WASP course – Machine Learning

Exercice 1. Consider the following generative linear model: $y_i = a.x_i + n_i, i = 1, \ldots, n$, where a is an unknown scalar, and the noise process n_i is such that $E(n_i) = 0$ and $Cov(n_i, n_j) = \sigma^2 \delta_{ij}$.
a. Derive the LSE estimator $\hat{a}$.
b. We propose a different estimator $\tilde{a}$ constructed by considering the straight line going through the origin and the centre of gravity of the data. Compute $\tilde{a}$.
c. Are $\tilde{a}$ and $\hat{a}$ biased?
*d. Compare the variance of the two estimators.

Exercice 2. Reinforcement learning. An agent has an asset to sell. She sequentially receives N offers $w_1, \ldots, w_N$ i.i.d. with density $f(w)$. At time i, after receiving the offer w_i, the decision maker has to decide whether to accept the offer or to reject it. If the offer is accepted, the reward is $(1+r)^{N-i} w_i$, where $r > 0$ denotes the interest rate. Once she accepted an offer, the subsequent offers do not matter. The problem is to define a strategy maximizing the expected reward
a. Formulate the problem of identifying the best policy as a Markov Decision Process.
b. Compute the best ?threshold-based? policy using dynamic programming. Note: at step k, a threshold based policy accepts the offer if and only if the offer is greater than a threshold α_k.
c. The agent has now many identical assets to sell sequentially. For each asset, as before, she receives i.i.d. offers with unknown distribution (the density f is unknown). Propose a simple algorithm whose strategy converges to the optimal strategy identified in b.