SPandML Autumn 2023 EXERCISE 04: The annealed solution to randomised perceptron learning
David Villa Blanco

In this exercise we are considering a student perceptron which has N binary input elements and one binary output that is receiving the classification results for P = αN random examples from an identical teacher perceptron. The generalization error $\epsilon(\alpha)$ for the student achieving zero training error has the form:

$$\alpha(\epsilon) = (1 - \epsilon)\pi cot(\pi),$$

according to the annealed computation.

Our main goal is to plot generalization errors obtained at exercise 03 with this function $\alpha(\epsilon)$, which will be plotted in two different ways. The first one being parametic ploting, and the second one through a recursive formula given by:

$$\epsilon_{i+1} = (1 - a)\epsilon_i + af(\epsilon_i, \alpha),$$

specifying the values of ϵ_0, α and a, that will be chosen after analyzing the behavior of the method for various $f(\epsilon, \alpha)$.

1 Parametric plotting

Let $t = \epsilon \in [0 : 0.5]$, we have to plot $\alpha(\epsilon)$ with enough points (chosen 1000). On Figure 1 $\alpha(\epsilon)$ is represented for values of $\epsilon \in [0 : 0.5]$.

On the other hand, on Figure 2 $\alpha(\epsilon)$ is plotted for $\alpha \in [0, 10]$ and with the values of the generalization errors obtained at exercise 03 for $N = 20$ and $N = 40$

2 Iterative procedure

If we rewrite $\alpha(\epsilon)$ as $\epsilon = f(\epsilon, \alpha)$ and choose a value for ϵ_0, we can obtain the following ϵ) value through the relation:

$$\epsilon_{i+1} = (1 - a)\epsilon_i + af(\epsilon_i, \alpha).$$

Ploting ϵ_i vs i for different values of ϵ_0, α and a we can analise the convergence and precision of the method for a given $f(\epsilon_i, \alpha)$ and chosen parameters.

From $\alpha(\epsilon)$ equation we can satisfy $\epsilon = f(\epsilon, \alpha)$ with two differents $f(\epsilon, \alpha)$,

$$f_1(\epsilon, \alpha) = 1 - \alpha tan(\pi\epsilon)/\pi,$$

$$f_2(\epsilon, \alpha) = arctan(\pi(1 - \epsilon)/\alpha)/\pi.$$

Given the values $\alpha = 5$, $\alpha_0 = 0.25$ and $a \in [0.1, 0.5, 0.9]$, on Figure 3 we easily see the problems of choosing $f_1(\epsilon, \alpha)$ as $f(\epsilon, \alpha)$. Due to the behavior of the function tan, ϵ_i has specific points where its value is not logical.

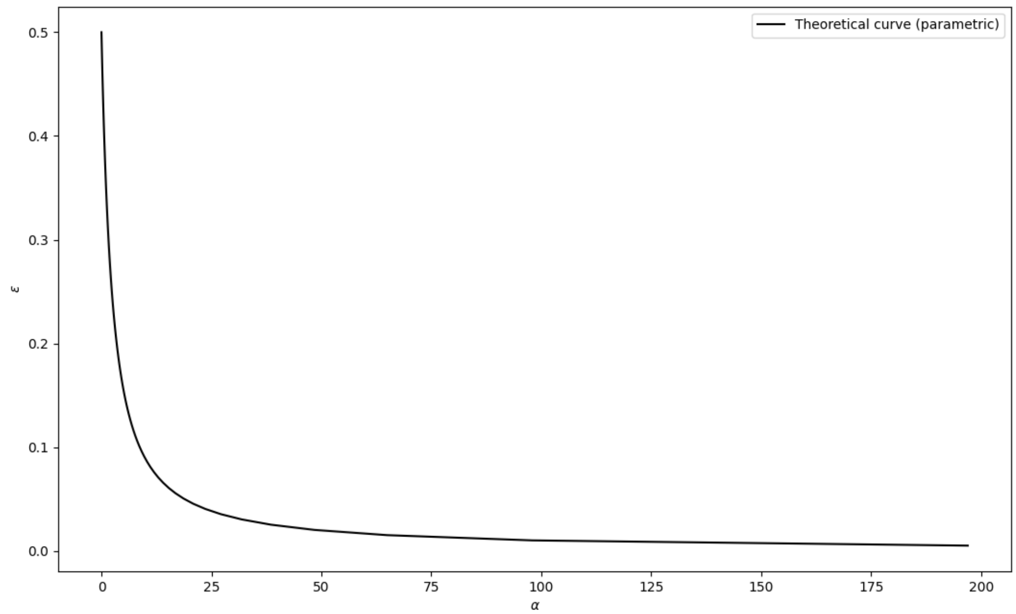

Figure 1: Plot of the parametric curve $\alpha(\epsilon) = (1 - \epsilon)\pi cot(\pi)$ for a given values of ϵ.

On Figure 4 on contrary, the values of ϵ_i seem to converge smoothly. In this figure its also more clear that when a is closer to 0 ϵ_i values tend to converge more slowly but more smoothly while where a is closer to 1 ϵ_i values tend to converge more rapidly but also in a more caotic way.

This will be cruzial for choosing the optimal value of a for the best fit on the values of the generalization errors obtained at exercise 03 because while a values closer to 0 tend to converge more gently, if we iterate enough times on the iteration procedure (this is, getting a enough amount of different ϵ_i values) the plot would be smooth enough, hence the only downside will be the computation cost, which in this specific exercise, is not that important. Given all this, I choose the values of $\alpha_0 = 0.50$ (because our first points for $N = 20$ and $N = 40$ are near 0.5 instead of 0.25), $a = 0.9$ and $\alpha \in [0 : 10]$ and the function $f_1(\epsilon, \alpha) = 1 - \alpha tan(\pi\epsilon)/\pi$ for adjusting the experimental values.

Finally, on Figure 5 both parametric and iterative curves are plotted with the experimental results of the previous exercise. Differences between curve and experimental values could be explained due to the finite dimension of the training sets, while we see that parametric and iterative curse seem to be very similar (in the order of magnitud of the difference between the parametric curve and the experimental values).

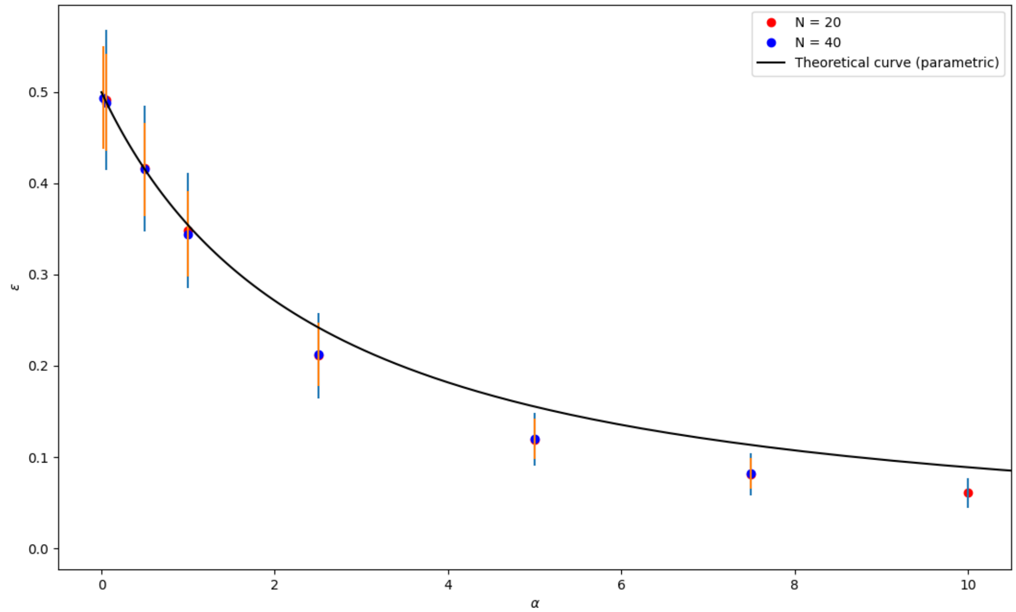

Figure 2: Plot of the parametric curve $\alpha(\epsilon) = (1 - \epsilon)\pi cot(\pi)$ for a given values of ϵ with $\alpha \in [0, 10]$ and showing also the generalization errors obtained at exercise 03.

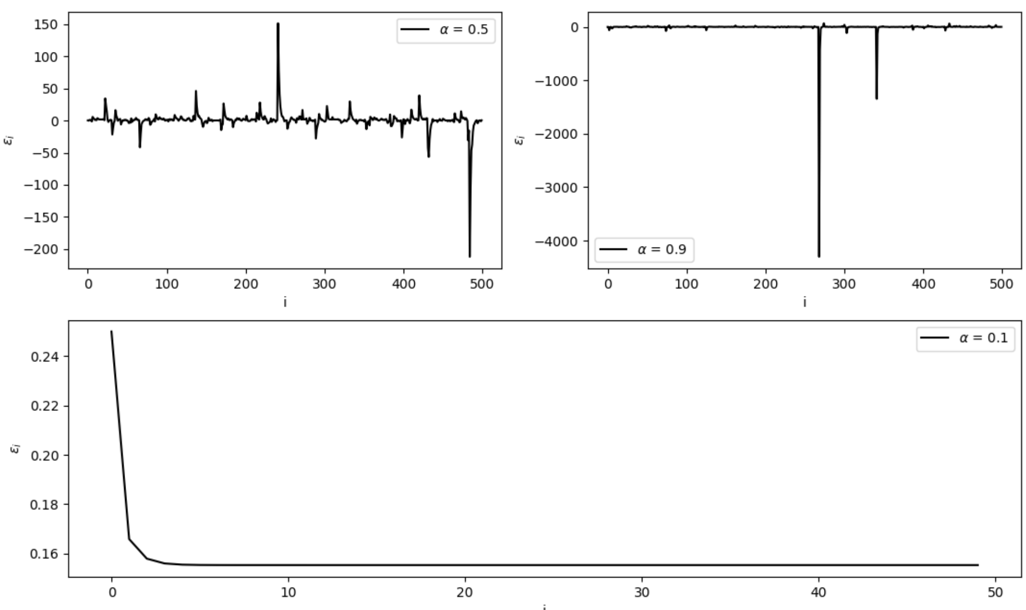

Figure 3: Plots of the iterative curve $\epsilon(\alpha)$ for $f_1(\epsilon, \alpha) = 1 - \alpha tan(\pi\epsilon)/\pi$ $\alpha = 5$, $\alpha_0 = 0.25$ and $a \in [0.1, 0.5, 0.9]$.

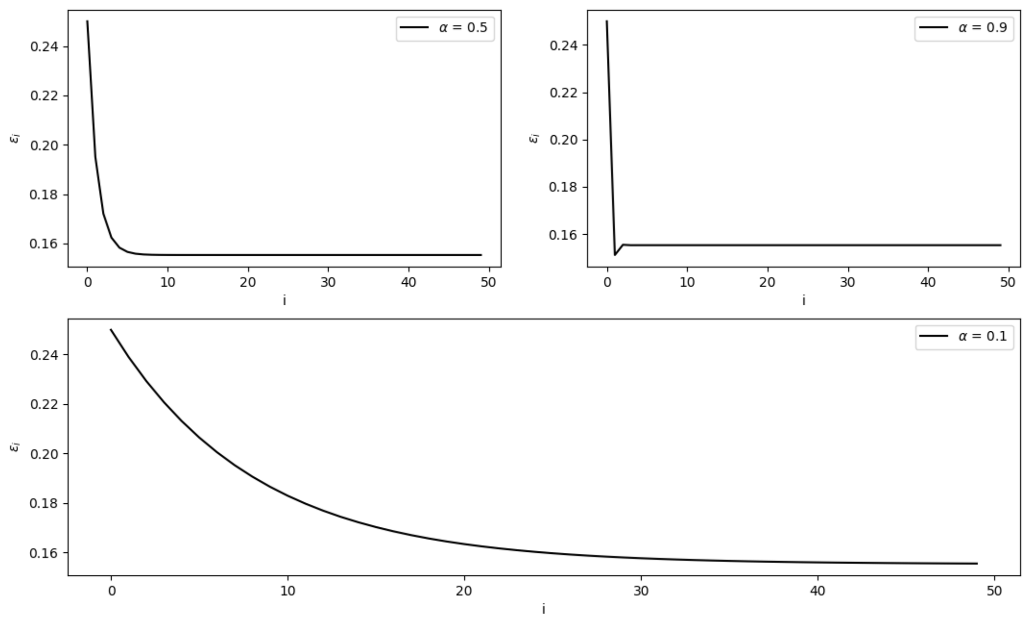

Figure 4: Plots of the iterative curve $\epsilon(\alpha)$ for $f_2(\epsilon, \alpha) = arctan(\pi(1-\epsilon)/\alpha)/\pi$ $\alpha = 5$, $\alpha_0 = 0.25$ and $a \in [0.1, 0.5, 0.9]$.

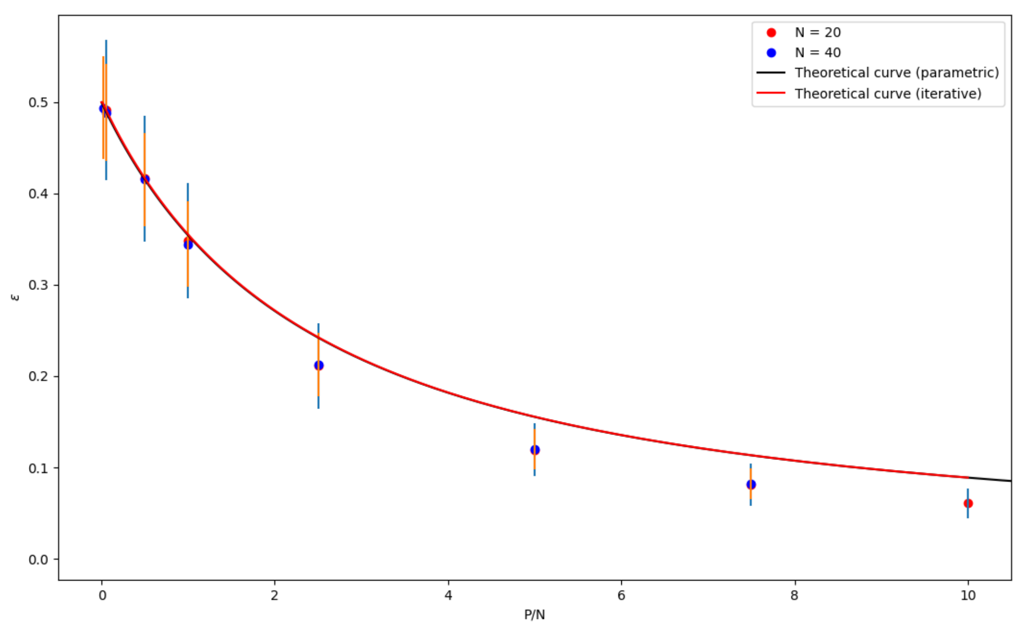

Figure 5: Plots of the iterative curve $\epsilon(\alpha)$ for $f_2(\epsilon, \alpha) = arctan(\pi(1-\epsilon)/\alpha)/\pi$, $\alpha_0 = 0.25$ and $a = 0.9$, the parametric curve $\alpha(\epsilon) = (1-\epsilon)\pi cot(\pi)$ and the experimental values.