Questions on Large-scale Machine Learning

Q1. Consider the following data matrix:

x1	x2	y
0	0	1
0	1	-1
1	0	-1
1	1	1

Run SGD on this data for training a linear classifier with hinge loss function, but a constant learning rate of 1. Will this ever converge?

Q2. On the above problem, run SGD with step size reducing as:

i. 1/t
ii. $1/\sqrt{t}$

Report the error as a function of iterations for the following situations:

i. Final result is the average of last 100 parameters.
ii. Final result is the last 1 parameter.

Q3. Formulate and write the ADMM iterations for the pure consensus problem, where objective function is constant for each sub-problem:

$$\min_{x_1,\ldots,x_n,z} 1$$
$$subject\ to: x_i = z\ \forall i$$

This should lead to an algorithm for n computers trying to arrive at a consensus. What is the consensus value?

Q4. Formulate the dual decomposition algorithm for the above problem.