EXERCISE 04: The annealed solution to randomised perceptron learning

Consider a student perceptron with N binary input elements and one binary output receiving the classification results for $P = \alpha N$ random examples from an identical teacher perceptron. According to the annealed computation, the typical student achieving zero training error also achieves generalisation error $\epsilon(\alpha)$ such that

$$\alpha = (1-\epsilon)\pi \cot(\pi\epsilon) \ . \tag{1}$$

Superimpose to the numerical results obtained in Exercise 03 the result of equation (1) obtained by the following methods:

- plot the two dimensional parametric curve describing $\epsilon(\alpha)$ with parameter t (the natural choice is to set $t = \epsilon \in [0 : 0.5]$).
- consider a strategy to solve explicitly equation (1) in correspondence of a given α by setting up an iterative procedure:
  - rewrite the equation as $\epsilon = f(\epsilon, \alpha)$
  - start with ϵ_0
  - evaluate $\epsilon_{i+1} = (1-a)\epsilon_i + af(\epsilon, \alpha)$
  - plot ϵ_i *vs* i to check if it converges
- check that a possible choice of $f(\epsilon, \alpha)$ is $f(\epsilon, \alpha) = 1 - \alpha \tan(\pi\epsilon)/\pi$ and show the plot ϵ_i *vs* i for
  - $\alpha = 5$, $\epsilon_0 = 0.25$, $a = 0.5$
  - $\alpha = 5$, $\epsilon_0 = 0.25$, $a = 0.9$
  - $\alpha = 5$, $\epsilon_0 = 0.25$, $a = 0.1$
- check that another possible choice of $f(\epsilon, \alpha)$ is $f(\epsilon, \alpha) = \arctan(\pi(1-\epsilon)/\alpha)/\pi$ and show the plot ϵ_i *vs* i for
  - $\alpha = 5$, $\epsilon_0 = 0.25$, $a = 0.5$
  - $\alpha = 5$, $\epsilon_0 = 0.25$, $a = 0.9$
  - $\alpha = 5$, $\epsilon_0 = 0.25$, $a = 0.1$
- how do you explain the results?
- play further with the parameters ϵ_0 and a at different α to decide which $f(\epsilon, \alpha)$ is best suited for the iterative equation
- use the chosen $f(\epsilon, \alpha)$ to set up an automatic procedure to determine the result $\epsilon(\alpha)$ with high precision in correspondence of many $\alpha \in [0 : 10]$
- plot the obtained curve $\epsilon(\alpha)$ on top of the parametric plot and of the numerical result from exercise 03.