EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Storage capacity of a Quantum Perceptron

Giovanni Gramegna

In collaboration with Fabio Benatti and Stefano Mancini

SM&FT 2022

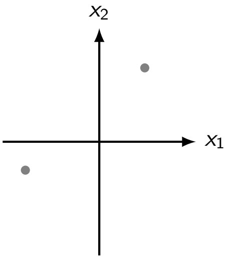
December 20, 2022

Outline

- Basic element of neural networs: perceptron
  - Linear separability problem
- Storage capacity
  - Geometric approach and Cover counting argument
  - Statistical Physics approach
- Quantum Perceptron
  - Model
  - Methods (statistical physics approach)
  - Results

The classical perceptron

- The classical perceptron realizes the mapping input-output $\boldsymbol{x} \in \mathbb{R}^N \mapsto \sigma \in \{-1, 1\}$, via



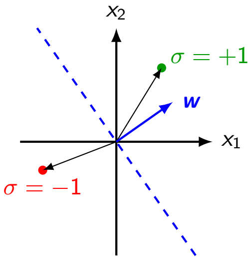
F. Rosenblatt, "The Perceptron: A perceiving and recognizing automaton", Tech. Rep. Inc. Report No. 85-460-1 (Cornell Aeronautical Laboratory, 1957)

The classical perceptron

- The classical perceptron realizes the mapping input-output $\boldsymbol{x} \in \mathbb{R}^N \mapsto \sigma \in \{-1, 1\}$, via

$$\sigma = \operatorname{sgn}\left(\frac{\boldsymbol{w} \cdot \boldsymbol{x}}{\|\boldsymbol{w}\|}\right) = \operatorname{sgn}\left(\frac{1}{\|\boldsymbol{w}\|}\sum_{j=1}^{N} w_j x_j\right),$$

where $\boldsymbol{w} \in \mathbb{R}^N$ and $\operatorname{sgn}(z)$ is the sign function



F. Rosenblatt, "The Perceptron: A perceiving and recognizing automaton", Tech. Rep. Inc. Report No. 85-460-1 (Cornell Aeronautical Laboratory, 1957)

The classical perceptron

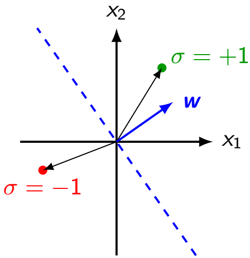
- The classical perceptron realizes the mapping input-output $\boldsymbol{x} \in \mathbb{R}^N \mapsto \sigma \in \{-1, 1\}$, via

$$\sigma = \operatorname{sgn}\left(\frac{\boldsymbol{w} \cdot \boldsymbol{x}}{\|\boldsymbol{w}\|}\right) = \operatorname{sgn}\left(\frac{1}{\|\boldsymbol{w}\|}\sum_{j=1}^{N} w_j x_j\right),$$

where $\boldsymbol{w} \in \mathbb{R}^N$ and $\operatorname{sgn}(z)$ is the sign function

- A classification $\{\boldsymbol{x}^\mu, \xi^\mu\}$, $\mu = 1, \ldots, p$ can be realized by a classical perceptron if for some $\boldsymbol{w} \in \mathbb{R}^N$ such that

$$\xi^\mu = \operatorname{sgn}\left(\frac{\boldsymbol{w} \cdot \boldsymbol{x}^\mu}{\|\boldsymbol{w}\|}\right), \qquad \forall \mu = 1, \ldots, p$$



F. Rosenblatt, "The Perceptron: A perceiving and recognizing automaton", Tech. Rep. Inc. Report No. 85-460-1 (Cornell Aeronautical Laboratory, 1957)

The classical perceptron

- The classical perceptron realizes the mapping input-output $\boldsymbol{x} \in \mathbb{R}^N \mapsto \sigma \in \{-1, 1\}$, via

$$\sigma = \mathrm{sgn}\left(\frac{\boldsymbol{w} \cdot \boldsymbol{x}}{\|\boldsymbol{w}\|}\right) = \mathrm{sgn}\left(\frac{1}{\|\boldsymbol{w}\|} \sum_{j=1}^{N} w_j x_j\right),$$

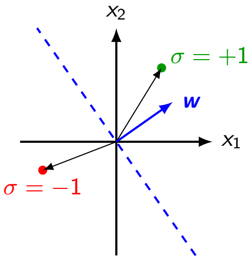
  where $\boldsymbol{w} \in \mathbb{R}^N$ and $\mathrm{sgn}(z)$ is the sign function

- A classification $\{\boldsymbol{x}^\mu, \xi^\mu\}$, $\mu = 1, \ldots, p$ can be realized by a classical perceptron if for some $\boldsymbol{w} \in \mathbb{R}^N$ such that

$$\xi^\mu = \mathrm{sgn}\left(\frac{\boldsymbol{w} \cdot \boldsymbol{x}^\mu}{\|\boldsymbol{w}\|}\right), \qquad \forall \mu = 1, \ldots, p$$

- Example: the XOR function can *not* be computed with a single perceptron:

$$\begin{aligned} \boldsymbol{x}^1 &= (-1, -1) & \xi^1 &= -1 \\ \boldsymbol{x}^2 &= (-1, 1) & \xi^2 &= 1 \\ \boldsymbol{x}^3 &= (1, -1) & \xi^3 &= 1 \\ \boldsymbol{x}^4 &= (1, 1) & \xi^4 &= -1 \end{aligned}$$



F. Rosenblatt, "The Perceptron: A perceiving and recognizing automaton", Tech. Rep. Inc. Report No. 85-460-1 (Cornell Aeronautical Laboratory, 1957)

The classical perceptron

- The classical perceptron realizes the mapping input-output $\boldsymbol{x} \in \mathbb{R}^N \mapsto \sigma \in \{-1, 1\}$, via

$$\sigma = \operatorname{sgn}\left(\frac{\boldsymbol{w} \cdot \boldsymbol{x}}{\|\boldsymbol{w}\|}\right) = \operatorname{sgn}\left(\frac{1}{\|\boldsymbol{w}\|}\sum_{j=1}^{N} w_j x_j\right),$$

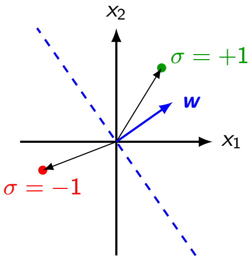
  where $\boldsymbol{w} \in \mathbb{R}^N$ and $\operatorname{sgn}(z)$ is the sign function

- A classification $\{\boldsymbol{x}^\mu, \xi^\mu\}$, $\mu = 1, \ldots, p$ can be realized by a classical perceptron if for some $\boldsymbol{w} \in \mathbb{R}^N$ such that

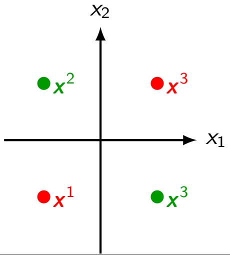
$$\xi^\mu = \operatorname{sgn}\left(\frac{\boldsymbol{w} \cdot \boldsymbol{x}^\mu}{\|\boldsymbol{w}\|}\right), \qquad \forall \mu = 1, \ldots, p$$

- Example: the XOR function can *not* be computed with a single perceptron:

$$\begin{aligned} \boldsymbol{x}^1 &= (-1, -1) & \xi^1 &= -1 \\ \boldsymbol{x}^2 &= (-1, 1) & \xi^2 &= 1 \\ \boldsymbol{x}^3 &= (1, -1) & \xi^3 &= 1 \\ \boldsymbol{x}^4 &= (1, 1) & \xi^4 &= -1 \end{aligned}$$



The XOR problem:



F. Rosenblatt, "The Perceptron: A perceiving and recognizing automaton", Tech. Rep. Inc. Report No. 85-460-1 (Cornell Aeronautical Laboratory, 1957)

A counting argument by Cover

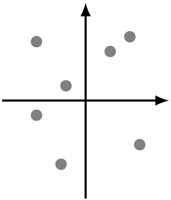
For a large number of inputs $N \to \infty$, how many patterns can we store?

A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?

- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

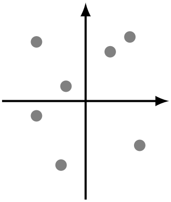
A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?

- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ $(\xi^\mu \in \{+1, -1\})$ are linearly separable?



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

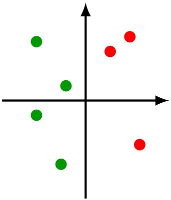
A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?

- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

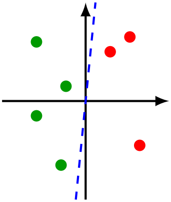
A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?

- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

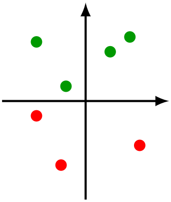
A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?

- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

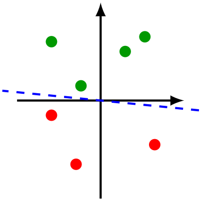
A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?

- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

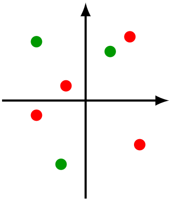
A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?

- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

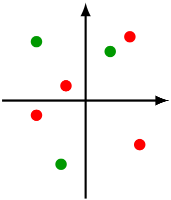
- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?



- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

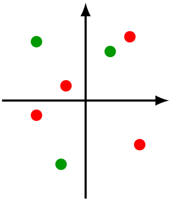
- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?

$$C(p, N) := \#\{\boldsymbol{\xi} \in \{-1, 1\}^p : \xi^\mu = \mathrm{sgn}(\boldsymbol{w} \cdot \boldsymbol{x}^\mu)\ \forall \mu = 1, \dots, p,\ \ \boldsymbol{w} \in \mathbb{R}^N\}$$

T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?



- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?

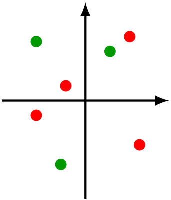
$$C(p, N) := \#\{\boldsymbol{\xi} \in \{-1, 1\}^p : \xi^\mu = \operatorname{sgn}(\boldsymbol{w} \cdot \boldsymbol{x}^\mu) \ \forall \mu = 1, \dots, p, \ \boldsymbol{w} \in \mathbb{R}^N\}$$

- Recursive formula:
  $C(p, N) = C(p-1, N) + C(p, N-1)$

T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?



- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

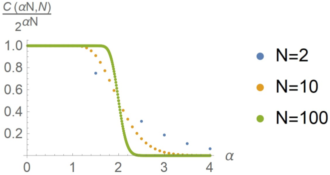
- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ $(\xi^\mu \in \{+1, -1\})$ are linearly separable?

$$C(p, N) := \#\{\boldsymbol{\xi} \in \{-1, 1\}^p : \xi^\mu = \mathrm{sgn}(\boldsymbol{w} \cdot \boldsymbol{x}^\mu)\ \forall \mu = 1, \dots, p,\ \ \boldsymbol{w} \in \mathbb{R}^N\}$$

- Recursive formula:
$C(p, N) = C(p-1, N) + C(p, N-1)$

$$\Rightarrow C(p, N) = 2 \sum_{j=0}^{N-1} \binom{p-1}{j}$$

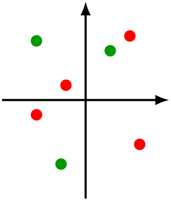
(with the convention $\binom{n}{m} = 0$ for $m > n$)



T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

A counting argument by Cover

For a large number of inputs $N \to \infty$, how many patterns can we store?



- Assume we have p patterns $\{\boldsymbol{x}^\mu\}$, in "generic positions":

$$\{\boldsymbol{x}^\mu, \boldsymbol{x}^\nu\} \text{ are lin. ind. } \forall \mu \neq \nu$$

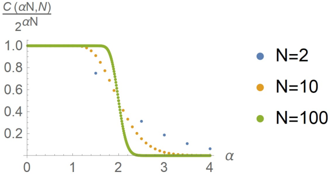
- How many of the 2^p possible choices of $\boldsymbol{\xi} = (\xi^\mu)_{\mu=1}^p$ ($\xi^\mu \in \{+1, -1\}$) are linearly separable?

$$C(p, N) := \#\{\boldsymbol{\xi} \in \{-1, 1\}^p : \xi^\mu = \mathrm{sgn}(\boldsymbol{w} \cdot \boldsymbol{x}^\mu)\ \forall \mu = 1, \ldots, p,\ \ \boldsymbol{w} \in \mathbb{R}^N\}$$

- Recursive formula:
$C(p, N) = C(p-1, N) + C(p, N-1)$

$$\Rightarrow C(p, N) = 2 \sum_{j=0}^{N-1} \binom{p-1}{j}$$

(with the convention $\binom{n}{m} = 0$ for $m > n$)



- Large N limit (keeping $\alpha = p/N$ fixed):

$$\frac{C(p, N)}{2^p} \xrightarrow{N \to \infty} \begin{cases} 1 & \text{if } \alpha < 2 \\ 0 & \text{if } \alpha > 2 \end{cases}$$

$$\alpha_c = \frac{p_c}{N} = 2$$

"storage capacity"

T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers. EC-14 (3): 326–334 (1965)

Statistical Physics Approach

- Gardner approach: relative volume of weights satisfying the classification condition

$$V_N\left(\{\xi^\mu, \boldsymbol{x}^\mu\}_{\mu=1}^p\right) = \int_{\mathbb{R}^N} \mathrm{d}\mu(\boldsymbol{w}) \prod_{\mu=1}^{p} \theta\left(\xi^\mu \frac{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}{\|\boldsymbol{w}\|} - \kappa\right)$$

$\kappa > 0$ stability parameter,

$$\mathrm{d}\mu(\boldsymbol{w}) = \left(\int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}\,\delta(\|\boldsymbol{w}\|^2 - N)\right)^{-1} \delta(\|\boldsymbol{w}\|^2 - N)$$

E. Gardner, "The space of interactions in neural network models", J. Phys. A: Math. Gen. **21** 257-270 (1988)

- Gardner approach: relative volume of weights satisfying the classification condition

$$V_N\left(\{\xi^\mu, \boldsymbol{x}^\mu\}_{\mu=1}^p\right) = \int_{\mathbb{R}^N} \mathrm{d}\mu(\boldsymbol{w}) \prod_{\mu=1}^p \theta\left(\xi^\mu \frac{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}{\|\boldsymbol{w}\|} - \kappa\right)$$

$\kappa > 0$ stability parameter,

$$\mathrm{d}\mu(\boldsymbol{w}) = \left(\int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}\,\delta(\|\boldsymbol{w}\|^2 - N)\right)^{-1} \delta(\|\boldsymbol{w}\|^2 - N)$$

V_N is a "low-temperature partition function":

$$V_N = \lim_{\beta\to\infty} \int_{\mathbb{R}^N} \mathrm{d}\mu(\boldsymbol{w})\, \mathrm{e}^{-\beta E(\boldsymbol{w})}, \quad E(\boldsymbol{w}) = \sum_{\mu=1}^p \left[1 - \theta\left(\xi^\mu \frac{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}{\|\boldsymbol{w}\|} - \kappa\right)\right]$$

E. Gardner, "The space of interactions in neural network models", J. Phys. A: Math. Gen. **21** 257-270 (1988)

- Gardner approach: relative volume of weights satisfying the classification condition

$$V_N\left(\{\xi^\mu, \boldsymbol{x}^\mu\}_{\mu=1}^p\right) = \int_{\mathbb{R}^N} \mathrm{d}\mu(\boldsymbol{w}) \prod_{\mu=1}^{p} \theta\left(\xi^\mu \frac{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}{\|\boldsymbol{w}\|} - \kappa\right)$$

$\kappa > 0$ stability parameter,

$$\mathrm{d}\mu(\boldsymbol{w}) = \left(\int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}\,\delta(\|\boldsymbol{w}\|^2 - N)\right)^{-1} \delta(\|\boldsymbol{w}\|^2 - N)$$

V_N is a "low-temperature partition function":

$$V_N = \lim_{\beta\to\infty} \int_{\mathbb{R}^N} \mathrm{d}\mu(\boldsymbol{w})\, \mathrm{e}^{-\beta E(\boldsymbol{w})}, \quad E(\boldsymbol{w}) = \sum_{\mu=1}^{p}\left[1 - \theta\left(\xi^\mu \frac{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}{\|\boldsymbol{w}\|} - \kappa\right)\right]$$

- Idea from **spin glass theory**: "average" V_N over random configurations of patterns and classifications $\{\boldsymbol{x}^\mu, \xi^\mu\}_{\mu=1}^p$

$$\mathrm{P}\left(x_j^\mu = \pm 1\right) = \frac{1}{2}, \qquad \mathrm{P}\left(\xi^\mu = \pm 1\right) = \frac{1}{2}$$

E. Gardner, "The space of interactions in neural network models", J. Phys. A: Math. Gen. **21** 257-270 (1988)

Storage capacity as a critical value

- The statistical relevant quantity is $\langle \ln V_N \rangle$ ($\langle \cdot \rangle$: average over $\{\boldsymbol{x}^\mu, \xi^\mu\}_{\mu=1}^p$)

Storage capacity as a critical value

- The statistical relevant quantity is $\langle \ln V_N \rangle$ ($\langle \cdot \rangle$: average over $\{\boldsymbol{x}^\mu, \xi^\mu\}_{\mu=1}^p$)

The random variable $\ln V_N$ is "self-averaging":

$\langle \ln V_N \rangle$ represents the typical value of $\ln V_N$

M. Shcherbina and B. Tirozzi, Rigorous Solution of the Gardner Problem, Commun. Math. Phys. **234**, 383–422 (2003)

Storage capacity as a critical value

- The statistical relevant quantity is $\langle \ln V_N \rangle$ ($\langle \cdot \rangle$: average over $\{\mathbf{x}^\mu, \xi^\mu\}_{\mu=1}^p$)

> **The random variable $\ln V_N$ is "self-averaging":**
> $\langle \ln V_N \rangle$ represents the typical value of $\ln V_N$

- A critical behaviour in $\alpha = \frac{p}{N}$ arises in the thermodynamic limit:
  - For $\alpha < \alpha_c(\kappa)$:

$$\lim_{N\to\infty} \frac{\langle \ln V_N \rangle}{N} = -\mathcal{F}(\alpha,\kappa) \quad \Rightarrow \quad V_N \simeq \mathrm{e}^{-N\mathcal{F}(\alpha,\kappa)}$$

  - For $\alpha > \alpha_c(\kappa)$:

$$\lim_{N\to\infty} \frac{\langle \ln V_N \rangle}{N} = -\infty \quad \Rightarrow \quad V_N \simeq o\left(\mathrm{e}^{-\mathcal{O}(N)}\right)$$

M. Shcherbina and B. Tirozzi, Rigorous Solution of the Gardner Problem, Commun. Math. Phys. **234**, 383–422 (2003)

Storage capacity as a critical value

- The statistical relevant quantity is $\langle \ln V_N \rangle$ ($\langle \cdot \rangle$: average over $\{\boldsymbol{x}^\mu, \xi^\mu\}_{\mu=1}^p$)

> **The random variable $\ln V_N$ is "self-averaging":**
> $\langle \ln V_N \rangle$ represents the typical value of $\ln V_N$

- A critical behaviour in $\alpha = \frac{p}{N}$ arises in the thermodynamic limit:
  - For $\alpha < \alpha_c(\kappa)$:

$$\lim_{N\to\infty} \frac{\langle \ln V_N \rangle}{N} = -\mathcal{F}(\alpha, \kappa) \quad \Rightarrow \quad V_N \simeq \mathrm{e}^{-N\mathcal{F}(\alpha,\kappa)}$$

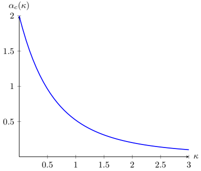
  - For $\alpha > \alpha_c(\kappa)$:

$$\lim_{N\to\infty} \frac{\langle \ln V_N \rangle}{N} = -\infty \quad \Rightarrow \quad V_N \simeq o\left(\mathrm{e}^{-\mathcal{O}(N)}\right)$$

- The critical value

$$\alpha_c(\kappa) = \left[\int_{-\kappa}^{\infty} \frac{\mathrm{d}t}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} (t+\kappa)^2\right]^{-1}$$

is the "storage capacity" of the perceptron.
Note: $\alpha_c(0) = 2$ ✓ Cover result



M. Shcherbina and B. Tirozzi, Rigorous Solution of the Gardner Problem, Commun. Math. Phys. **234**, 383–422 (2003)

Quantum perceptron: Model

- Input pattern $\boldsymbol{x}$ encoded in a gaussian state of the form:

$$|\Psi\rangle = \bigotimes_{j=1}^{N} |x_j\rangle, \qquad |x_j\rangle = \frac{1}{(2\pi\sigma_j^2)^{1/4}} \int_{-\infty}^{+\infty} dq_j \, \exp\left(-\frac{(q_j - x_j)^2}{4\sigma_j^2}\right) |q_j\rangle$$

- Squeezing operator:

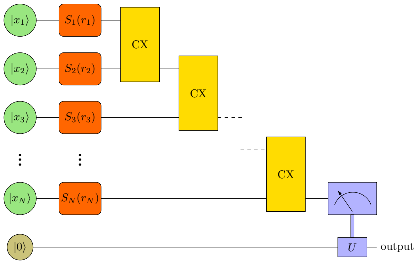
$$S_j(r_j) = \mathrm{e}^{\mathrm{i}\, r_j\,(q_j p_j + p_j q_j)}\,,\, \mathrm{e}^{-2r_j} = w_j$$

$$S_j(r)\,|q_j\rangle = \sqrt{w_j}\,|w_j q_j\rangle$$

- Controlled shift:

$$\mathrm{CX} := \exp\left(-\mathrm{i}\, q_j \otimes p_{j+1}\right)$$

$$\mathrm{CX}\,|q_j, q_{j+1}\rangle = |q_j, q_j + q_{j+1}\rangle$$



- Output state position eigenfunction:

$$\psi_{\boldsymbol{w},\boldsymbol{x}^\mu}(s) = \frac{1}{(2\pi \sum_j w_j^2 \sigma_j^2)^{1/4}} \exp\left(-\frac{(s - \boldsymbol{w}\cdot\boldsymbol{x}^\mu)^2}{4\sum_j w_j^2\sigma_j^2}\right)$$

F. Benatti, S. Mancini and S. Mangini, Continuous variable quantum perceptron, International Journal of Quantum Information **17**, 1941009 (2019)

- The probability to correctly classify the pattern μ is:

$$R^{\mu}(\kappa) = \int_{-\infty}^{+\infty} \mathrm{d}s \; P_{\boldsymbol{w},\boldsymbol{x}^{\mu},\sigma}(s) \theta \left(\xi^{\mu} \frac{s}{\|\boldsymbol{w}\|} - \kappa \right),$$

where

$$P_{\boldsymbol{w},\boldsymbol{x}^{\mu},\sigma}(s) = |\psi_{\boldsymbol{w},\boldsymbol{x}^{\mu},\sigma}(s)|^2 = \frac{1}{\sqrt{2\pi}\|\boldsymbol{w}\|\sigma} \exp\left(-\frac{(s - \boldsymbol{w}\cdot\boldsymbol{x}^{\mu})^2}{2\|\boldsymbol{w}\|^2\sigma^2} \right)$$

- Classical limit ($\sigma \to 0$):

$$P_{\boldsymbol{w},\boldsymbol{x}^{\mu},\sigma} \to \delta(s - \boldsymbol{x}\cdot\boldsymbol{w}), \qquad R^{\mu}(\kappa) = \theta\left(\xi^{\mu} \frac{\boldsymbol{w}\cdot\boldsymbol{x}^{\mu}}{\|\boldsymbol{w}\|} - \kappa \right)$$

- We introduce the upper bound ϵ on the acceptable error
- Relative volume in the quantum case:

$$V_N(\{\boldsymbol{x}^{\mu}, \xi^{\mu}\}_{\mu=1}^{p}) = \frac{1}{C_N} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w} \; \delta(\|\boldsymbol{w}\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R^{\mu}(\kappa) - 1 + \epsilon \right)$$

Methods: Replica method

The quantity $\langle \ln V_N \rangle$ is computed with the **replica trick**:

$$\langle \ln V_N \rangle = \lim_{n \to 0} \frac{\langle V_N^n \rangle - 1}{n}$$

- Compute $\langle V_N^n \rangle$ for n integer (*relatively easy*):

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \; \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) \right\rangle$$

  each $\boldsymbol{w}^\gamma$ can be interpreted as a “replica”
- Take the limit $n \to 0$ as an “analytic continuation”

Methods: Replica method

The quantity $\langle \ln V_N \rangle$ is computed with the **replica trick**:

$$\langle \ln V_N \rangle = \lim_{n \to 0} \frac{\langle V_N^n \rangle - 1}{n}$$

- Compute $\langle V_N^n \rangle$ for n integer (*relatively easy*):

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \ \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) \right\rangle$$

  each $\boldsymbol{w}^\gamma$ can be interpreted as a "replica"
- Take the limit $n \to 0$ as an "analytic continuation"

Not mathematically rigorous, but successful!

M. Mézard, G. Parisi, and M. A. Virasoro, "Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications", Vol. 9. World Scientific Publishing Company, 1987

Saddle-point approximation

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \ \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta \left(R_\gamma^\mu(\kappa) - 1 + \epsilon \right) \right\rangle$$

"Relatively easy":

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \; \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) \right\rangle$$

"Relatively easy":

$$\theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) = \int_{1-\epsilon}^{\infty} \frac{\mathrm{d}z_\gamma^\mu}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}y_\gamma^\mu \; \mathrm{e}^{\mathrm{i}y_\gamma^\mu z_\gamma^\mu} \mathrm{e}^{-\mathrm{i}y_\gamma^\mu R_\gamma^\mu(\kappa)}$$

Saddle-point approximation

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \; \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) \right\rangle$$

"Relatively easy":

$$\theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) = \int_{1-\epsilon}^{\infty} \frac{\mathrm{d}z_\gamma^\mu}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}y_\gamma^\mu \; \mathrm{e}^{\mathrm{i}y_\gamma^\mu z_\gamma^\mu} \mathrm{e}^{-\mathrm{i}y_\gamma^\mu R_\gamma^\mu(\kappa)}$$

$$\delta\left(\sum_{j=1}^{N} (w_j^\gamma)^2 - N\right) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}E_\gamma}{4\pi} \exp\left[\mathrm{i}\frac{E_\gamma}{2}\left(N - \sum_{j=1}^{N} (w_j^\gamma)^2\right)\right]$$

Saddle-point approximation

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \; \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) \right\rangle$$

"Relatively easy":

$$\theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) = \int_{1-\epsilon}^{\infty} \frac{\mathrm{d}z_\gamma^\mu}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}y_\gamma^\mu \; \mathrm{e}^{\mathrm{i}y_\gamma^\mu z_\gamma^\mu} \mathrm{e}^{-\mathrm{i}y_\gamma^\mu R_\gamma^\mu(\kappa)}$$

$$\delta\left(\sum_{j=1}^{N} (w_j^\gamma)^2 - N\right) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}E_\gamma}{4\pi} \exp\left[\mathrm{i}\frac{E_\gamma}{2}\left(N - \sum_{j=1}^{N} (w_j^\gamma)^2\right)\right]$$

$$\delta\left(q_{\gamma\delta} - \frac{1}{N}\sum_{j=1}^{N} w_j^\gamma w_j^\delta\right) = N \int_{-\infty}^{+\infty} \frac{\mathrm{d}F_{\gamma\delta}}{2\pi} \exp\left(-\mathrm{i}N q_{\gamma\delta} F_{\gamma\delta} + \mathrm{i}F_{\gamma\delta} \sum_{j=1}^{N} w_j^\gamma w_j^\delta\right)$$

Saddle-point approximation

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \; \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) \right\rangle$$

"Relatively easy":

$$\theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) = \int_{1-\epsilon}^{\infty} \frac{\mathrm{d}z_\gamma^\mu}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}y_\gamma^\mu \; \mathrm{e}^{\mathrm{i}y_\gamma^\mu z_\gamma^\mu} \mathrm{e}^{-\mathrm{i}y_\gamma^\mu R_\gamma^\mu(\kappa)}$$

$$\delta\left(\sum_{j=1}^{N} (w_j^\gamma)^2 - N\right) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}E_\gamma}{4\pi} \exp\left[\mathrm{i}\frac{E_\gamma}{2}\left(N - \sum_{j=1}^{N} (w_j^\gamma)^2\right)\right]$$

$$\delta\left(q_{\gamma\delta} - \frac{1}{N}\sum_{j=1}^{N} w_j^\gamma w_j^\delta\right) = N\int_{-\infty}^{+\infty} \frac{\mathrm{d}F_{\gamma\delta}}{2\pi} \exp\left(-\mathrm{i}Nq_{\gamma\delta}F_{\gamma\delta} + \mathrm{i}F_{\gamma\delta}\sum_{j=1}^{N} w_j^\gamma w_j^\delta\right)$$

yields:

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \int \left(\prod_{\gamma=1}^{n} \mathrm{d}E_\gamma\right)\left(\prod_{\substack{\gamma,\delta=1 \\ \gamma<\delta}}^{n} \mathrm{d}q_{\gamma\delta}\,\mathrm{d}F_{\gamma\delta}\right) \mathrm{e}^{NG(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\})}$$

Saddle-point approximation

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^{n} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{w}^\gamma \; \delta(\|\boldsymbol{w}^\gamma\|^2 - N) \prod_{\mu=1}^{p} \theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) \right\rangle$$

"Relatively easy":

$$\theta\left(R_\gamma^\mu(\kappa) - 1 + \epsilon\right) = \int_{1-\epsilon}^{\infty} \frac{\mathrm{d}z_\gamma^\mu}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}y_\gamma^\mu \; \mathrm{e}^{\mathrm{i}y_\gamma^\mu z_\gamma^\mu} \mathrm{e}^{-\mathrm{i}y_\gamma^\mu R_\gamma^\mu(\kappa)}$$

$$\delta\left(\sum_{j=1}^{N} (w_j^\gamma)^2 - N\right) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}E_\gamma}{4\pi} \exp\left[\mathrm{i}\frac{E_\gamma}{2}\left(N - \sum_{j=1}^{N} (w_j^\gamma)^2\right)\right]$$

$$\delta\left(q_{\gamma\delta} - \frac{1}{N}\sum_{j=1}^{N} w_j^\gamma w_j^\delta\right) = N \int_{-\infty}^{+\infty} \frac{\mathrm{d}F_{\gamma\delta}}{2\pi} \exp\left(- \mathrm{i}N q_{\gamma\delta} F_{\gamma\delta} + \mathrm{i}F_{\gamma\delta} \sum_{j=1}^{N} w_j^\gamma w_j^\delta\right)$$

yields:

$$\langle V_N^n \rangle = \frac{1}{C_N^n} \int \left(\prod_{\gamma=1}^{n} \mathrm{d}E_\gamma\right)\left(\prod_{\substack{\gamma,\delta=1 \\ \gamma<\delta}}^{n} \mathrm{d}q_{\gamma\delta}\, \mathrm{d}F_{\gamma\delta}\right) \mathrm{e}^{NG(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\})}$$

Saddle point approximation:

$$\langle V_N^n \rangle \simeq \frac{1}{C_N^n} \mathrm{e}^{NG(z_S)} \sqrt{\frac{2\pi}{N|\det G''(z_S)|}},$$

where z_S is the saddle point

Giovanni Gramegna

$$G(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\}) = \alpha G_1(\{q_{\gamma\delta}\}) + G_2(\{F_{\gamma\delta}\},\{E_\gamma\}) + G_3(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\})$$

Replica symmetry

$$G(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\}) = \alpha G_1(\{q_{\gamma\delta}\}) + G_2(\{F_{\gamma\delta}\},\{E_\gamma\}) + G_3(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\})$$

$$G_1(\{q_{\gamma\delta}\}) = \ln\left[\int_{1-\epsilon}^{\infty}\left(\prod_{\gamma=1}^{n}\frac{\mathrm{d}z_\gamma}{2\pi}\right)\int\left(\prod_{\gamma=1}^{n}\frac{\mathrm{d}\lambda_\gamma \mathrm{d}y_\gamma \mathrm{d}\omega_\gamma}{2\pi}\right)\mathrm{e}^{K(\{\lambda_\gamma\},\{y_\gamma\},\{\omega_\gamma\},\{q_{\gamma\delta}\})}\right]$$

$$K(\{\lambda_\gamma\},\{y_\gamma\},\{\omega_\gamma\},\{q_{\gamma\delta}\}) \equiv \mathrm{i}\sum_{\gamma=1}^{n} y_\gamma\left[z_\gamma - \Phi(\lambda_\gamma)\right] - \mathrm{i}\sum_{\gamma=1}^{n}\left(\frac{\kappa}{\sigma}+\lambda_\gamma\right)\omega_\gamma^\mu - \frac{1}{2\sigma^2}\sum_{\gamma,\delta=1}^{n} q_{\gamma\delta}\omega_\gamma\omega_\delta$$

$$G_2(\{F_{\gamma\delta}\},\{E_\gamma\}) = \ln\left[\int\left(\prod_{\gamma=1}^{n}\mathrm{d}w^\gamma\right)\exp\left(-\frac{\mathrm{i}}{2}\sum_{\gamma=1}^{n}E_\gamma (w^\gamma)^2 + \mathrm{i}\sum_{\substack{\gamma,\delta=1\\ \gamma<\delta}}^{n} F_{\gamma\delta} w^\gamma w^\delta\right)\right]$$

$$G_3(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\}) = -\mathrm{i}\sum_{\substack{\gamma,\delta=1\\ \gamma<\delta}}^{n} F_{\gamma\delta}q_{\gamma\delta} + \frac{\mathrm{i}}{2}\sum_{\gamma=1}^{n}E_\gamma\,.$$

Replica symmetry

$$G(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\}) = \alpha G_1(\{q_{\gamma\delta}\}) + G_2(\{F_{\gamma\delta}\},\{E_\gamma\}) + G_3(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\})$$

$$G_1(\{q_{\gamma\delta}\}) = \ln\left[\int_{1-\epsilon}^{\infty}\left(\prod_{\gamma=1}^{n}\frac{\mathrm{d}z_\gamma}{2\pi}\right)\int\left(\prod_{\gamma=1}^{n}\frac{\mathrm{d}\lambda_\gamma \mathrm{d}y_\gamma \mathrm{d}\omega_\gamma}{2\pi}\right)\mathrm{e}^{K(\{\lambda_\gamma\},\{y_\gamma\},\{\omega_\gamma\},\{q_{\gamma\delta}\})}\right]$$

$$K(\{\lambda_\gamma\},\{y_\gamma\},\{\omega_\gamma\},\{q_{\gamma\delta}\}) \equiv \mathrm{i}\sum_{\gamma=1}^{n} y_\gamma\left[z_\gamma - \Phi(\lambda_\gamma)\right] - \mathrm{i}\sum_{\gamma=1}^{n}\left(\frac{\kappa}{\sigma}+\lambda_\gamma\right)\omega_\gamma^\mu - \frac{1}{2\sigma^2}\sum_{\gamma,\delta=1}^{n} q_{\gamma\delta}\omega_\gamma\omega_\delta$$

$$G_2(\{F_{\gamma\delta}\},\{E_\gamma\}) = \ln\left[\int\left(\prod_{\gamma=1}^{n}\mathrm{d}w^\gamma\right)\exp\left(-\frac{\mathrm{i}}{2}\sum_{\gamma=1}^{n}E_\gamma(w^\gamma)^2 + \mathrm{i}\sum_{\substack{\gamma,\delta=1\\ \gamma<\delta}}^{n}F_{\gamma\delta}w^\gamma w^\delta\right)\right]$$

$$G_3(\{q_{\gamma\delta}\},\{F_{\gamma\delta}\},\{E_\gamma\}) = -\mathrm{i}\sum_{\substack{\gamma,\delta=1\\ \gamma<\delta}}^{n}F_{\gamma\delta}q_{\gamma\delta} + \frac{\mathrm{i}}{2}\sum_{\gamma=1}^{n}E_\gamma .$$

Replica symmetry ansatz:

$$q_{\gamma\delta} = q \qquad F_{\gamma\delta} = F \qquad E_\gamma = E,$$

for all $\gamma,\delta = 1,\dots,n$, with $\gamma \neq \delta$

$$G^{\mathrm{RS}}(q,F,E) = \alpha G_1^{\mathrm{RS}}(q) + G_2^{\mathrm{RS}}(F,E) + G_3^{\mathrm{RS}}(q,F,E)$$

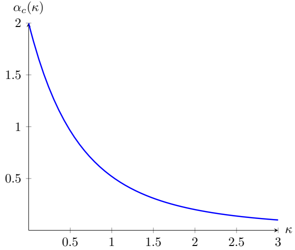
Quantum perceptron: Results

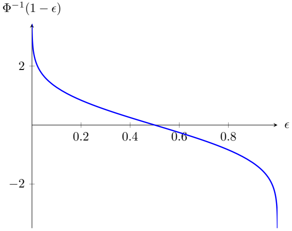
We find the quantum storage capacity:

$$\alpha_c^q(\kappa, \epsilon, \sigma) = \alpha_c(\widetilde{\kappa}), \qquad \alpha_c(\kappa) = \left[\int_{-\kappa}^{\infty} \frac{\mathrm{d}t}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}(t+\kappa)^2\right]^{-1}$$

where the "effective stability" parameter $\widetilde{\kappa}$ is given by:

$$\widetilde{\kappa} = \kappa + \sigma\Phi^{-1}(1-\epsilon), \qquad \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\mathrm{d}t$$





- The performances of the quantum perceptron are always worse in the meaningful regime $0 \leqslant \epsilon \leqslant 1/2$
- In the classical limit $\sigma \to 0$ we retrieve the previous results: $\widetilde{\kappa} \to \kappa$

F. Benatti, G. Gramegna, and S. Mancini, Journal of Physics A: Mathematical and Theoretical **55**, 155301 (2022)

Conclusions and Outlook

- The statistical approach is a powerful tool to compute the storage capacity of both classical and quantum perceptron
- The storage capacity of the quantum perceptron considered is always worse than its classical counterpart

Conclusions and Outlook

- The statistical approach is a powerful tool to compute the storage capacity of both classical and quantum perceptron
- The storage capacity of the quantum perceptron considered is always worse than its classical counterpart

BUT:

- We have not considered other kinds of quantum advantages (e.g: learning speed)
- One might consider other models of quantum perceptron

  A. Gratsea, V. Kasper and M. Lewenstein, Storage properties of a quantum perceptron, https://arxiv.org/abs/2111.08414 (2021)
- When considering multiple-layer neural networks the build-up of quantum coherences might be advantageous
- If one allows for some patterns to be stored "unreliably", the quantum perceptron might still perform better

Conclusions and Outlook

- The statistical approach is a powerful tool to compute the storage capacity of both classical and quantum perceptron
- The storage capacity of the quantum perceptron considered is always worse than its classical counterpart

BUT:

- We have not considered other kinds of quantum advantages (e.g: learning speed)
- One might consider other models of quantum perceptron

  A. Gratsea, V. Kasper and M. Lewenstein, Storage properties of a quantum perceptron, https://arxiv.org/abs/2111.08414 (2021)
- When considering multiple-layer neural networks the build-up of quantum coherences might be advantageous
- If one allows for some patterns to be stored "unreliably", the quantum perceptron might still perform better

Thanks for the attention!