal distribution of the visible units as the single-particle functions in the Slater determinant, and see if them can model the correlations

$$\phi_n(\boldsymbol{r}) \propto H_n(\sqrt{\omega}\boldsymbol{r})P(\boldsymbol{r};\boldsymbol{\theta})$$

where $P(\boldsymbol{r})$ is the marginal distribution of the visible units.

$$|\hat{D}(\boldsymbol{r};\boldsymbol{\theta})| \propto P(\boldsymbol{r};\boldsymbol{\theta}) \begin{vmatrix} H_1(\boldsymbol{r}_1) & H_2(\boldsymbol{r}_1) & \dots & H_N(\boldsymbol{r}_1) \\ H_1(\boldsymbol{r}_2) & H_2(\boldsymbol{r}_2) & \dots & H_N(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ H_1(\boldsymbol{r}_N) & H_2(\boldsymbol{r}_N) & \dots & H_N(\boldsymbol{r}_N) \end{vmatrix}$$

Jastrow Factor

The Jastrow factor is added to account for the correlations
Simple Jastrow factor

$$J(\boldsymbol{r};\boldsymbol{\beta}) = \exp\left(\sum_{i=1}^{N}\sum_{j>i}^{N}\beta_{ij}r_{ij}\right).$$

Padé-Jastrow factor

$$J(\boldsymbol{r};\beta) = \exp\left(\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{a_{ij}r_{ij}}{1+\beta r_{ij}}\right).$$

Our Trial Wave Function Ansätze

- $\Psi_{\text{RBM}}(\boldsymbol{R}) = |\hat{D}_{\text{RBM}}(\boldsymbol{R})|$
- $\Psi_{\text{RBM+SJ}}(\boldsymbol{R}) = |\hat{D}_{\text{RBM}}(\boldsymbol{R})|J(\boldsymbol{R};\boldsymbol{\beta})$
- $\Psi_{\text{RBM+PJ}}(\boldsymbol{R}) = |\hat{D}_{\text{RBM}}(\boldsymbol{R})|J(\boldsymbol{R};\beta)$
- $\Psi_{\text{VMC}}(\boldsymbol{R}) = |\hat{D}_{\text{Gauss}}(\boldsymbol{R})|J(\boldsymbol{R};\beta)$

Software

Three principal aims

- Efficient

Software

Three principal aims

- Efficient
- Flexible

Software

Three principal aims

- Efficient
- Flexible
- Straightforward to use

Results

Ground State Energy

Number of electrons: $N = 2$. Frequency: ω.

ω	RBM	RBM+SJ	RBM+PJ	VMC	HF *	Exact †
1/6	0.7036(1)	0.67684(7)	0.66715(6)	0.66710(1)	0.768675	2/3
0.28	1.07050(4)	1.03470(7)	1.021668(7)	1.02192(1)	1.14171	
1	3.0803(2)	3.02108(5)	2.999587(5)	2.99936(1)	3.16190	3

*Computation of the Hartree-Fock limit by Mariadason, 2018 [8].

†Semi-analytical ground state energy calculated by Taut, 1993 [9].

Ground State Energy

Number of electrons: $N = 20$. Frequency: ω.

ω	RBM	RBM+SJ	RBM+PJ	VMC	HF [‡]	DMC [§]
0.1	30.824(2)	30.567(3)	30.1553(9)	30.0403(2)	31.1902	29.9779(1)
1.0	159.428(3)	156.816(4)	156.104(1)	155.8900(4)	158.004	155.8822(1)

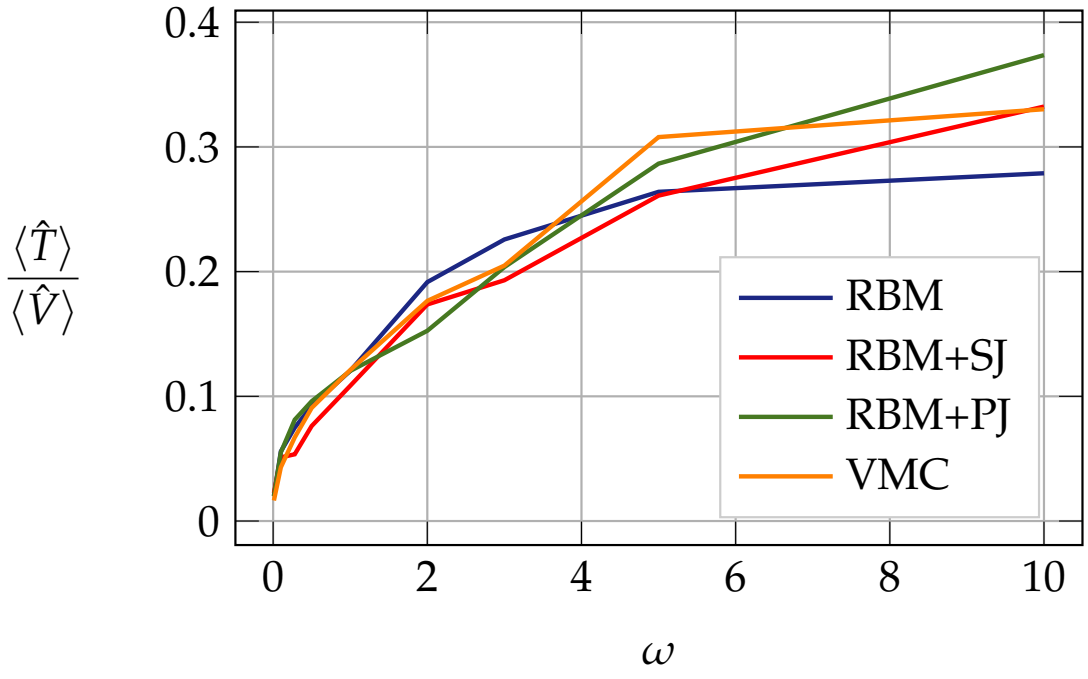
[‡]Computation of the Hartree-Fock limit by Mariadason, 2018 [8].

[§]Ground state energy estimate using the diffusion Monte Carlo method. By Høgberget, 2013 [10].

Energy distribution

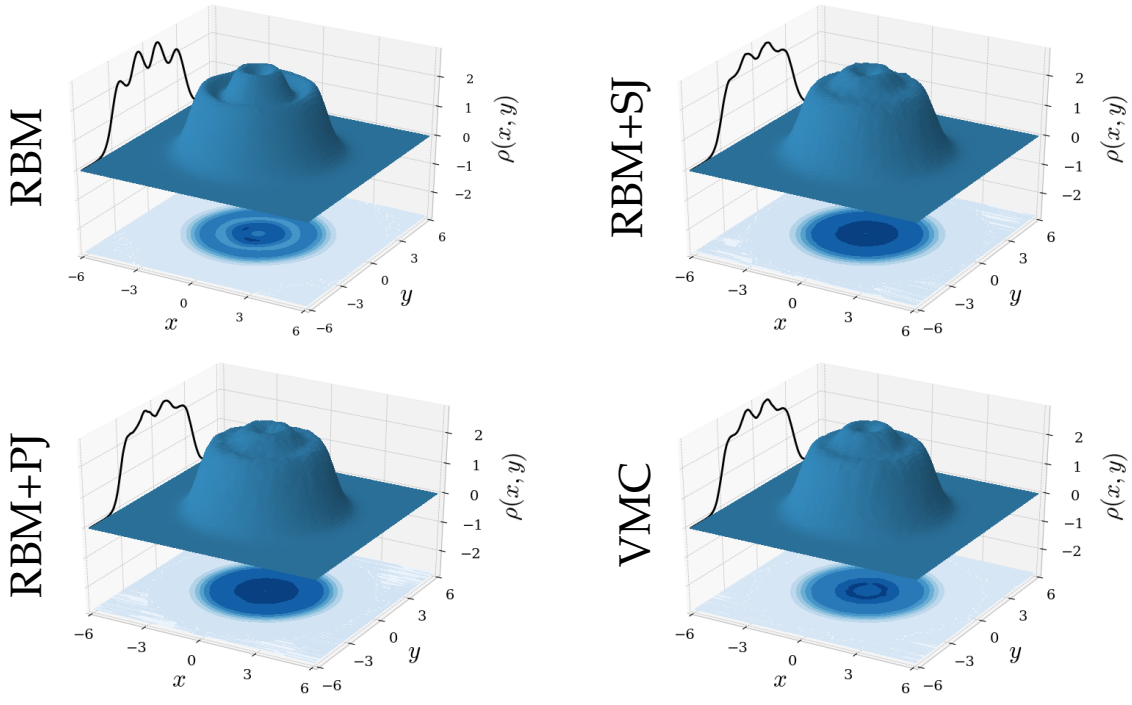
Number of electrons: $N = 20$. Frequency: ω.

Ratio between the kinetic energy, $\langle \hat{T} \rangle$, and the total potential energy, $\langle \hat{V} \rangle$.



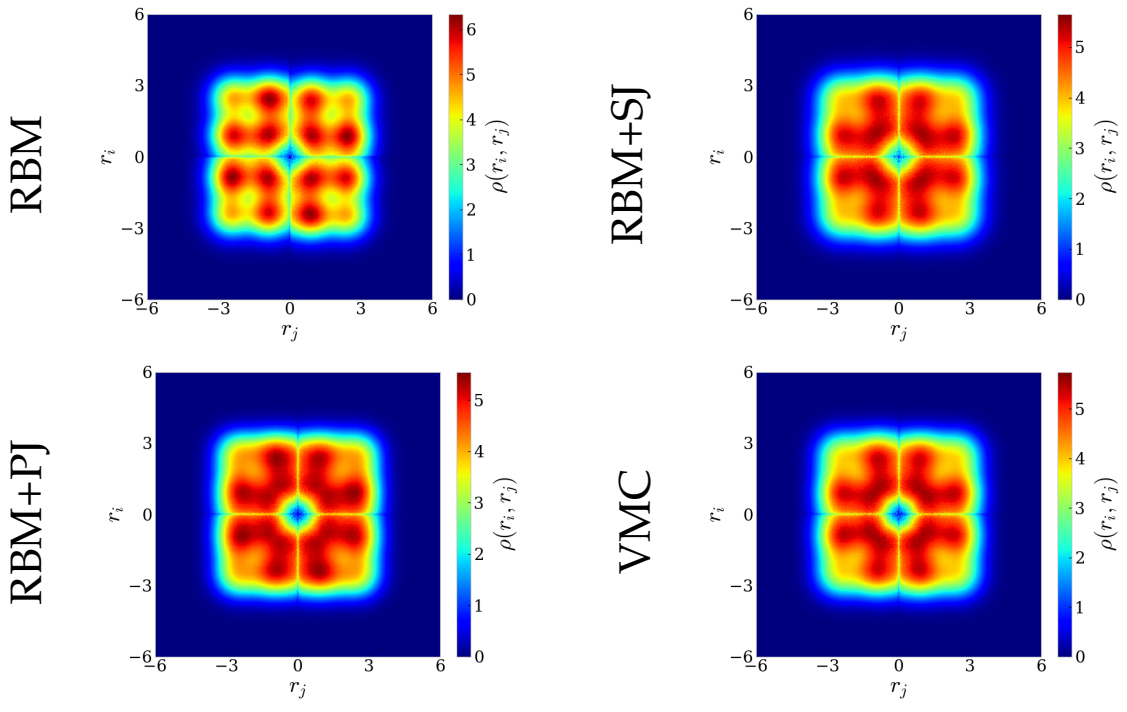
One-body Density

Number of electrons: $N = 20$. Frequency: $\omega = 1.0$.



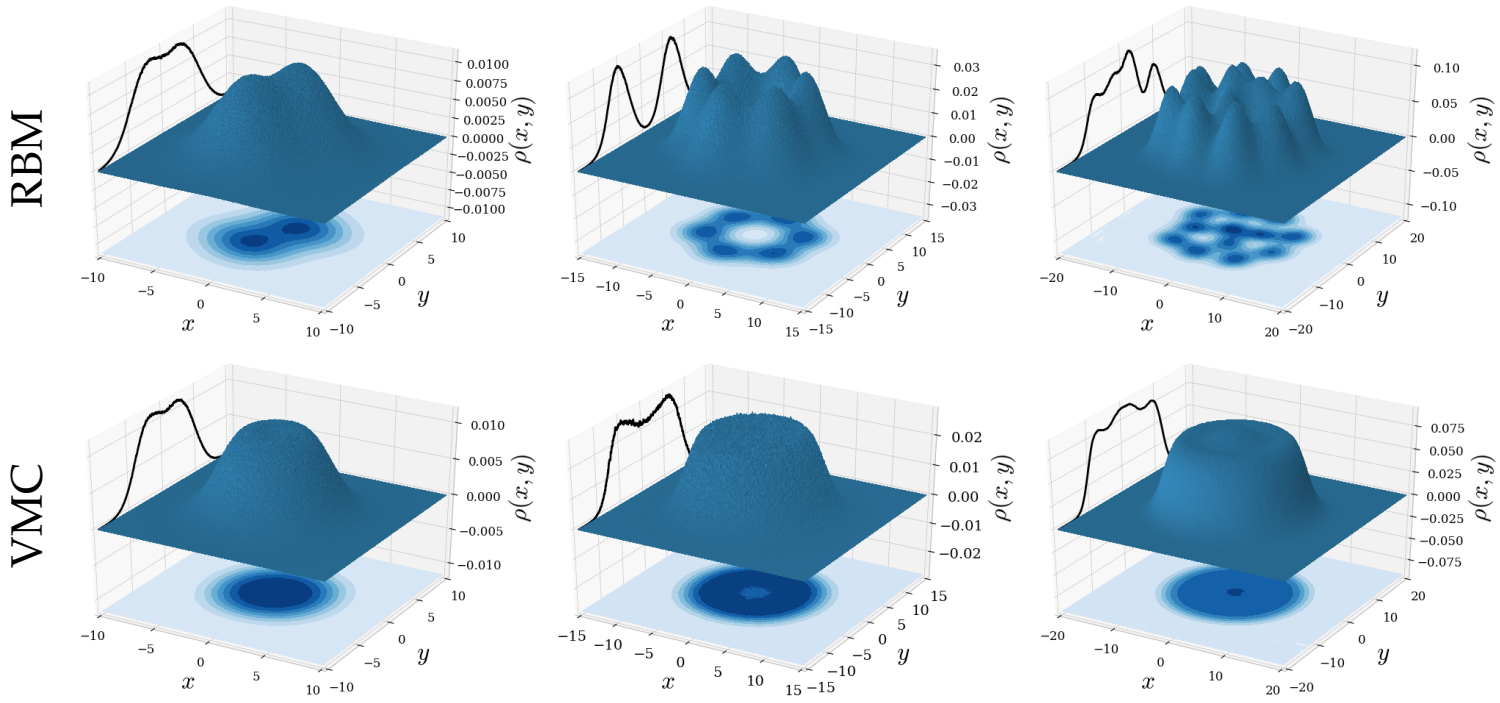
Two-body Density

Number of electrons: $N = 20$. Frequency: $\omega = 1.0$.



Low-frequency Dots

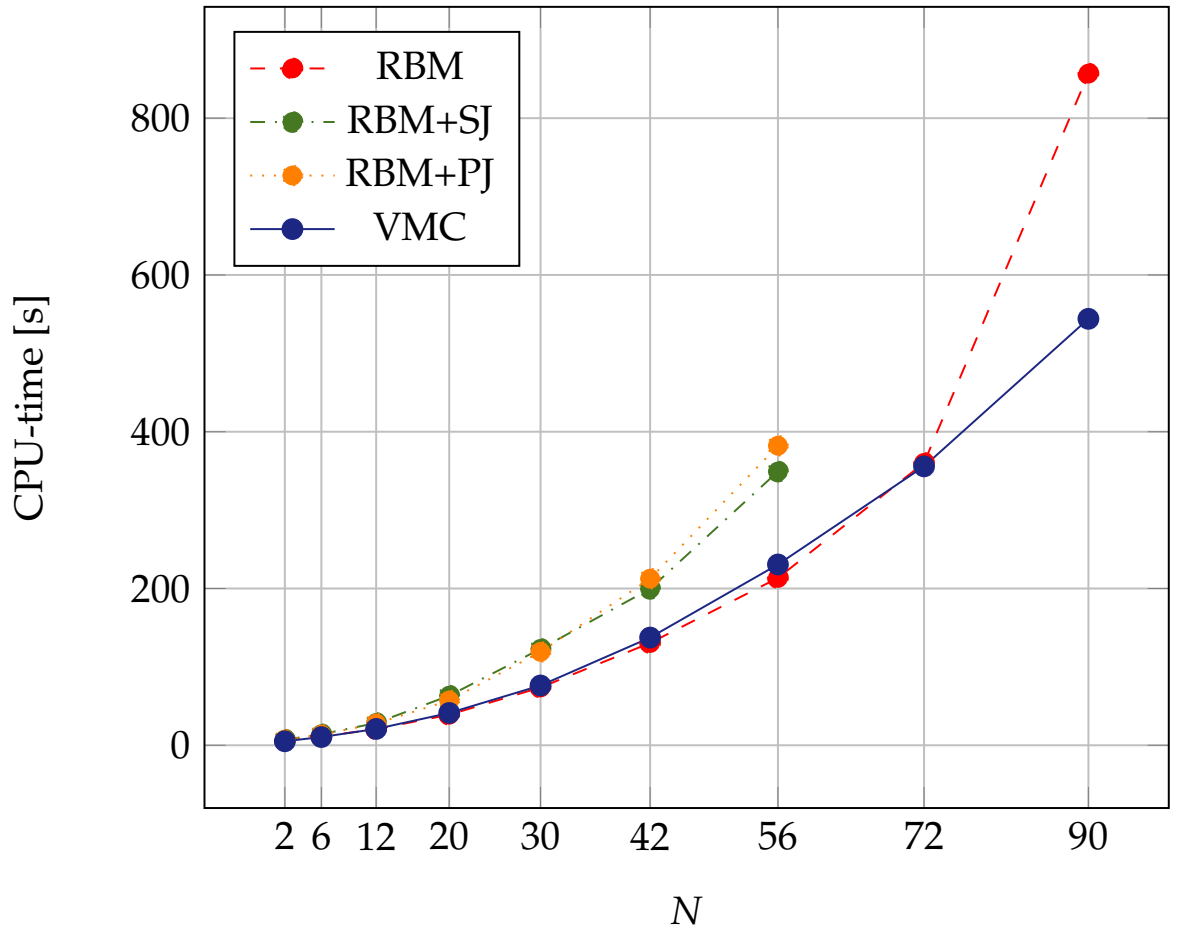
Number of electrons: N. Frequency: $\omega = 0.1$.



UiO : University of Oslo

Computational Cost

Number of electrons: N.



Conclusions

Findings

- The RBM ansatz is able to account for most of the correlations
- The RBM+PJ ansatz might give a better ground state estimate of small quantum dots, compared to the traditional VMC ansatz
- The energy distribution is different for the different ansätze, indicating different electron configurations.
- The ground state energy might not be the best way to evaluate the various ansätze

Future Work

- Investigate restricted Boltzmann machines with other architectures

Future Work

- Investigate restricted Boltzmann machines with other architectures
- Try other optimization algorithms

Future Work

- Investigate restricted Boltzmann machines with other architectures
- Try other optimization algorithms
- Pass more information to the restricted Boltzmann machine

Future Work

- Investigate restricted Boltzmann machines with other architectures
- Try other optimization algorithms
- Pass more information to the restricted Boltzmann machine
- Apply the method on more complex systems

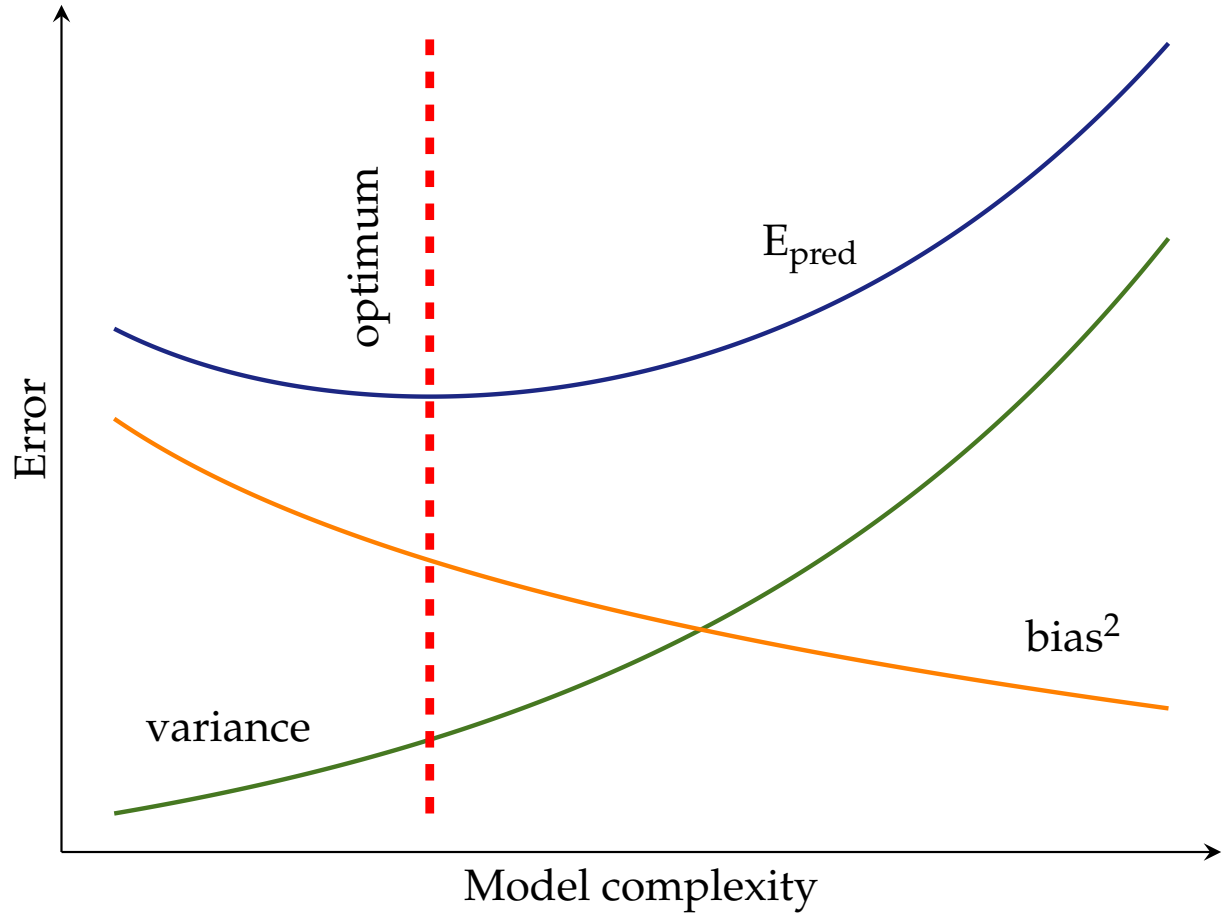
Thank you!

References

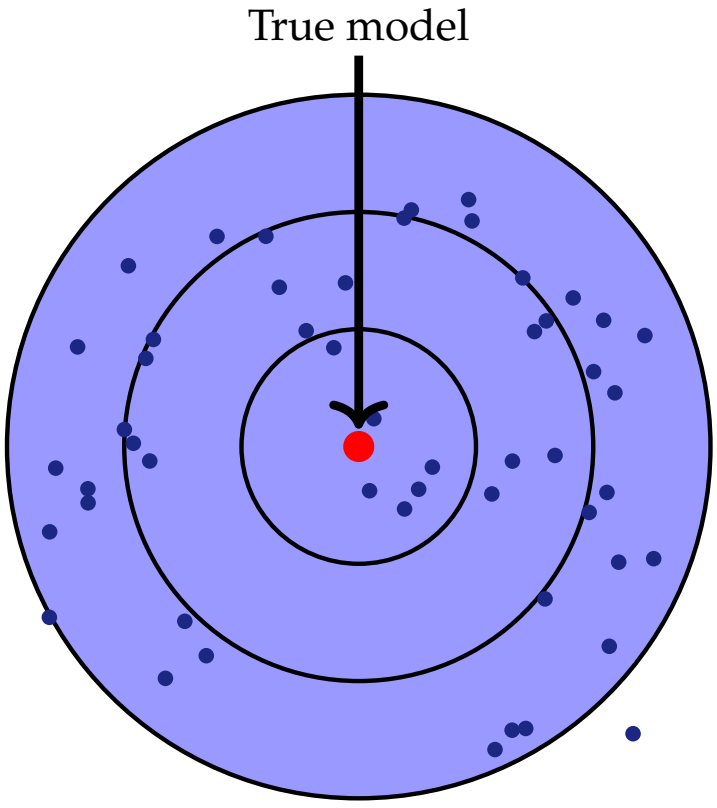
1. *Samsung QLED TV | The Frame | Premium UHD TV*. http://www.samsung.com/global/tv/blog/why-are-quantum-dot-displays-so-good/.
2. Manders, J. R. *et al.* 8.3: Distinguished Paper: Next-Generation Display Technology: Quantum-Dot LEDs. *SID Symposium Digest of Technical Papers* **46,** 73 (2015).
3. Brunner, K., Abstreiter, G., Böhm, G., Tränkle, G. & Weimann, G. Sharp-Line Photoluminescence and Two-Photon Absorption of Zero-Dimensional Biexcitons in a GaAs/AlGaAs Structure. *Physical Review Letters* **73,** 1138 (1994).
4. *Machine Learning - Stanford University*. http://mlclass.stanford.edu/.
5. Carleo, G. & Troyer, M. Solving the Quantum Many-Body Problem with Artificial Neural Networks. *Science* **355,** 602 (2017).
6. Flugsrud, V. M. *Solving Quantum Mechanical Problems with Machine Learning*. MA thesis (2018).
7. Pfau, D., Spencer, J. S., Matthews, A. G. G. & Foulkes, W. M. C. Ab-Initio Solution of the Many-Electron Schrödinger Equation with Deep Neural Networks. arXiv: 1909.02487 (2019).
8. Mariadason, A. A. *Quantum Many-Body Simulations of Double Dot System*. MA thesis (2018).
9. Taut, M. Two electrons in an external oscillator potential: Particular analytic solutions of a Coulomb correlation problem. *Physical Review A* **48,** 3561 (1993).
10. Høgberget, J. *Quantum Monte-Carlo Studies of Generalized Many-body Systems*. MA thesis (2013).

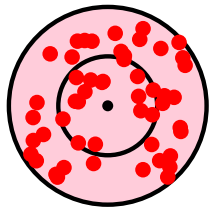
Machine Learning

Bias-variance Decomposition



Different models

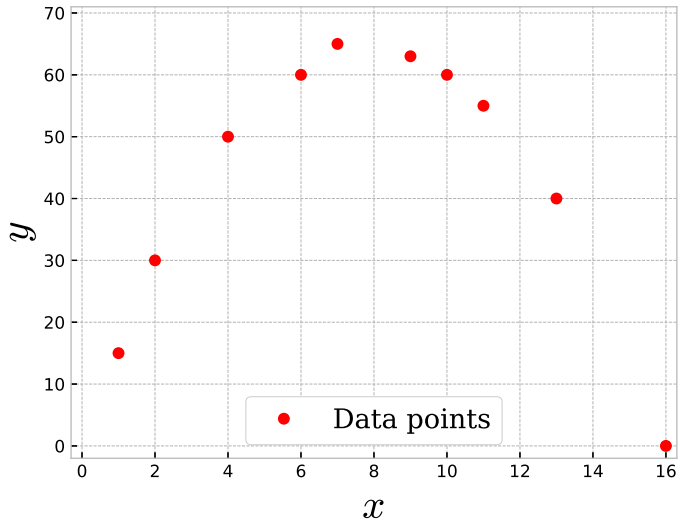




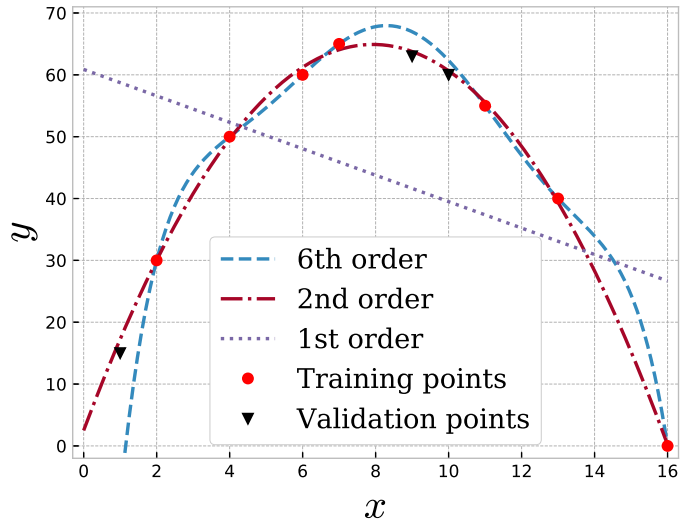
Low variance high-bias model

High variance low-bias model

Polynomial Regression



(y) Data set



(z) Data set with fitted polynomials

Ordinary Linear Regression

The output from ordinary linear regression is given by

$$f(x_i) = \sum_{j=0}^{p} X_{ij}(x_i)\theta_j.$$

Using the mean square error as the cost function, we obtain

$$\mathcal{C}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} X_{ij}\theta_j \right)^2,$$

which is equivalent to

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}.$$

Singular Value Decomposition

Decomposing a matrix into three matrices

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T.$$

Ridge Regression

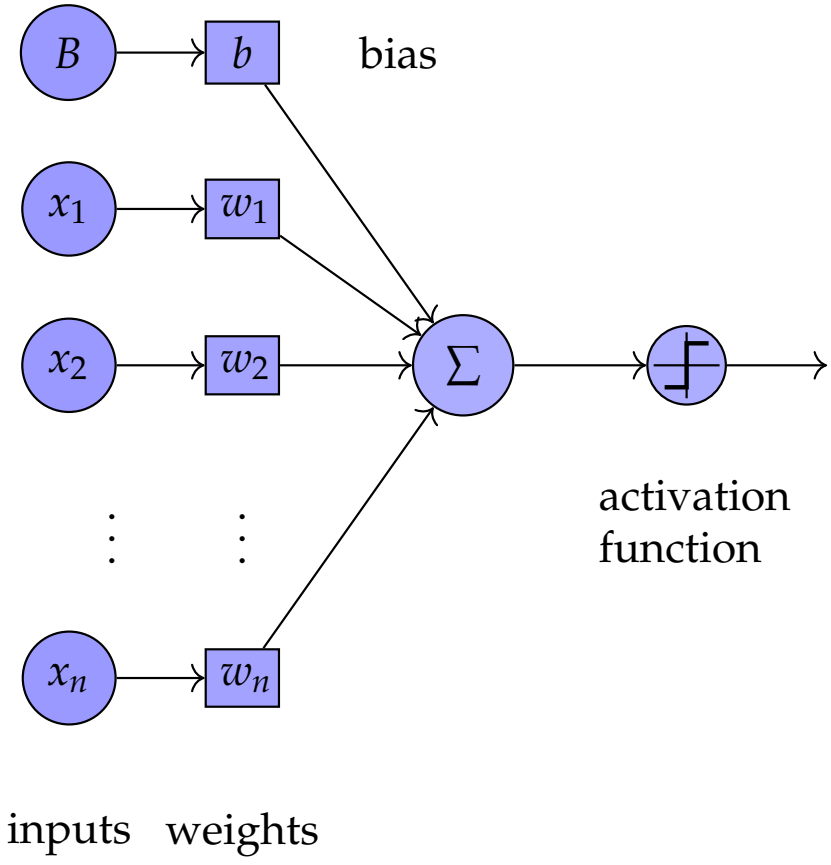
$$\mathcal{C}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{p} X_{ij}\theta_j\right)^2 + \lambda\sum_{j=1}^{p}|\theta_j|^2,$$

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\mathbb{1})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

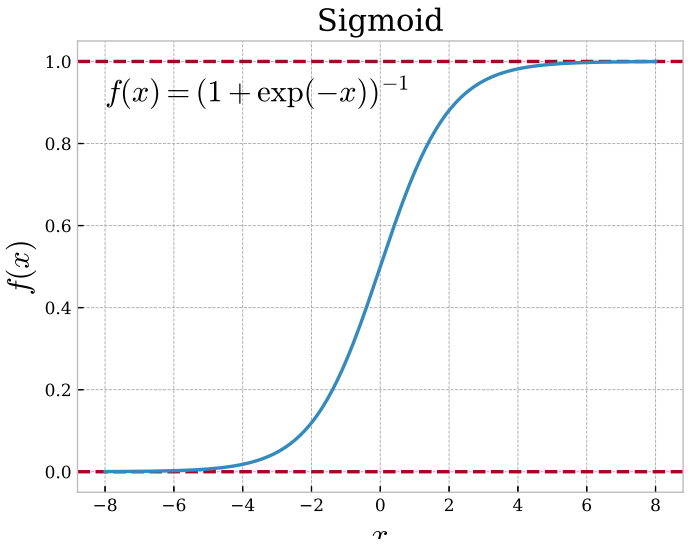
Lasso Regression

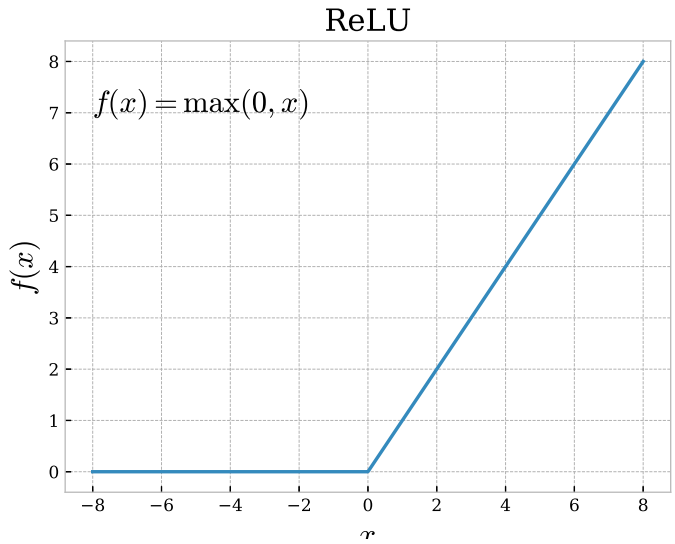
$$\mathcal{C}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} X_{ij}\theta_j \right)^2 + \lambda \sum_{j=1}^{p} |\theta_j|.$$

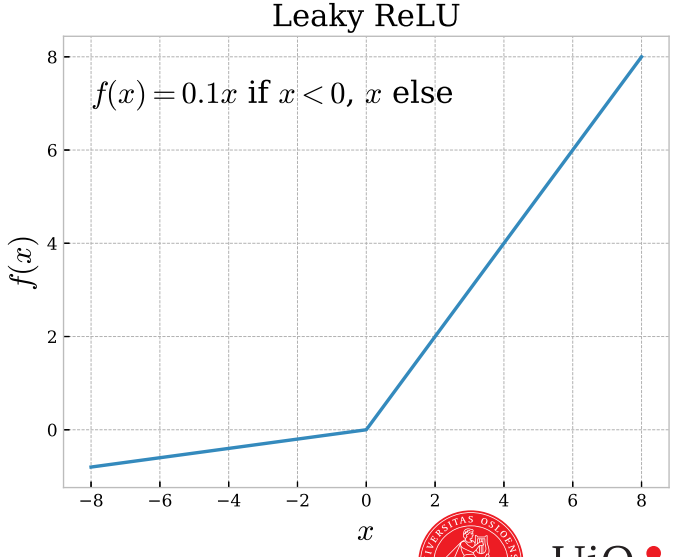
Logistic Regression

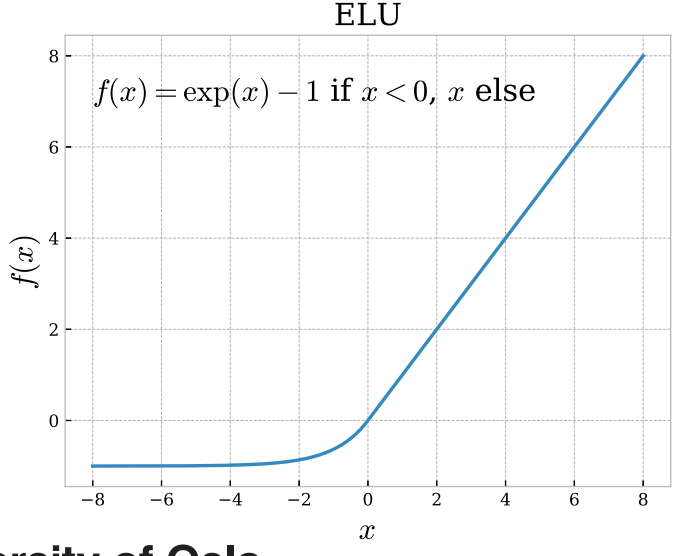


Activation Functions









FNN: Forward Phase

The output before activation reads

$$z_j^{(l+1)} = \sum_{i=1}^{N_l+1} a_i^{(l)} w_{ij}^{(l)}$$

while after the activation we have

$$a_j^{(l+1)} = f(z_j^{(l+1)}) = f\left(\sum_{i=1}^{N_l+1} a_i^{(l)} w_{ij}^{(l)}\right)$$

Backpropagation

$$\frac{\partial \mathcal{C}(\boldsymbol{w})}{\partial w_{jk}^{(l)}} = \delta_j^{(l+1)} a_k^{(l)}$$

$$\delta_j^{(l)} = \sum_k \delta_k^{(l+1)} w_{kj}^{(l)} f'(z_j^{(l)})$$