EXERCISE 03: The randomised perceptron learning rule

Consider a perceptron which takes as input two binary numbers of $N/2$ digits each and restitutes as an output either $+1$ or -1. The input vector $\vec{S}$ has N elements $S_i = \{\pm 1\}$. The first $N/2$ elements represent the first number n_T and the second $N/2$ elements represent the second number n_B. The perceptron has N synaptic couplings J_i and must be able to tell which number is larger by giving as an output $\sigma(\{S_i\},\{J_i\}) = \mathrm{sign}(\sum_{i=1}^N S_i J_i)$ such that $\sigma = 1$ if $n_T > n_B$ and $\sigma = -1$ if $n_T \leq n_B$.

(1) Use the perfect perceptron (with synaptic weights $J_i^* = 2^{N/2-i} \quad 1 \leq i \leq N/2$ and $J_i^* = -J_{i-N/2}^* \quad N/2+1 \leq i \leq N$) as teacher perceptron: $T_i = J_i^*$ and $\sigma_T^\mu = \sigma(\{S_i^\mu\},\{T_i\})$.

(2) Initialise a student perceptron with the synaptic weights obtained as random values extracted from a Gaussian distribution $\mathcal{N}(0,1)$.

(3) For N=20 produce 7 sets of $P = 1, 10, 20, 50, 100, 150, 200$ examples ξ^μ with correct classifications σ_T^μ obtained by the teacher perceptron.

(4) For each training set train the perceptron using the following perceptron update rule: only if $\sigma(\{\xi_i^\mu\},\{J_i\}) \neq \sigma_T^\mu$ update the synaptic couplings $J_i(t+1) = J_i(t) + \frac{1}{\sqrt{N}}\sigma_T^\mu \xi_i^\mu * (1 + r_i)$ with r_i a Gaussian random number with null mean and variance $\sigma^2 = 50$. The rule should be applied across an entire cycle of the inputs in the training set, a training error ϵ_t is evaluated at the end of the cycle and the process repeated until $\epsilon_t = 0$.

(5) Evaluate the corresponding test error $\epsilon = P_{test}^{-1}\sum_\mu \theta(\sigma(\{S_i^\mu\},\{J_i\}),\sigma_T^\mu)$ of the trained student perceptrons with $P_{test} = 1000$.

(6) Repeat the procedure 1000 times and evaluate the average test error (and its error bar).

(7) Plot ϵ vs P.

(8) Repeat the entire study (points from (3) to (6)) for $N = 40$ and $P = 1, 2, 20, 40, 100, 200, 300$.

(9) Plot ϵ vs P for $N = 40$ together with the data obtained with $N = 20$.

(10) Find a way to represent the data so that they superimpose and comment on what done.