ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Bachelor Project

Neural Network Capacity

Author:
Ridha Chahed

Supervisors:
Dr. Nicolas Macris
Clément Luneau

Lausanne, December 2018

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Table of contents

1 Introduction

Artificial Neural Network provides a practical method for learning certain rules, mappings associating any output to a certain space of inputs. A network is said to implement the rule if for any input the map is correctly achieved.

In the realization of this input-output function, we will interest ourselves to measure the capacity of a neural network, that is, the number of random input-output pairs that we expect to store reliably in a given network.

To that end, we will consider a simple feedforward architecture, the Perceptron. We will present geometrical proof [1] of it's critical capacity and then we will see how E.Gardner [2] used the replica method to calculate the volume of weights capable of storing a certain number of patterns and derive from it a more general result.

2 The Perceptron

2.1 Definition

The perceptron is a supervised learning algorithm that is used for classification of linearly separable patterns. It's formed of a single neuron with adjustable synaptic weights. The single neuron composition limits the pattern classification to only 2 classes.

A perceptron takes a vector of real valued inputs, calculates a linear combination of them and outputs 1 if the result is greater than a threshold and -1 otherwise.

$$o(x_1, ..., x_n) = \begin{cases} 1, & \text{if } w_0 + w_1 * x_1 + ... + w_n * x_n > 0 \\ -1, & \text{otherwise} \end{cases} \tag{1}$$

where each w_i is a weight that determines the contribution of x_i and $-w_0$ is the threshold.

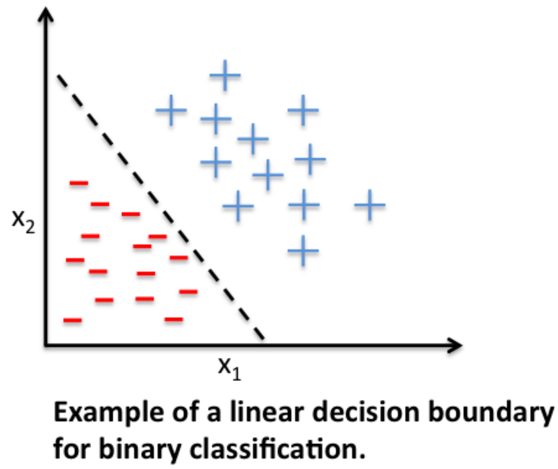

Figure 1:

As we have that

$$o(\overrightarrow{x}) = sign(\overrightarrow{w} \cdot \overrightarrow{x}) \tag{2}$$

$$\overrightarrow{w} \cdot (\overrightarrow{x} * o(\overrightarrow{x})) > 0 \tag{3}$$

we can see the perceptron as representing a hyperplane decision surface (of equation $\overrightarrow{w} \cdot \overrightarrow{x} = 0$) where for each x_i it cuts the space into half.

2.2 Cover's Counting Theorem

Let's consider a perceptron with N continuous valued inputs and ± 1 output unit with a threshold equal to zero. Therefore, the problem of finding the capacity of the network boils down to find how many points can we expect to put randomly in a N-dimensional space and find a hyperplane through the origin that divides the 2 different classes of points.

With continuous inputs the result would be meaningless as every infinitesimal changes would give a new partition. To overcome this problem we will constraint the inputs to a fixed finite set of p input vectors $x^1, ..., x^p$ which gives us 2^p possibilities.

For a given set of p randomly placed points in a N dimensional plane, for how many possible combinations of input-output can we find a hyperplane that divides the 2 classes. Let's call the answer C(p,n) which represents the number of possible linearly separable partition.

For a $p <= N$ it's obviously possible to find hyperplane for any possible combination as we a have a system of p equations with N degrees of freedom.

For a large p it's more difficult and we need to have the points in general position wich implies that every set of N points or fewer must be independent, otherwise they are not linearly separable.

Suppose we start with p points and we add a new point P. We can have 2 possible cases:

(i) If the hyperplane pass through P, there will be 2 new dichotomies as we can move the hyperplane infinitesimally to both side. In this case we constraint the hyperplane to go through a specific point which is equivalent to remove one degree of freedom: C(p,N-1) .

(ii) For another dichotomy only one class will fit.

We therefore get the following recurrence relation:

$$C(p+1, N) = C(p, N) + C(p, N-1) \tag{4}$$

$$\begin{aligned}
C(p+1) &= C(p, N) + c(p, N-1) \\
&= C(p-1, N) + 2 * C(p-1, N-1) + C(p-1, N-2) \\
&= C(p-2, N) + 3 * C(p-2, N) + 3 * C(p-2, N-2) + C(p-1, N-3) \\
&\implies C(p+1, N) = \binom{p}{0} C(1, N) + \binom{p}{1} C(1, N-1) + \ldots + \binom{p}{p} C(1, N-p)
\end{aligned}$$

We know that C(1,k)= 0 for $k < 1$ and C(1,N)=2 as there are only 2 possibilities to classify a single point. We get:

$$C(p, N) = 2 \sum_{i=0}^{N-1} \binom{p-1}{i} \tag{5}$$

If p=2N :

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

By the binomial theorem we have that:

$$\begin{aligned}
(1+1)^{2N-1} &= \sum_{k=0}^{2N-1}\binom{2N-1}{k} \\
&= \sum_{k=0}^{N-1}\binom{2N-1}{k} + \sum_{k=N}^{2N-1}\binom{2N-1}{k} \\
&= \sum_{k=0}^{N-1}\binom{2N-1}{k} + \sum_{k'=0}^{N-1}\binom{2N-1}{2N-1-k'} \\
\binom{n}{k} = \binom{n}{n-k} \implies & \sum_{k=0}^{N-1}\binom{2N-1}{k} + \sum_{k'=0}^{N-1}\binom{2N-1}{k'} \\
&= 2\sum_{k=0}^{N-1}\binom{2N-1}{k}
\end{aligned}$$

$$C(2N,N) = 2^{2N-1} \tag{6}$$

If we allow p and N to approach ∞ while keeping a constant ratio $\alpha = \frac{p}{N}$, as N grows, C(p,N) becomes a step function where the step occurs at $\alpha = 2$ as $\frac{C(2N,N)}{2^N} = 0.5$. Below this value of alpha we can see that in general all dichotomies are possible while above this value the probability goes to zero as shown in figure 2.

By this elegant proof, Cover showed that in a random case the maximum number of patterns that can be learned by a perceptron is 2N. However, this proof doesn't hold for more complicated architecture. We will therefore examine an approach attempted by E. Gardner to calculate the volume of the network weight space of a peceptron capable of storing a certain number of patterns.

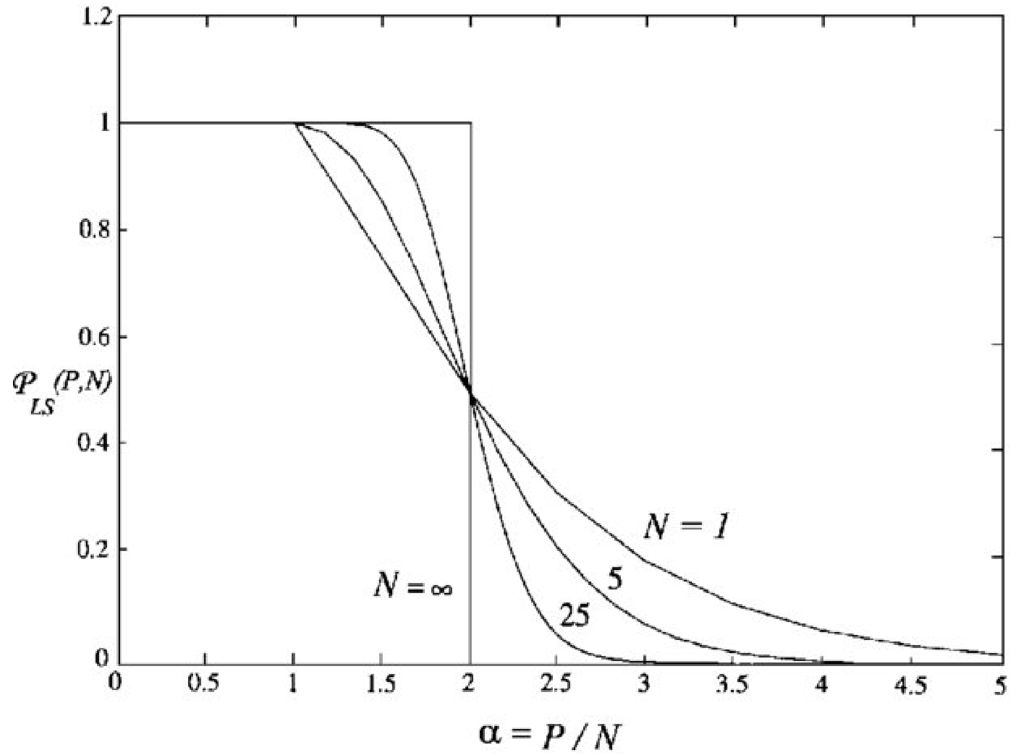

Figure 2: Fraction of linearly separable dichotomies $\frac{C(2N,N)}{2^N}$ vs. P/N

3 Gardner's replica theory

3.1 Replica method to find the critical capacity

Gardner's idea is to consider the fraction of weight space that implements a particular input-output function.

We consider a perceptron with N binary inputs $\epsilon_j = \pm 1$ and M binary threshold units that compute the outputs

$$O_i = sign(\frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_{ij} \epsilon_j) \tag{7}$$

We are given the a set of p random patterns $(\epsilon_j^\mu \rightarrow \zeta_i^\mu) \forall \mu \in [1, p]$ and we want to know the fraction of weight space that satisfies the equations:

$$\zeta_i^\mu = sign(\frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_{ij} \epsilon_j^\mu) \tag{8}$$

wich is equivalent to :

$$\zeta_i^\mu * (\frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_{ij} \epsilon_j^\mu) > 0 \tag{9}$$

and we add a margin size κ to get:

$$\zeta_i^\mu * (\frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_{ij} \epsilon_j^\mu) > \kappa \tag{10}$$

To keep the weights within bounds we add a normalization constraint:

$$\sum_{j=1}^{N} w_{ij}^2 = N \tag{11}$$

We finally get an expression of the fraction of the weight space that implements the function corresponding to the patterns.

$$V_T = \frac{\int dw (\prod_{i\mu} \Theta(\zeta_i^\mu * (\frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_{ij} \epsilon_j^\mu) - \kappa)) \prod_i \delta(\sum_{j=1}^{N} w_{ij}^2 - N)}{\int dw \prod_i \delta(\sum_{j=1}^{N} w_{ij}^2 - N)} \tag{12}$$

Where we select in the numerator with the use of the δ-function and the Heaviside function Θ(x) the part of the weight space that ensures the constraints.

We observe that we have a product on identical term for each i.

$$V_T = \prod_{i=1}^{M} V_i \tag{13}$$

Therefore we can drop the i index, reducing without loss of generality the calculation for a single output unit.

We are interested in the entropy density Z defined as

$$Z = \lim_{N \to \infty} \frac{1}{N} \ln(V) \tag{14}$$

We suppose that this quantity is self averaging and therefore we only need to calculate $\langle \ln(V) \rangle$ the average of $\ln(V)$ over the patterns distribution. This mean is very difficult to calculate directly.

However by the replica trick we have with n close to 0:

$$\ln(V) = \lim_{n \to 0} \frac{V^n - 1}{n} \tag{15}$$

which allows us to compute $\langle \ln(V) \rangle$:

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

$$\langle V \rangle = \frac{\langle \prod_{\alpha=1}^{n} \int dw^{\alpha} (\prod_{\mu=1}^{p} \Theta(\zeta^{\mu} * (\frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_j^{\alpha} \epsilon_j^{\mu}) - \kappa)) \delta(\sum_{j=1}^{N} (w_j^{\alpha})^2 - N) \rangle}{\prod_{\alpha=1}^{n} \int dw^{\alpha} \delta(\sum_{j=1}^{N} (w_j^{\alpha})^2 - N)} \tag{16}$$

We can approximate the step function by:

$$\Theta(z - k) = \int_{k}^{\infty} d\lambda \delta(\lambda - z) \tag{17}$$

Using the delta function integral approximation we get

$$\Theta(z-k) = \int_k^{\infty} d\lambda \int \frac{dx}{2\pi} e^{ix(\lambda - z)} \tag{18}$$

$$\Theta(\zeta^\mu * (\frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_j^\alpha \epsilon_j^\mu) - \kappa) = \int_\kappa^{\infty} \frac{d\lambda_\alpha^\mu}{2\pi} \int dx_\alpha^\mu e^{ix_\alpha^\mu \lambda_\alpha^\mu} e^{-ix_\alpha^\mu z_\alpha^\mu} \tag{19}$$

with $z_\alpha^\mu = \frac{\zeta^\mu}{\sqrt{N}} \sum_{j=1}^{N} w_j^\alpha \epsilon_j^\mu$

$$\begin{aligned}
\langle \prod_{\mu\alpha} e^{-ix_\alpha^\mu z_\alpha^\mu} \rangle &= \prod_{\mu\alpha} \langle e^{-ix_\alpha^\mu \frac{\zeta^\mu}{\sqrt{N}} \sum_{j=1}^{N} w_j^\alpha \epsilon_j^\mu} \rangle \\
&= \prod_{\mu j} \langle e^{-i\zeta^\mu \epsilon_j^\mu \frac{1}{\sqrt{N}} \sum_{\alpha=1}^{n} w_j^\alpha x_j^\mu} \rangle \\
&= \prod_{\mu j} \frac{1}{2} (e^{\frac{-i}{\sqrt{N}} \sum_{\alpha=1}^{n} w_j^\alpha x_j^\mu} + e^{\frac{i}{\sqrt{N}} \sum_{\alpha=1}^{n} w_j^\alpha x_j^\mu}) \\
&= \prod_{\mu j} \cos(\frac{1}{\sqrt{N}} \sum_{\alpha=1}^{n} w_j^\alpha x_j^\mu) \\
&= \exp(\sum_{\mu j} \log(\cos(\frac{1}{\sqrt{N}} \sum_{\alpha=1}^{n} w_j^\alpha x_j^\mu)))
\end{aligned}$$

By Taylor series approximation we get:

$$\cos(x) = 1 - \frac{x^2}{2} + o(x^2) \tag{20}$$

$$\log(1+x) = x + o(x) \tag{21}$$

$$\log(\cos(x)) = \log(1 - \frac{x^2}{2} + o(x^2)) = -\frac{x^2}{2} + o(x^2) \tag{22}$$

$$\exp(\sum_{\mu j}\log(\cos(\frac{1}{\sqrt{N}}\sum_{\alpha=1}^{n}w_j^\alpha x_j^\mu))) = \exp(-\frac{1}{2N}\sum_{\mu j}(\sum_{\alpha}w_j^\alpha x_j^\mu)^2) \tag{23}$$

$$(\sum_{\alpha}x_\alpha^\mu w_j^\alpha)^2 = (\sum_{\alpha}x_\alpha^\mu w_j^\alpha)(\sum_{\beta}x_\beta^\mu w_j^\beta) \tag{24}$$

$$= \sum_{\alpha}\sum_{\beta}x_\alpha^\mu x_\beta^\mu w_j^\alpha w_j^\beta \tag{25}$$

$$\exp(-\frac{1}{2N}\sum_{\mu j}(\sum_{\alpha=1}^{n}w_j^\alpha x_j^\mu)^2) = \exp(-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sum_{\mu=1}^{p}x_\alpha^\mu x_\beta^\mu(\frac{1}{N}\sum_{j=1}^{N}w_j^\alpha w_j^\beta)) \tag{26}$$

$$= \prod_{\mu=1}^{p}\exp(-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}x_\alpha^\mu x_\beta^\mu q_{\alpha\beta}) \tag{27}$$

with

$$q_{\alpha\beta} = \frac{1}{N}\sum_{j=1}^{N}w_j^\alpha w_j^\beta \tag{28}$$

that gives $q_{\alpha\alpha} = \frac{1}{N}\sum_j (w_j^\alpha)^2 = 1$ by (11) and $q_{\alpha\beta} = q_{\beta\alpha}$ so we get

$$\prod_{\mu=1}^{p}\exp(-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}x_\alpha^\mu x_\beta^\mu q_{\alpha\beta}) = \prod_{\mu=1}^{p}\exp(-\frac{1}{2}(\sum_{\alpha=1}^{n}(x_\alpha^\mu)^2 + 2\sum_{\alpha=1}^{n}\sum_{\beta=1}^{\alpha-1}q_{\alpha\beta}x_\alpha^\mu x_\beta^\mu)) \tag{29}$$

$$= \prod_{\mu=1}^{p}\exp(-\frac{1}{2}\sum_{\alpha=1}^{n}(x_\alpha^\mu)^2 - \sum_{\alpha<\beta}q_{\alpha\beta}x_\alpha^\mu x_\beta^\mu) \tag{30}$$

We can finally write:

$$\begin{aligned}\langle\prod_{\mu\alpha}\Theta(\zeta^\mu * (\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^\alpha\epsilon_j^\mu) - \kappa)\rangle = \prod_{\mu=1}^{p}\int_{\kappa}^{\infty}(\prod_{\alpha=1}^{n}\frac{d\lambda_\alpha^\mu}{2\pi})\int(\prod_{\alpha=1}^{n}dx_\alpha^\mu)\\ \exp(i\sum_{\alpha=1}^{n}x_\alpha^\mu\lambda_\alpha^\mu - \frac{1}{2}\sum_{\alpha=1}^{n}(x_\alpha^\mu)^2 - \sum_{\alpha<\beta}q_{\alpha\beta}x_\alpha^\mu x_\beta^\mu)\end{aligned} \tag{31}$$

We can drop the $\mu's$ and write:

$$\langle\prod_{\mu\alpha}\Theta(\zeta^{\mu}*(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^{\alpha}\epsilon_j^{\mu})-\kappa)\rangle=[\int_{\kappa}^{\infty}(\prod_{\alpha=1}^{n}\frac{d\lambda_{\alpha}}{2\pi})\int(\prod_{\alpha=1}^{n}dx_{\alpha})e^{K(\lambda,x,q)}]^{p} \quad (32)$$

with $K(\lambda,x,q)=i\sum_{\alpha=1}^{n}x_{\alpha}\lambda_{\alpha}-\frac{1}{2}\sum_{\alpha=1}^{n}(x_{\alpha})^2-\sum_{\alpha<\beta}q_{\alpha\beta}x_{\alpha}x_{\beta}$

We now turn our attention to the constraints represented by a δ function. First of all, we know that a δ function can be represented as an integral as follows:

$$\delta(z)=\int\frac{dr}{2\pi i}e^{-rz} \quad (33)$$

Which allows us to get:

$$\delta(\sum_{j}(w_j)^2-N)=\int\frac{dr}{2\pi i}e^{-r(\sum_j(w_j^{\mu})^2-N)} \quad (34)$$

we set $r=\frac{E_{\alpha}}{2}$

$$\delta(\sum_{j}(w_j)^2-N)=\int\frac{dE_{\alpha}}{4\pi i}e^{-\frac{E_{\alpha}}{2}(\sum_j(w_j^{\alpha})^2-N)} \quad (35)$$

The definition of $q_{\alpha\beta}$ adds an additional constraint for each pair (α,β) with $\alpha<\beta$ we have

$$\delta(q_{\alpha\beta}-\frac{1}{N}\sum_{j}w_j^{\alpha}w_j^{\beta})=\int\frac{dr}{2\pi i}e^{-r(q_{\alpha\beta}-\frac{1}{N}\sum_j w_j^{\alpha}w_j^{\beta})} \quad (36)$$

we set $r=NF_{\alpha\beta}$

$$\delta(q_{\alpha\beta}-\frac{1}{N}\sum_{j}w_j^{\alpha}w_j^{\beta})=N\int\frac{dF_{\alpha\beta}}{2\pi i}e^{-NF_{\alpha\beta}(q_{\alpha\beta}-\frac{1}{N}\sum_j w_j^{\alpha}w_j^{\beta})} \quad (37)$$

The volume of the space weight that enforce these constraints is

$$\prod_{\alpha}\prod_{\alpha<\beta}\int(\prod_{j=1}^{N}dw_j^{\alpha})dq_{\alpha\beta}\delta(\sum_{j=1}^{N}(w_j^{\alpha})^2-N)\delta(q_{\alpha\beta}-\frac{1}{N}\sum_{j=1}^{N}w_j^{\alpha}w_j^{\beta})$$

$$=\prod_{\alpha}\prod_{\alpha<\beta}\int(\prod_{j=1}^{N}dw_j^{\alpha})dq_{\alpha\beta}\frac{dE_{\alpha}}{4\pi i}\frac{dF_{\alpha\beta}}{2\pi i}\exp{(-\frac{E_{\alpha}}{2}(\sum_{j}(w_j^{\alpha})^2-N)-NF_{\alpha\beta}(q_{\alpha\beta}-\frac{1}{N}\sum_{j}w_j^{\alpha}w_j^{\beta}))}$$

$$=\prod_{\alpha}\prod_{\alpha<\beta}\int\exp{(\frac{NE_{\alpha}}{2}-NF_{\alpha\beta}q_{\alpha\beta})}\frac{dE_{\alpha}}{4\pi i}\frac{dF_{\alpha\beta}}{2\pi i}dq_{\alpha\beta}\int(\prod_{j=1}^{N}dw_j^{\alpha})\prod_{j=1}^{N}\exp{(-\frac{E_{\alpha}}{2}(w_j^{\alpha})^2+F_{\alpha\beta}w_j^{\alpha}w_j^{\beta})}$$

$$=\prod_{\alpha}\prod_{\alpha<\beta}\int\exp{(\frac{NE_{\alpha}}{2}-NF_{\alpha\beta}q_{\alpha\beta})}\frac{dE_{\alpha}}{4\pi i}\frac{dF_{\alpha\beta}}{2\pi i}dq_{\alpha\beta}\prod_{j=1}^{N}\int(dw_j^{\alpha})\exp{(-\frac{E_{\alpha}}{2}(w_j^{\alpha})^2+F_{\alpha\beta}w_j^{\alpha}w_j^{\beta})}$$

$$=\int\exp{(\frac{NE_{\alpha}}{2}-NF_{\alpha\beta}q_{\alpha\beta})}(\prod_{\alpha}\frac{dE_{\alpha}}{4\pi i})(\prod_{\alpha<\beta}\frac{dF_{\alpha\beta}}{2\pi i}dq_{\alpha\beta})$$
$$\prod_{j=1}^{N}\int(\prod_{\alpha}dw_j^{\alpha})\exp{(-\sum_{\alpha}\frac{E_{\alpha}}{2}(w_j^{\alpha})^2+\sum_{\alpha<\beta}F_{\alpha\beta}w_j^{\alpha}w_j^{\beta})}$$

$$\boxed{\begin{aligned}&=\int\exp{(\frac{NE_{\alpha}}{2}-NF_{\alpha\beta}q_{\alpha\beta})}(\prod_{\alpha}\frac{dE_{\alpha}}{4\pi i})(\prod_{\alpha<\beta}\frac{dF_{\alpha\beta}}{2\pi i}dq_{\alpha\beta})\\&[\int(\prod_{\alpha}dw_{\alpha})\exp{(-\sum_{\alpha}\frac{E_{\alpha}}{2}(w_{\alpha})^2+\sum_{\alpha<\beta}F_{\alpha\beta}w_{\alpha}w_{\beta})}]^N\end{aligned}} \tag{38}$$

We finally get that:

$$\langle V^n\rangle=\frac{\int(\prod_{\alpha}dE_{\alpha})(\prod_{\alpha<\beta}dF_{\alpha\beta}dq_{\alpha\beta})e^{NG(q,F,E)}}{\int(\prod_{\alpha}dE_{\alpha})e^{NH(E)}} \tag{39}$$

with

$$G(q,F,E) = \frac{p}{N}\log[\int_{\kappa}^{\infty}(\prod_{\alpha}\frac{d\lambda_\alpha}{2\pi})\int(\prod_{\alpha}dx_\alpha)e^{K(\lambda,x,q)}] \tag{40}$$

$$+\log[\int(\prod_{\alpha}dw^{\alpha})\exp(-\sum_{\alpha}\frac{E_\alpha}{2}(w_\alpha)^2+\sum_{\alpha<\beta}F_{\alpha\beta}w_\alpha w_\beta)] \tag{41}$$

$$+\sum_{\alpha}\frac{E_\alpha}{2}-\sum_{\alpha<\beta}F_{\alpha\beta}q_{\alpha\beta} \tag{42}$$

$$H(E) = \log[\int(\prod_{\alpha}dw_\alpha)e^{-\sum_\alpha E_\alpha\frac{w_\alpha^2}{2}}]+\sum_{\alpha}\frac{E_\alpha}{2} \tag{43}$$

To calculate these integrals we will use Laplace's method that gives us that

$$I(\lambda) = \int_a^b e^{-\lambda p(t)}q(t)dt = e^{\lambda p(t_0)}q(t_0)\sqrt{\frac{2\pi}{\lambda p''(t_0)}} \tag{44}$$

with the point $t = t_0$ a minimum of p(t).

Therefore, to calculate the integral we need to find the saddle point of G over q, F and E. To that extent we will assume the replica symmetric ansatz and for all α and β with $\alpha < \beta$ we have:

$$q^{\alpha\beta} = q$$
$$F^{\alpha\beta} = F$$
$$E^{\alpha} = E$$

To compute K(λ,x,q) we first acknowledge:

$$(\sum_{\alpha}x_\alpha)^2 = \sum_{\alpha}x_\alpha^2 + 2\sum_{\alpha<\beta}x_\alpha x_\beta$$
$$\sum_{\alpha<\beta}x_\alpha x_\beta = \frac{1}{2}[(\sum_{\alpha}x_\alpha)^2 - \sum_{\alpha}x_\alpha^2]$$

which gives us

$$
\begin{aligned}
K(\lambda, x, q) &= i\sum_\alpha x_\alpha\lambda_\alpha - \frac{1}{2}\sum_\alpha x_\alpha^2 - \sum_{\alpha<\beta} x_\alpha x_\beta q_{\alpha\beta} \\
&= i\sum_\alpha x_\alpha\lambda_\alpha - \frac{1}{2}\sum_\alpha x_\alpha^2 - \frac{q}{2}[(\sum_\alpha x_\alpha)^2 - \sum_\alpha x_\alpha^2] \\
&= i\sum_\alpha x_\alpha\lambda_\alpha - \frac{1-q}{2}\sum_\alpha x_\alpha^2 - \frac{q}{2}(\sum_\alpha x_\alpha)^2
\end{aligned}
$$

Now we want to linearize the $e^{-\frac{1}{2}q(\sum_\alpha x_\alpha)^2}$ term. We use the Gaussian integral trick:

$$
\boxed{\int \exp(-ay^2 + xy)dy = \sqrt{\frac{\pi}{a}}\exp(\frac{1}{4a}x^2)} \tag{45}
$$

$$
\begin{aligned}
\exp(-\frac{1}{2}q(\sum_\alpha x_\alpha)^2) &= \exp(\frac{1}{4(-\frac{1}{2q})}(\sum_\alpha x_\alpha)^2) \\
&= \frac{1}{i\sqrt{2\pi q}}\int \exp(\frac{1}{2q}y^2 + y\sum_\alpha x_\alpha)dy
\end{aligned}
$$

We perform a change of variable:

$$
t = \frac{y}{i\sqrt{q}} \implies dy = i\sqrt{q}dt
$$

$$
\int \frac{1}{\sqrt{2\pi}}\exp(-\frac{t^2}{2} + i\sqrt{q}t\sum_\alpha x_\alpha)dt \tag{46}
$$

The first term of G(q,F,E) becomes

$$\frac{p}{N}\log[\int_{\kappa}^{\infty}(\prod_{\alpha}\frac{d\lambda_{\alpha}}{2\pi})\int(\prod_{\alpha}dx_{\alpha})e^{K(\lambda,x,q)}]$$
$$=\frac{p}{N}\log[\int_{\kappa}^{\infty}(\prod_{\alpha}\frac{d\lambda_{\alpha}}{2\pi})\int(\prod_{\alpha}dx_{\alpha})\frac{1}{\sqrt{2\pi}}\exp{(i\sum_{\alpha}x_{\alpha}\lambda_{\alpha}-\frac{1-q}{2}\sum_{\alpha}x_{\alpha}^{2}-\frac{t^{2}}{2}+i\sqrt{q}t\sum_{\alpha}x_{\alpha})}dt]$$
$$=\frac{p}{N}\log[\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^{2}}{2}}[\int_{\kappa}^{\infty}\frac{d\lambda}{2\pi}\int dx\exp{(ix\lambda-\frac{1-q}{2}x^{2}+i\sqrt{q}tx)}]^{n}]$$
$$=\frac{p}{N}\log[\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^{2}}{2}}[\int_{\kappa}^{\infty}\frac{d\lambda}{2\pi}\int dx\exp{(-\frac{1-q}{2}x^{2}+xi(\lambda+\sqrt{q}t))}]^{n}]$$

To calculate the integral over x we use the Gaussian integral trick (45) with $a=\frac{1-q}{2}$ and we fix the ratio $\alpha\equiv\frac{p}{N}$

$$\frac{p}{N}\log[\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^{2}}{2}}[\int_{\kappa}^{\infty}\frac{d\lambda}{2\pi}\int dx\exp{(-\frac{1-q}{2}x^{2}+xi(\lambda+\sqrt{q}t))}]^{n}]$$
$$=\alpha\log[\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^{2}}{2}}[\int_{\kappa}^{\infty}\frac{d\lambda}{\sqrt{2\pi}}\frac{1}{\sqrt{1-q}}\exp{(-\frac{(\lambda+\sqrt{q}t)^{2}}{2(1-q)})}]^{n}]$$

By Taylor series approximation we know that

$$e^{x}=1+x+o(x)$$
$$\log(1+x)=x+o(x)$$

as $n\to 0$ we get:

$$\begin{aligned}
&\alpha \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} [\int_{\kappa}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} \frac{1}{\sqrt{1-q}} \exp{(-\frac{(\lambda+\sqrt{q}t)^2}{2(1-q)})}]^n] \\
&= \alpha \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \exp{(n \log[\int_{\kappa}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} \frac{1}{\sqrt{1-q}} \exp{(-\frac{(\lambda+\sqrt{q}t)^2}{2(1-q)})}])}] \\
&= \alpha \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} (1 + n \log[\int_{\kappa}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} \frac{1}{\sqrt{1-q}} \exp{(-\frac{(\lambda+\sqrt{q}t)^2}{2(1-q)})}])] \\
&= \alpha \log[1 + \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} (n \log[\int_{\kappa}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} \frac{1}{\sqrt{1-q}} \exp{(-\frac{(\lambda+\sqrt{q}t)^2}{2(1-q)})}])] \\
&= n\alpha \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \log[\int_{\kappa}^{\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} \exp{(-\frac{(\lambda+\sqrt{q}t)^2}{2(1-q)})}]
\end{aligned}$$

We do the same for the second term of G(q,F,E)

$$\begin{aligned}
&\log[\int (\prod_{\alpha} dw_{\alpha}) \exp{(-\sum_{\alpha} \frac{E_{\alpha}}{2} (w_{\alpha})^2 + \sum_{\alpha<\beta} F_{\alpha\beta} w_{\alpha} w_{\beta})}] \\
&= \log[\int (\prod_{\alpha} dw_{\alpha}) \exp{(-\frac{E}{2} \sum_{\alpha} (w_{\alpha})^2 + F \sum_{\alpha<\beta} w_{\alpha} w_{\beta})}]
\end{aligned}$$

We use that:

$$\begin{aligned}
(\sum_{\alpha} w_{\alpha})^2 &= \sum_{\alpha} w_{\alpha}^2 + 2 \sum_{\alpha<\beta} w_{\alpha} w_{\beta} \\
\sum_{\alpha<\beta} w_{\alpha} w_{\beta} &= \frac{1}{2} [(\sum_{\alpha} w_{\alpha})^2 - \sum_{\alpha} w_{\alpha}^2]
\end{aligned}$$

$$\begin{aligned}
&\log[\int (\prod_{\alpha} dw_{\alpha}) \exp{(-\frac{E}{2} \sum_{\alpha} w_{\alpha}^2 + \frac{F}{2} [(\sum_{\alpha} w_{\alpha})^2 - \sum_{\alpha} w_{\alpha}^2]]} \\
&\log[\int (\prod_{\alpha} dw_{\alpha}) \exp{(-\frac{E+F}{2} \sum_{\alpha} w_{\alpha}^2 + \frac{F}{2} (\sum_{\alpha} w_{\alpha})^2]}
\end{aligned}$$

As previously we linearize the $e^{\frac{F}{2}(\sum_\alpha w_\alpha)^2}$ term with the Gaussian integral trick (45) with $a = \frac{1}{2F}$.

$$\begin{aligned}
\exp\left(\frac{F}{2}(\sum_\alpha w_\alpha)^2\right) &= \exp\left(\frac{1}{4(\frac{1}{2F})}(\sum_\alpha w_\alpha)^2\right) \\
&= \frac{1}{\sqrt{2\pi F}}\int \exp\left(-\frac{1}{2F}y^2 + y\sum_\alpha w_\alpha\right)dy
\end{aligned}$$

We perform a change of variable:

$$y^2 = t^2 F \implies y = t\sqrt{F} \implies dy = dt\sqrt{F}$$

$$\exp\left(\frac{F}{2}(\sum_\alpha w_\alpha)^2\right) = \frac{1}{\sqrt{2\pi}}\int \exp\left(-\frac{t^2}{2} + t\sqrt{F}\sum_\alpha w_\alpha\right)dt \tag{47}$$

The second term of G(q,F,E) becomes

$$\begin{aligned}
&log[\int (\prod_\alpha dw_\alpha)\frac{dt}{\sqrt{2\pi}} \exp\left(-\frac{E+F}{2}\sum_\alpha w_\alpha^2 + -\frac{t^2}{2} + t\sqrt{F}\sum_\alpha w_\alpha\right)] \\
&= \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int (\prod_\alpha dw_\alpha) \prod_\alpha \exp\left(-\frac{E+F}{2}w_\alpha^2 + t\sqrt{F}w_\alpha\right)] \\
&= \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} [\int dw \exp\left(-\frac{E+F}{2}w^2 + t\sqrt{F}w\right)]^n]
\end{aligned}$$

To calculate the integral over w we use the Gaussian integral trick (45) with $a = \frac{E+F}{2}$

$$\begin{aligned}
&\log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} [\int dw \exp\left(-\frac{E+F}{2}w^2 + t\sqrt{F}w\right)]^n] \\
&= \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} [\sqrt{\frac{2\pi}{E+F}} \exp\left(\frac{t^2 F}{2(E+F)}\right)]^n]
\end{aligned}$$

By Taylor series approximation as $n \to 0$ we get:

$$\begin{aligned}
&\log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \exp\left(n \log\left(\sqrt{\frac{2\pi}{E+F}} \exp\left(\frac{t^2 F}{2(E+F)}\right)\right)\right)] \\
&= \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \left(1 + n \log\left(\sqrt{\frac{2\pi}{E+F}} \exp\left(\frac{t^2 F}{2(E+F)}\right)\right)\right)] \\
&= \log[1 + \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \left(n \log\left(\sqrt{\frac{2\pi}{E+F}} \exp\left(\frac{t^2 F}{2(E+F)}\right)\right)\right)] \\
&= n \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \left(\log\left(\sqrt{\frac{2\pi}{E+F}} \exp\left(\frac{t^2 F}{2(E+F)}\right)\right)\right) \\
&= \frac{n}{2}\left(\log(2\pi) - \log(E+F) + \frac{F}{E+F} \int \frac{dt}{\sqrt{2\pi}} t^2 e^{-\frac{t^2}{2}}\right) \\
&= \frac{n}{2}\left(\log(2\pi) - \log(E+F) + \frac{F}{E+F}\right)
\end{aligned}$$

The third term gives us

$$\sum_{\alpha} \frac{E_\alpha}{2} - \sum_{\alpha<\beta} F_{\alpha\beta} q_{\alpha\beta} = \frac{n}{2}(E + qF) \tag{48}$$

We finally get

$$\boxed{\begin{aligned} G(q,F,E) = n\alpha \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \log[\int_{\kappa}^{\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} \exp\left(-\frac{(\lambda + \sqrt{q}t)^2}{2(1-q)}\right)] \\ + \frac{n}{2}\left(\log(2\pi) - \log(E+F) + \frac{F}{E+F} + (E+qF)\right) \end{aligned}} \tag{49}$$

It's now possible to find the saddle point of G with respect to q,F and E:

$$\begin{aligned}
\frac{\delta G}{\delta E} = 0 &\Leftrightarrow -\frac{1}{E+F} - \frac{F}{(E+F)^2} + 1 = 0 \\
&\Leftrightarrow -E - F - F + (E+F)^2 = 0 \\
&\Leftrightarrow (E+F)^2 = E + 2F
\end{aligned}$$

$$
\begin{aligned}
\frac{\delta G}{\delta F} = 0 &\Leftrightarrow -\frac{1}{E+F} + \frac{E}{(E+F)^2} + q \\
&\Leftrightarrow -E - F + E + q(E+F)^2 = 0 \\
&\Leftrightarrow -F + q(E+F)^2 = 0 \\
&\Leftrightarrow -F + q(E+2F) = 0 \\
&\Leftrightarrow F(2q-1) + qE = 0 \\
&\Leftrightarrow F = \frac{qE}{1-2q}
\end{aligned}
$$

$$
\begin{aligned}
E + 2F - (E+F)^2 = 0 &\Leftrightarrow E + \frac{2qE}{1-2q} - (E + \frac{qE}{1-2q})^2 = 0 \\
&\Leftrightarrow E(1 + \frac{2q}{1-2q} - E - \frac{2qE}{1-2q} - \frac{q^2E}{(1-2q)^2}) = 0 \\
&\Leftrightarrow (1 + \frac{2q}{1-2q} - E - \frac{2qE}{1-2q} - \frac{q^2E}{(1-2q)^2}) = 0 \\
&\Leftrightarrow E[-1 - \frac{2q}{1-2q} - \frac{q^2}{1-2q}] = -1 - \frac{2q}{1-2q} \\
&\Leftrightarrow E(1 + \frac{q}{1-2q})^2 = 1 + \frac{2q}{1-2q} \\
&\Leftrightarrow E(\frac{1-q}{1-2q})^2 = \frac{1}{1-2q} \\
&\Leftrightarrow \boxed{E = \frac{1-2q}{(1-q)^2}} \implies F = \frac{qE}{1-2q} = \boxed{\frac{q}{(1-q)^2}}
\end{aligned}
$$

We introduce these values in the expression of G in (49).

$$
\begin{aligned}
\frac{1}{n}G(q) = \alpha \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \log[\int_{\kappa}^{\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} \exp{(-\frac{(\lambda + \sqrt{q}t)^2}{2(1-q)})}] \\
+ \frac{1}{2}(\log(2\pi) + \log(1-q) + \frac{q}{(1-q)} + 1)
\end{aligned}
$$

We perform a change of variable:

$$
z = \frac{\lambda + \sqrt{q}t}{\sqrt{(1-q)}} \implies dz = \frac{d\lambda}{\sqrt{1-q}}
$$

$$\frac{1}{n}G(q) = \alpha\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\log[\int_{\frac{\kappa+\sqrt{q}t}{\sqrt{(1-q)}}}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}]+\frac{1}{2}(\log(2\pi)+\log(1-q)+\frac{q}{(1-q)}+1) \tag{50}$$

We now look for $\frac{\delta G}{\delta q} = 0$:

$$\frac{\delta G}{\delta q} = 0 \Leftrightarrow -n\alpha\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}[\int_{\frac{\kappa+\sqrt{q}t}{\sqrt{(1-q)}}}^{\infty}e^{-\frac{z^2}{2}}]^{-1}e^{-\frac{\frac{\kappa+\sqrt{q}t}{\sqrt{(1-q)}}^2}{2}}\frac{t+\kappa\sqrt{q}}{2\sqrt{q}(1-q)^{\frac{3}{2}}}+\frac{nq}{2(1-q)^2} = 0$$

$$\Leftrightarrow \alpha\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}[\int_{\frac{\kappa+\sqrt{q}t}{\sqrt{(1-q)}}}^{\infty}e^{-\frac{z^2}{2}}]^{-1}e^{-\frac{\frac{\kappa+\sqrt{q}t}{\sqrt{(1-q)}}^2}{2}}\frac{t+\kappa\sqrt{q}}{2\sqrt{q}(1-q)^{\frac{3}{2}}} = \frac{q}{2(1-q)^2}$$

The parameter q characterizes the overlap between pairs of solutions.As α increases different solutions become more correlated and q increases. We can therefore determine α_c the maximum storage capacity of the network as q tends to its maximum value which is 1.

L'Hospital's rule gives as $q \rightarrow 1$:

$$\boxed{\alpha_c(\kappa) = [\int\frac{dt}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}(t+\kappa)]^{-1}} \tag{51}$$

This result satisfies Cover's result as $\alpha_c(0) = 2$.

3.2 Derivation of the replica method with a general activation function.

We will now apply Gardner's method with the same architecture but with a general bijective odd activation function $\varphi(x)$.

As above, given the a set of p random patterns $(x_j^\mu \rightarrow y^\mu)\forall\mu \in [1,p]$ we want to know the fraction of weight space that satisfies the equations:

$$y^\mu = \varphi(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j x_j^\mu) \tag{52}$$

We introduce n replica:

$$\langle V\rangle = \frac{\langle\prod_{\alpha=1}^{n}\int dw^{\alpha}(\prod_{\mu=1}^{p}\delta(y^{\mu}-\varphi(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^{\alpha}x_j^{\mu})))\delta(\sum_{j=1}^{N}(w_j^{\alpha})^2-N)\rangle}{\prod_{\alpha=1}^{n}\int dw^{\alpha}\delta(\sum_{j=1}^{N}(w_j^{\alpha})^2-N)} \tag{53}$$

Let's first focus on $\langle\prod_{\alpha,\mu}\delta(y^{\mu}-\varphi(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^{\alpha}x_j^{\mu}))\rangle$ which is difficult to calculate in that form.

We know that

$$\delta(f(x)) = \sum_i \frac{1}{|f'(x_i)|}\delta(x-x_i) \tag{54}$$

with $f(x_i)=0$

We set $z=\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^{\alpha}x_j^{\mu}$ and $f(z)=y^{\mu}-\delta(z)$

$$\prod_{\alpha,\mu}\delta(y^{\mu}-\varphi(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^{\alpha}x_j^{\mu})) = \prod_{\alpha,\mu}\frac{1}{|\varphi'(\varphi^{-1}(y^{\mu}))|}\delta(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^{\alpha}x_j^{\mu}-\varphi^{-1}(y^{\mu})) \tag{55}$$

We now use (33) to put the delta function in its integral form and we get:

$$\prod_{\alpha,\mu}\frac{1}{2\pi|\varphi'(\varphi^{-1}(y^{\mu}))|}\int d\lambda_{\alpha}^{\mu}\exp(i\lambda_{\alpha}^{\mu}(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^{\alpha}x_j^{\mu}-\varphi^{-1}(y^{\mu})))$$
$$=\prod_{\mu=1}^{p}\int(\prod_{\alpha=1}^{n}\frac{d\lambda_{\alpha}^{\mu}}{2\pi})\frac{1}{|\varphi'(\varphi^{-1}(y^{\mu}))|^{n}}\exp(-i\sum_{\alpha=1}^{n}\lambda_{\alpha}^{\mu}\varphi^{-1}(y^{\mu}))\prod_{j=1}^{N}\exp(i\sum_{\alpha=1}^{n}\frac{\lambda_{\alpha}^{\mu}}{\sqrt{N}}w_j^{\alpha}x_j^{\mu})$$

In this form we can now without too much difficulties compute the mean over the patterns distribution. The patterns are chosen randomly, it implies that we have a Bernoulli distribution with $p=\frac{1}{2}$.

First of all, for the y's we get:

$$\frac{1}{2}[\frac{1}{|\varphi'(\varphi^{-1}(1))|^n}\exp\left(-i\varphi^{-1}(1)\sum_{\alpha=1}^{n}\lambda_\alpha^\mu\right)+\frac{1}{|\varphi'(\varphi^{-1}(-1))|^n}\exp\left(-i\varphi^{-1}(-1)\sum_{\alpha=1}^{n}\lambda_\alpha^\mu\right)]$$

We know that φ is a bijective odd function: $\varphi(-x)=-\varphi(x)$

$$\implies \varphi^{-1}(-y)=\varphi^{-1}(-\varphi(x))=\varphi^{-1}(\varphi(-x))=-x=-\varphi^{-1}(y)$$
$$\implies \varphi^{-1}(x) \text{ is odd}$$

$$\varphi(-x)=-\varphi(x) \implies -\varphi'(-x)=-\varphi'(x) \implies \varphi'(-x)=\varphi'(x)$$
$$\implies \varphi'(x) \text{ is even}$$

Therefore the mean over the y's is:

$$\frac{1}{|\varphi'(\varphi^{-1}(1))|^n}\cos(\varphi^{-1}(1)\sum_{\alpha=1}^{n}\lambda_\alpha^\mu) \tag{56}$$

By Taylor series approximation (22) we get:

$$\prod_{\mu=1}^{p}\frac{1}{|\varphi'(\varphi^{-1}(1))|^n}\exp(\log(\cos(\varphi^{-1}(1)\sum_{\alpha=1}^{n}\lambda_\alpha^\mu))) \tag{57}$$
$$=\boxed{\prod_{\mu=1}^{p}\frac{1}{|\varphi'(\varphi^{-1}(1))|^n}\exp(-\frac{1}{2}(\varphi^{-1}(1))^2(\sum_{\alpha=1}^{n}\lambda_\alpha^\mu)^2)} \tag{58}$$

We now turn our attention to the mean over the x's :

$$\frac{1}{2}[\exp\left(\frac{i}{\sqrt{N}}\sum_{\alpha=1}^{n}\lambda_\alpha^\mu w_j^\alpha\right)+\exp\left(\frac{-i}{\sqrt{N}}\sum_{\alpha=1}^{n}\lambda_\alpha^\mu w_j^\alpha\right)]$$
$$=\cos(\frac{1}{\sqrt{N}}\sum_{\alpha=1}^{n}\lambda_\alpha^\mu w_j^\alpha)$$

By Taylor approximation (22) we get:

$$\begin{aligned}\prod_{\mu=1}^{p}\prod_{j=1}^{N}\cos(\frac{1}{\sqrt{N}}\sum_{\alpha=1}^{n}\lambda_\alpha^\mu w_j^\alpha) &= \prod_{\mu=1}^{p}\exp\,(\sum_{j=1}^{N}\log(\cos(\frac{1}{\sqrt{N}}\sum_{\alpha=1}^{n}\lambda_\alpha^\mu w_j^\alpha)))\\ &= \prod_{\mu=1}^{p}\exp\,(\sum_{j=1}^{N}\log(\cos(\frac{1}{\sqrt{N}}\sum_{\alpha=1}^{n}\lambda_\alpha^\mu w_j^\alpha)))\\ &= \prod_{\mu=1}^{p}\exp\,(-\frac{1}{2N}\sum_{j=1}^{N}(\sum_{\alpha=1}^{n}\lambda_\alpha^\mu w_j^\alpha)^2)\\ &= \prod_{\mu=1}^{p}\exp\,(-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha^\mu\lambda_\beta^\mu q_{\alpha\beta})\end{aligned}$$

We do the same steps as in Gardner's calculus (from 23 to 30) to get:

$$\prod_{\mu=1}^{p}\exp\,(-\frac{1}{2}\sum_{\alpha=1}^{n}(\lambda_\alpha^\mu)^2-\sum_{\alpha<\beta}\lambda_\alpha^\mu\lambda_\beta^\mu q_{\alpha\beta}) \tag{59}$$

We finally get :

$$\langle\delta(y^\mu-\varphi(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_j^\alpha x_j^\mu))\rangle = [\int(\prod_{\alpha=1}^{n}\frac{d\lambda_\alpha}{2\pi})\frac{1}{|\varphi'(\varphi^{-1}(y^\mu))|^n}e^{K(\lambda,q,\varphi)}]^p \tag{60}$$

with

$$K(\lambda,q,\varphi) = -\frac{1}{2}\sum_{\alpha=1}^{n}(\lambda_\alpha)^2-\sum_{\alpha<\beta}q_{\alpha\beta}\lambda_\alpha\lambda_\beta-\frac{1}{2}\varphi^{-1}(1)^2(\sum_{\alpha=1}^{n}\lambda_\alpha)^2 \tag{61}$$

For the mean over the $\delta-$functions representing the constraints of the weights normalization (35) and $q_{\alpha\beta}$'s definition (37) we have exactly the same thing as Gardner and we obtain (38).

We finally get that:

$$\langle V^n\rangle = \frac{\int(\prod_\alpha dE_\alpha)(\prod_{\alpha<\beta}dF_{\alpha\beta}dq_{\alpha\beta})e^{NG(q,F,E)}}{\int(\prod_\alpha dE_\alpha)e^{NH(E)}} \tag{62}$$

with

$$G(q,F,E) = \frac{p}{N}\log[\int(\prod_\alpha \frac{d\lambda_\alpha}{2\pi})\frac{e^{K(\lambda,q,\varphi)}}{|\varphi'(\varphi^{-1}(y^\mu))|^n}]$$
$$+\log[\int(\prod_\alpha dw^\alpha)\exp\left(-\sum_\alpha \frac{E_\alpha}{2}(w_\alpha)^2+\sum_{\alpha<\beta}F_{\alpha\beta}w_\alpha w_\beta\right)]$$
$$+\sum_\alpha \frac{E_\alpha}{2}-\sum_{\alpha<\beta}F_{\alpha\beta}q_{\alpha\beta}$$

$$H(E) = \log[\int(\prod_\alpha dw_\alpha)e^{-\sum_\alpha E_\alpha \frac{w_\alpha^2}{2}}]+\sum_\alpha \frac{E_\alpha}{2}$$

We apply the replica symmetric ansatz and for all α and β with $\alpha < \beta$ we have:

$$q^{\alpha\beta} = q$$
$$F^{\alpha\beta} = F$$
$$E^\alpha = E$$

The only modification in these expressions that we didn't already computed is the first term of G(q,F,E):

$$\frac{p}{N}\log[\int(\prod_\alpha \frac{d\lambda_\alpha}{2\pi})\frac{e^{K(\lambda,q,\varphi)}}{|\varphi'(\varphi^{-1}(y^\mu))|^n}]$$

with

$$K(\lambda,q,\varphi) = -\frac{1}{2}\sum_{\alpha=1}^n(\lambda_\alpha)^2 - q\sum_{\alpha<\beta}\lambda_\alpha\lambda_\beta - \frac{1}{2}\varphi^{-1}(1)^2(\sum_{\alpha=1}^n\lambda_\alpha)^2 \qquad (63)$$

knowing that

$$(\sum_{\alpha=1}^n\lambda_\alpha)^2 = \sum_{\alpha=1}^n\lambda_\alpha^2 + 2\sum_{\alpha<\beta}\lambda_\alpha\lambda_\beta$$

$$\Leftrightarrow \sum_{\alpha<\beta} \lambda_\alpha \lambda_\beta = \frac{1}{2}[(\sum_{\alpha=1}^{n} \lambda_\alpha)^2 - \sum_{\alpha=1}^{n} \lambda_\alpha^2]$$

it becomes

$$K(\lambda, q, \varphi) = -\frac{1-q}{2} \sum_{\alpha=1}^{n} \lambda_\alpha^2 - \frac{q + \varphi^{-1}(1)^2}{2} (\sum_{\alpha=1}^{n} \lambda_\alpha)^2 \tag{64}$$

To linearize it we use the Gaussian trick (45) with $a = -\frac{1}{2(q+\varphi^{-1}(1)^2)}$

$$\begin{aligned}
\exp(-\frac{q+\varphi^{-1}(1)^2}{2}(\sum_\alpha \lambda_\alpha)^2) &= \exp(\frac{1}{4(-\frac{1}{2(q+\varphi^{-1}(1)^2)})}(\sum_\alpha \lambda_\alpha)^2) \\
&= \frac{1}{i\sqrt{2\pi(q+\varphi^{-1}(1)^2)}} \int \exp(\frac{1}{2(q+\varphi^{-1}(1)^2)} y^2 + y \sum_\alpha \lambda_\alpha) dy
\end{aligned}$$

We perform a change of variable:

$$t = \frac{y}{i\sqrt{q+\varphi^{-1}(1)^2}} \implies dy = i\sqrt{q+\varphi^{-1}(1)^2} dt$$

$$\boxed{\int \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2} + it\sqrt{q+\varphi^{-1}(1)^2} \sum_\alpha \lambda_\alpha) dt} \tag{65}$$

$$\begin{aligned}
&\frac{p}{N} \log[\int (\prod_\alpha \frac{d\lambda_\alpha}{2\pi}) \frac{e^{K(\lambda,q,\varphi)}}{|\varphi'(\varphi^{-1}(y^\mu))|^n}] \\
&= \frac{p}{N} \log[\int (\prod_\alpha \frac{d\lambda_\alpha}{2\pi}) \frac{dt}{\sqrt{2\pi}|\varphi'(\varphi^{-1}(y^\mu))|^n} \exp(-\frac{1-q}{2} \sum_{\alpha=1}^{n} \lambda_\alpha^2 - \frac{t^2}{2} + it\sqrt{q+\varphi^{-1}(1)^2} \sum_\alpha \lambda_\alpha)] \\
&= \frac{p}{N} \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} [\int \frac{d\lambda}{2\pi|\varphi'(\varphi^{-1}(y^\mu))|} \exp(-\frac{1-q}{2}\lambda^2 + it\sqrt{q+\varphi^{-1}(1)^2}\lambda)]^n]
\end{aligned}$$

To calculate the integral over λ we use the Gaussian trick (45) with $a = \frac{1-q}{2}$ and we fix the ratio $\alpha \equiv \frac{p}{N}$

$$\alpha \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} [\frac{1}{\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|} \exp(-\frac{t^2(q+\varphi^{-1}(1)^2)}{2(1-q)})]^n] \quad (66)$$

We use Taylor series approximation as $n \to 0$ we get:

$$\begin{aligned}
&\alpha \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \exp(n \log([\frac{1}{\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|} \exp(-\frac{t^2(q+\varphi^{-1}(1)^2)}{2(1-q)})]))] \\
&= \alpha \log[\int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} (1 + n \log([\frac{1}{\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|} \exp(-\frac{t^2(q+\varphi^{-1}(1)^2)}{2(1-q)})]))] \\
&= \alpha \log[1 + n \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \log([\frac{1}{\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|} \exp(-\frac{t^2(q+\varphi^{-1}(1)^2)}{2(1-q)})])] \\
&= n\alpha \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \log([\frac{1}{\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|} \exp(-\frac{t^2(q+\varphi^{-1}(1)^2)}{2(1-q)})]) \\
&= n\alpha \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} (\log(\frac{1}{\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|}) + (-\frac{t^2(q+\varphi^{-1}(1)^2)}{2(1-q)})) \\
&= n\alpha (\log(\frac{1}{\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|}) + \int \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} (-\frac{t^2(q+\varphi^{-1}(1)^2)}{2(1-q)}))
\end{aligned}$$

The second term is the variance of a standard normal distribution so it equals 1.

$$\boxed{n\alpha[-\log(\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|) - \frac{(q+\varphi^{-1}(1)^2)}{2(1-q)}]} \quad (67)$$

The other terms are evaluated in the same way as we showed with Gardner. So we get:

$$\boxed{\begin{aligned} \frac{1}{n} G(q) = \alpha[-\log(\sqrt{2\pi(1-q)}|\varphi'(\varphi^{-1}(y^\mu))|) - \frac{(q+\varphi^{-1}(1)^2)}{2(1-q)}] \\ + \frac{1}{2}(\log(2\pi) + \log(1-q) + \frac{q}{(1-q)} + 1) \end{aligned}} \quad (68)$$

We set $\frac{\delta G}{\delta q} = 0$ to find the saddle point gives:

$$\begin{aligned}
&\alpha[\frac{1}{2(1-q)} - \frac{(1-q) + (q + \varphi^{-1}(1)^2)}{2(1-q)^2}] + \frac{1}{2}(-\frac{1}{1-q} + \frac{1}{(1-q)^2}) = 0 \\
&\Leftrightarrow \alpha[-1 + \frac{1 + \varphi^{-1}(1)^2}{(1-q)}] = -1 + \frac{1}{(1-q)} \\
&\Leftrightarrow \alpha[\frac{q + \varphi^{-1}(1)^2}{(1-q)}] = \frac{q}{1-q} \\
&\Leftrightarrow \alpha[q + \varphi^{-1}(1)^2] = q \\
&\Leftrightarrow \alpha = \frac{q}{q + \varphi^{-1}(1)^2}
\end{aligned}$$

We determine α_c the maximum storage capacity of the network as q tends to its maximum value which is 1.

$$\boxed{\alpha_c = \frac{1}{1 + \varphi^{-1}(1)^2}} \tag{69}$$

References

[1] Cover (1965). *Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition.*. IEEE Trans Electron Comput, 14:326-334.

[2] E. Gardner (1987). *The space of interactions in neural network models.* J. Phys A: Math.Gen 21 (1988) 257-270.

[3] Haim Sompolinsky, MIT. *Introduction: The Perceptron.*

[4] John Hertz, Anders Krogh, Richard G.Palmer. *Introduction to the theory of neural computation.* Santa Fe institute studies in the sciences of complexity.