Studies of Quantum Dots using Machine Learning

Even Marius Nordhagen

University of Oslo

evenmn@fys.uio.no

December 16, 2019

Studies of Quantum Dots using Machine Learning

Even Marius Nordhagen

University of Oslo

evenmn@fys.uio.no

December 16, 2019

- Welcome
- Compress

Outline

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

Outline

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

Results most interesting outcome

Motivation

Studies of Quantum Dots using Machine Learning

- Closer look at the title
- Decompose → Quantum dots

Studies of Quantum Dots using Machine Learning

- Closer look at the title
- Decompose → Quantum dots

Quantum Dots

- What are quantum dots?

Quantum Dots

- What are quantum dots?

1. – Small particles consisting of a bunch of subatomic particles confined in a external potential
   – Artificial atoms

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

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

1. – Small particles consisting of a bunch of subatomic particles confined in a external potential
   – Artificial atoms
2. – Emit one wave length → Samsung
   – Quantum circuits and quantum computers
   – Does also encourage → more specific
   – Wigner crystallization

Studies of Quantum Dots using Machine Learning

- Discussed quantum dot systems → Last term Machine learning

Studies of Quantum Dots using Machine Learning

- Discussed quantum dot systems → Last term 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".

1. – Many of you probably know
   – For our work → Definition by Stanford university
   – Has experienced a booming popularity over the past decade → neural networks
   – The field of image recognition (CNNs)
   – Voice recognition (RNNs)
   – Nothing to do with quantum mechanical problems

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".
Image recognition

1. – Many of you probably know
   – For our work → Definition by Stanford university
   – Has experienced a booming popularity over the past decade → neural networks
   – The field of image recognition (CNNs)
   – Voice recognition (RNNs)
   – Nothing to do with quantum mechanical problems

Machine Learning

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

Machine Learning
"Machine learning is the science of getting computers to act without being explicitly programmed".
Image recognition
Natural language processing

1. – Many of you probably know
   – For our work → Definition by Stanford university
   – Has experienced a booming popularity over the past decade → neural networks
   – The field of image recognition (CNNs)
   – Voice recognition (RNNs)
   – Nothing to do with quantum mechanical problems

Machine Learning + Quantum Mechanics

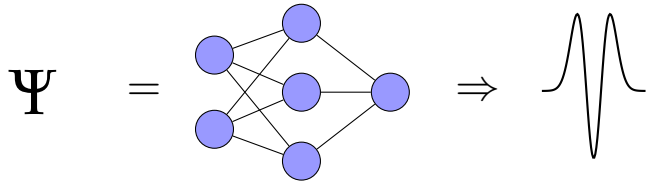
- Neural networks are eminent function approximators

Machine Learning + Quantum Mechanics
- Neural networks are eminent function approximators

- Impressive power
- According to the universal function approximation theorem
- Let the wave function be represented by a neural network
- Some popular quantum many-body methods, like VMC, are similar to machine learning algorithms
- Carleo & Troyer Ising model
- Flugsrud small quantum dots
- Pfau *et al.* neural networks to inestigate atoms and molecules

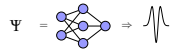
Machine Learning + Quantum Mechanics

- Neural networks are eminent function approximators



Machine Learning + Quantum Mechanics

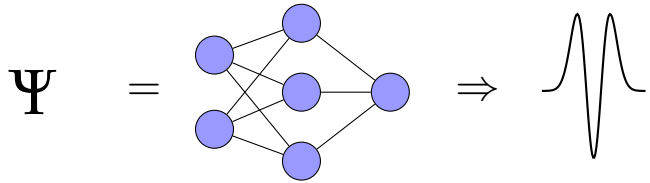
- Neural networks are eminent function approximators



- Impressive power
- According to the universal function approximation theorem
- Let the wave function be represented by a neural network
- Some popular quantum many-body methods, like VMC, are similar to machine learning algorithms
- Carleo & Troyer Ising model
- Flugsrud small quantum dots
- Pfau *et al.* neural networks to inestigate atoms and molecules

Machine Learning + Quantum Mechanics

- Neural networks are eminent function approximators



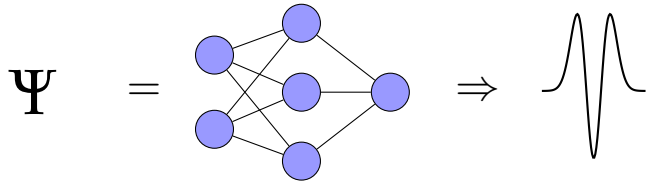
- Existing methods are reminiscent of machine learning algorithms

Machine Learning + Quantum Mechanics
- Neural networks are eminent function approximators
- Existing methods are reminiscent of machine learning algorithms

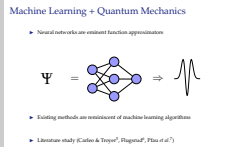
- Impressive power
- According to the universal function approximation theorem
- Let the wave function be represented by a neural network
- Some popular quantum many-body methods, like VMC, are similar to machine learning algorithms
- Carleo & Troyer Ising model
- Flugsrud small quantum dots
- Pfau *et al.* neural networks to inestigate atoms and molecules

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])



- Impressive power
- According to the universal function approximation theorem
- Let the wave function be represented by a neural network
- Some popular quantum many-body methods, like VMC, are similar to machine learning algorithms
- Carleo & Troyer Ising model
- Flugsrud small quantum dots
- Pfau *et al.* neural networks to inestigate atoms and molecules

Ethics in Science

- Respect for other's work

Ethics in Science

- Respect for other's work

- Whenever others work is used → Credit sources → text...
- When doing experiments → Always describe details in a such way
- Raw files are available on zenodo
- Open source code

Ethics in Science

- Respect for other's work
- Reproducibility

Ethics in Science

- Respect for other's work
- Reproducibility

- Whenever others work is used → Credit sources → text...
- When doing experiments → Always describe details in a such way
- Raw files are available on zenodo
- Open source code

Quantum Theory

Now we will give a breif introduction to the essential quantum theory.

The Schrödinger Equation

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

- Describes the mechanics of all QM systems
- Stationary systems → Time-independent SE
- Linear algebra terms
- Configuration interaction
- One year on solving
- Difficult to solve bco interactions between particles

The Schrödinger Equation

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

The Schrödinger Equation

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

$$\Downarrow$$

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

- Describes the mechanics of all QM systems
- Stationary systems → Time-independent SE
- Linear algebra terms
- Configuration interaction
- One year on solving
- Difficult to solve bco interactions between particles

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.

- To obtain the ground state energy
- States → minimizing

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

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

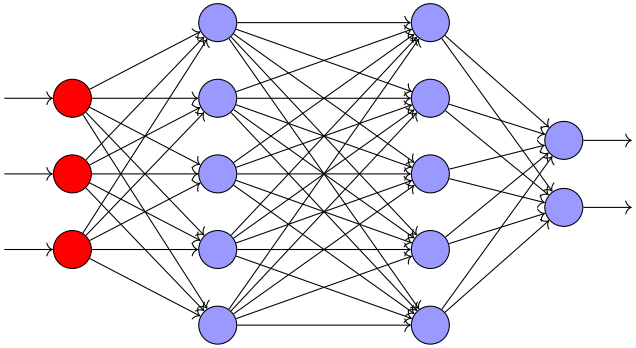
which are the magic numbers.

- Hamiltonian of the circular quantum dots consisting of electrons. In natural units.
- The magic numbers give the number of electron in each shell. Looked at closed-shell systems only

Machine Learning Theory

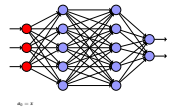
Now over to the machine learning theory. We have already mentioned the artificial neural networks, and we will now look at how they actually work.

Feed-forward Neural Network (FNN)



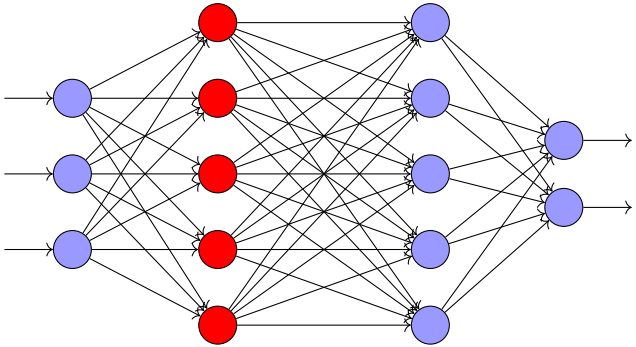
$a_0 = x$

Feed-forward Neural Network (FNN)



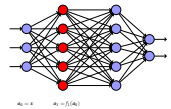
- FNNs are among the most popular neural networks
- Here a FNN
- Many different architectures
- Data set $\rightarrow$ propagating
-

Feed-forward Neural Network (FNN)



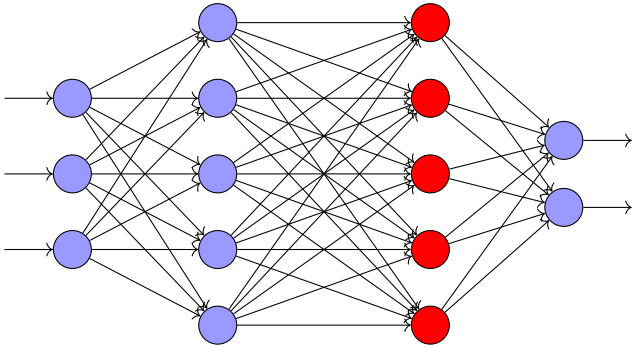
$a_0 = x$ $\quad$ $a_1 = f_1(a_0)$

Feed-forward Neural Network (FNN)



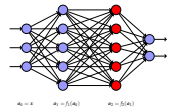
- FNNs are among the most popular neural networks
- Here a FNN
- Many different architectures
- Data set $\rightarrow$ propagating
-

Feed-forward Neural Network (FNN)



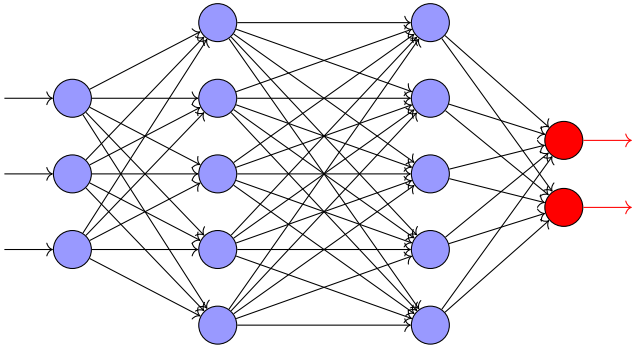
$a_0 = x$ $a_1 = f_1(a_0)$ $a_2 = f_2(a_1)$

Feed-forward Neural Network (FNN)



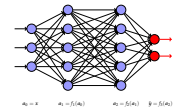
- FNNs are among the most popular neural networks
- Here a FNN
- Many different architectures
- Data set $\rightarrow$ propagating
-

Feed-forward Neural Network (FNN)



$$a_0 = x \qquad a_1 = f_1(a_0) \qquad a_2 = f_2(a_1) \qquad \tilde{y} = f_3(a_2)$$

Feed-forward Neural Network (FNN)



- FNNs are among the most popular neural networks
- Here a FNN
- Many different architectures
- Data set $\rightarrow$ propagating
-

Cost function

- The cost function defines the error

Cost function

- The cost function defines the error

- To decide how good the model performs
- Continuous model $\rightarrow$ MSE
- Want the error to be small $\rightarrow$ minimize the cost function

Cost function

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

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

- To decide how good the model performs
- Continuous model → MSE
- Want the error to be small → minimize the cost function

Cost function

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

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

- Attempt to minimize the cost function

Cost function

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

- To decide how good the model performs
- Continuous model → MSE
- Want the error to be small → minimize the cost function

Optimization Algorithms

- Minimize the cost function

Optimization Algorithms
- Minimize the cost function

- For this, we use optimization algorithms
- Plenty of methods → tradeoff between simplicity and performance
- GD perhaps the simplest → move in the direction that minimizes the cost function
- We have used the ADAM optimizer, which is slightly more complex. Contains momentum

Optimization Algorithms

- Minimize the cost function
- The gradient descent method:

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

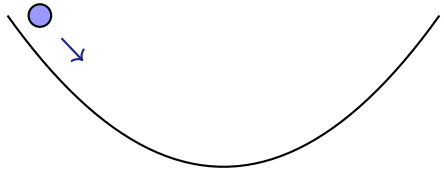
Optimization Algorithms
- Minimize the cost function
- The gradient descent method:
$\theta^{+} = \theta - \frac{\partial \mathcal{C}}{\partial \theta}$.

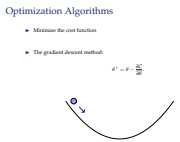
- For this, we use optimization algorithms
- Plenty of methods → tradeoff between simplicity and performance
- GD perhaps the simplest → move in the direction that minimizes the cost function
- We have used the ADAM optimizer, which is slightly more complex. Contains momentum

Optimization Algorithms

- Minimize the cost function
- The gradient descent method:

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



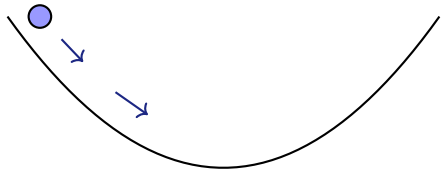


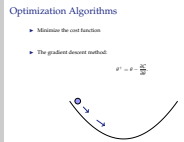
- For this, we use optimization algorithms
- Plenty of methods → tradeoff between simplicity and performance
- GD perhaps the simplest → move in the direction that minimizes the cost function
- We have used the ADAM optimizer, which is slightly more complex. Contains momentum

Optimization Algorithms

- Minimize the cost function
- The gradient descent method:

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



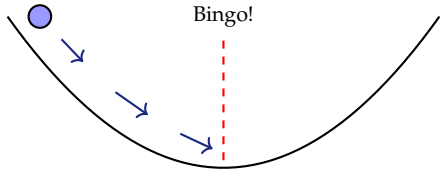


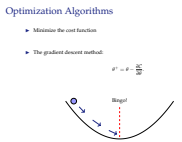
- For this, we use optimization algorithms
- Plenty of methods → tradeoff between simplicity and performance
- GD perhaps the simplest → move in the direction that minimizes the cost function
- We have used the ADAM optimizer, which is slightly more complex. Contains momentum

Optimization Algorithms

- Minimize the cost function
- The gradient descent method:

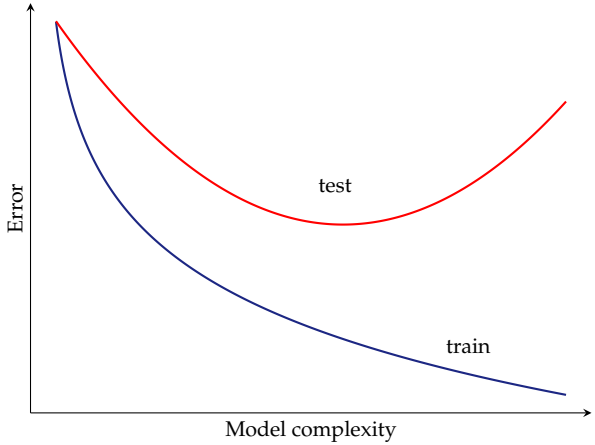
$$\theta^+ = \theta - \frac{\partial \mathcal{C}}{\partial \theta}.$$



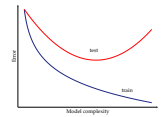


- For this, we use optimization algorithms
- Plenty of methods → tradeoff between simplicity and performance
- GD perhaps the simplest → move in the direction that minimizes the cost function
- We have used the ADAM optimizer, which is slightly more complex. Contains momentum

Find Appropriate Complexity

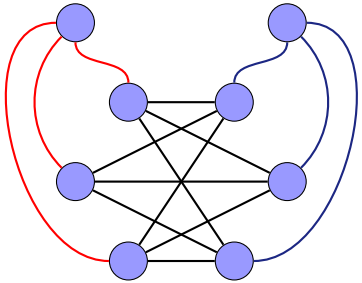


Find Appropriate Complexity

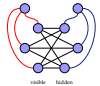


- Different architectures with different complexities
- Spkit data set in training and test set
- Want to minimize test error → trial and error

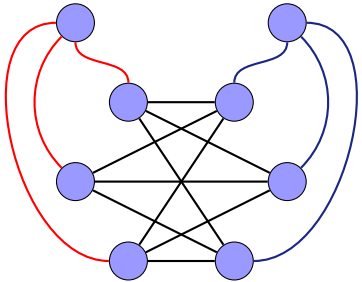
Restricted Boltzmann Machines



Restricted Boltzmann Machines



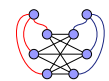
Restricted Boltzmann Machines



visible hidden

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

Restricted Boltzmann Machines



- What we have used in our work
- Energy based model
- Differs from FNN $\rightarrow$ obey unsupervised $\rightarrow$ no labeled data
- Finds the most likely configuration by minimizing the system energy

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

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

Probability Distribution

The joint probability distribution is given by the Boltzmann distribution:

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

The marginal distribution of the visible units is given by

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

- Named Boltzmann machine because of the joint probability distribution
- Find the marginal distribution of the visible units by integrating over all the hidden units

Methods

The applied methods will be discussed briefly

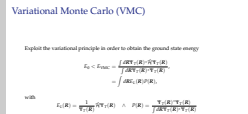
Variational Monte Carlo (VMC)

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

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



- Our work is based on VMC
- Exploits variational principle
- We start with rewriting the expression in terms of the local energy and the probability density function

Monte Carlo Integration

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

$$E_{\text{VMC}} = \int d\boldsymbol{R} E_L(\boldsymbol{R}) P(\boldsymbol{R}),$$
$$\approx \frac{1}{M} \sum_{i=1}^{M} E_L(\boldsymbol{R}_i).$$

Monte Carlo Integration

- The reason $\rightarrow$ On the form of a general expectation value
- Can be solved by Monte Carlo integration
- Only gives an energy
- Find the ground state energy by adjusting the trial wave fucntion with respect to minimizing the energy. Repeat exercise. When the energy has converged, we have a ground state energy estimate.

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.

- Arbitrary function → Few requirements → electron systems
- Standard Slater-Jastrow function

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,

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

- Hermite functions often used for circular quantum dots → quantities
- An important finding
- Slater determinant exchange correlation

Restricted Boltzmann Machine

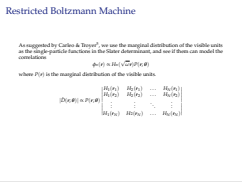
As suggested by Carleo & Troyer[5], we use the marginal distribution of the visible units as the single-particle functions in the Slater determinant, and see if them can model the correlations

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

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

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

Restricted Boltzmann Machine



- Our contribution → Marginal distribution
- Gives us a wave function where less physical intuition is needed
- Interesting because many systems, for instance nuclear systems, have very complex wave functions. We struggle with investigating those systems as we do not have the needed physical intuition

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

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

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

- Two Jastrow factors investigated
- Interesting as we want to see how much physical intuition we need to get acceptable results
- PJ is a complication of the simple Jastrow

Our Trial Wave Function Ansätze

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

Present ansätze

Our Trial Wave Function Ansätze

Software

Three principal aims

- Efficient

Software

Three principal aims

- Efficient
- Flexible

Software
Three principal aims
- Efficient
- Flexible

Software

Three principal aims

- Efficient
- Flexible
- Straightforward to use

- Before we move on to the results
- Develop from scratch → new approach RBM
- Object-oriented programming C++ open source

Software
Three principal aims
- Efficient
- Flexible
- Straightforward to use

Results

Finally ready to present the results

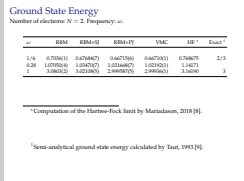
Ground State Energy

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].



- First look at ground state energy estimates
- Two electrons $\rightarrow$ some analytical results
- The RBM ansatz provides energy close to exact for $\omega = 1$
- Also closer than HF, which is interesting as they both attempt to approximate the wave function with a Slater determinant
- Same can be observed for $\omega = 1/6$
- For $\omega = 0.28$, the RBM+PJ provides significantly lower energu than VMC $\rightarrow$ indicates a better estimate

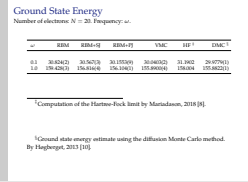
Ground State Energy

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].

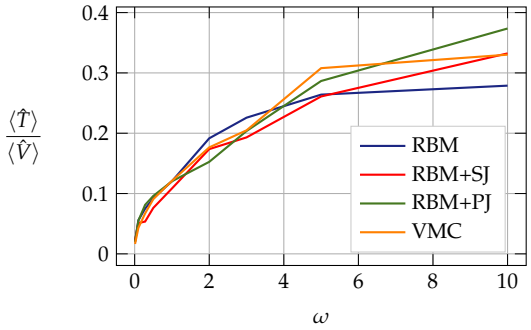


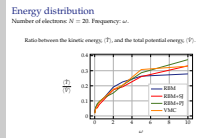
- No longer analytical results → Rely on diffusion Monte Carlo results which are expected to be almost exact
- The RBM+PJ ansatz now provides energies that are slightly larger than the VMC energy
- Large number of variational parameters (860)
- Low frequencies → RBM¡HF → better to model interactions
- High frequencies → RBM¿HF → Interactions less important

Energy distribution

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

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

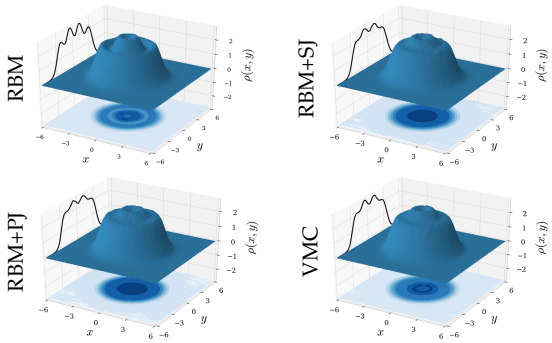


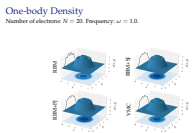


- Still looking at 20 electrons → Distribution between kinetic and potential energy
- Low frequency → Potential energy dominates over kinetic energy
- High frequency → Ansätze differ → Different electron configurations

One-body Density

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

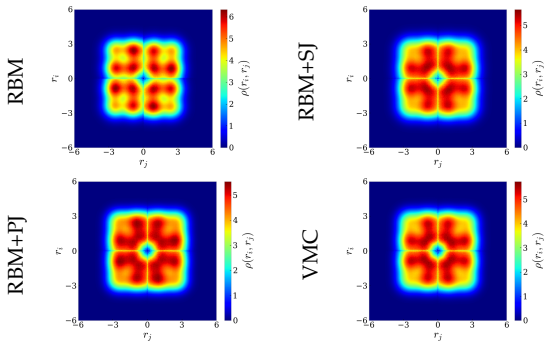


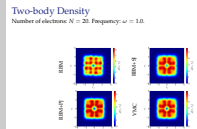


- One-body density describes how electrons distribute throughout the space
- Ansätze with Jastrow factor almost identical
- RBM provides more distinct peaks $\rightarrow$ model the electrons

Two-body Density

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

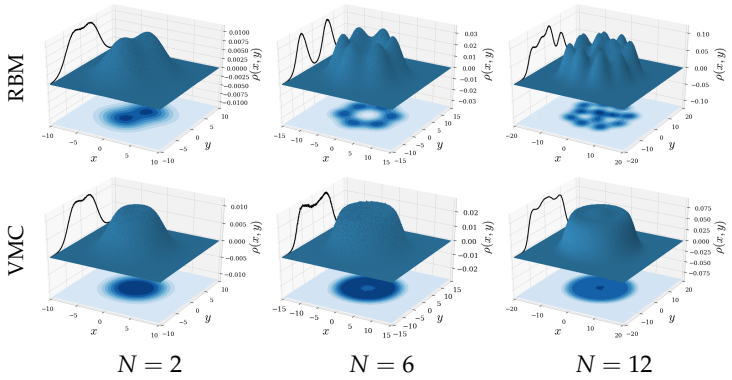


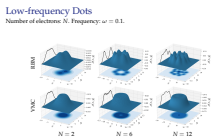


- Two-body density describes how the electrons distribute pairwise
- The same can be observed here, where RBM stands out

Low-frequency Dots

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

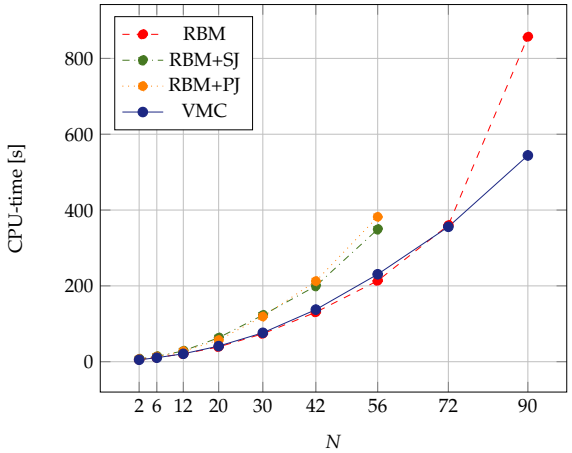




- Move on to low=frequency dots
- Interesting since RBM gives very different profiles from VMC
- N=peaks
- Indicates very localized electrons
- Indicates that energy is not the best way of evaluate an ansatz

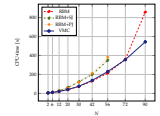
Computational Cost

Number of electrons: N.



UiO : University of Oslo

Computational Cost
Number of electrons: N.



- RBM and VMC pairwise
- RBM explodes for large systems → 10,000 parameters vs 2
- RBM+PJ and RBM+SJ pairwise most computationally intensive → No point to choose RBM+PJ considering the other results

Conclusions

Now we will address some brief 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

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

Even though the energy is quite similar for the various methods, the ratio between kinetic and potential energy reveals that the distribution between kinetic and potential energy is different. To change the potential energy, the electron configuration has to be different, meaning that the various ansätze provide different particle positions. We have also observed that different ansätze provide very different electron density plots even when the energy is similar. This indicates that the energy might not be the best way to evaluate various ansätze.

Future Work

- Investigate restricted Boltzmann machines with other architectures

Future Work

- Investigate restricted Boltzmann machines with other architectures

Future Work

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

└Future Work

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

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

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

Quantum dots are very simple systems, but the main application of this method is on systems where we don't have much information about the wave function. It is therefore obvious that it should be applied on more complex systems. It has shown that it can model electron-electron correlations, so it might be able to model three-body correlations as well, which is found in nuclear systems.

Thank you!

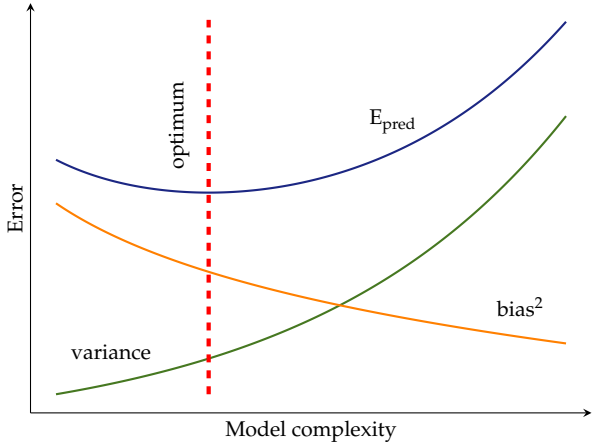
Thank you all for listening! Since I have a few more minutes, I will show how the developed software can be used.

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).

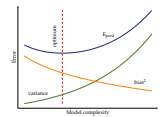
Machine Learning

Bias-variance Decomposition

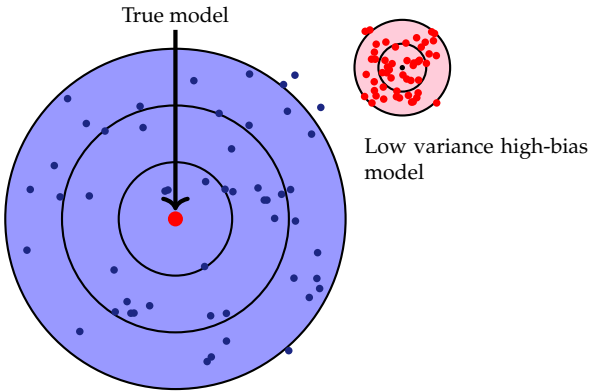


UiO : University of Oslo

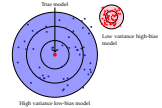
Bias-variance Decomposition



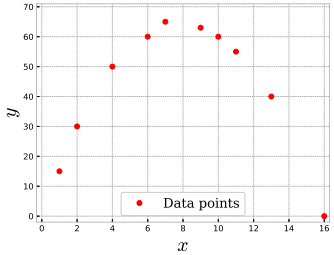
Different models



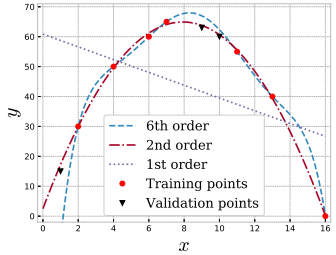
Different models



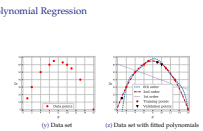
Polynomial Regression



(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

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

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

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

Singular Value Decomposition

Decomposing a matrix into three matrices

$X = U\Sigma V^T$.

Ridge Regression

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

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

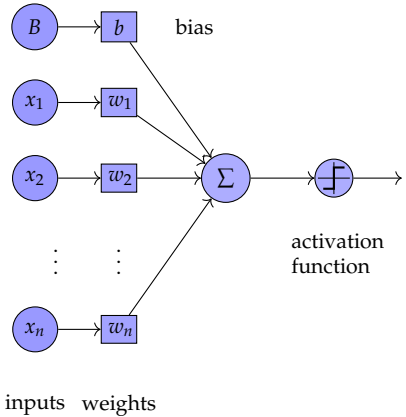
Ridge Regression

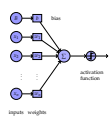
Lasso Regression

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

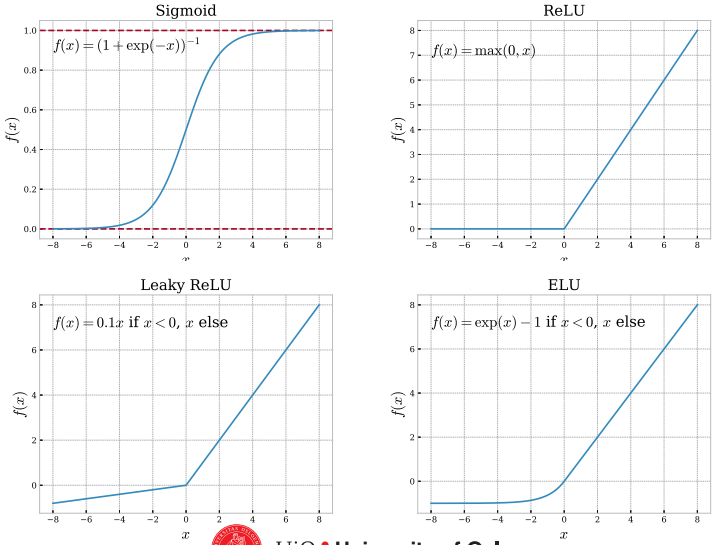
Lasso Regression

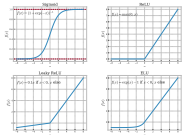
Logistic Regression





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

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

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

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

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

Backpropagation