3. (a) Give the pseudocode for stochastic gradient descent with a *linear unit* perceptron. Explain the differences between batch-mode and stochastic gradient descent. [8]

   (b) Give the primal optimisation problem for both the Hard Margin and Soft Margin SVM. Explain the role of the slack variables, and parameter C. How do we set C? [7]

   (c) Given the training examples in Table 3, a fixed learning rate of $\eta = 0.1$, and initial values for the parameters $w_0 = 0.1$, $w_1 = 0.1$, $w_2 = -0.1$, perform the first 3 parameter updates for the standard perceptron with a Heaviside step activation function. Show your working. [10]

x_1 (first attribute)	x_2 (second attribute)	**Class Variable**
0.5	0.5	-1
2	2	1
2	-1	1
1	-1.5	-1
-1	-1	-1

ıble 3: A toy dataset with 5 observations (rows), 2 attributes (x_1, x_2) and the target class (Class ariable)