Consider a perceptron algorithm (batch mode) implementation with initialization $\mathbf{w}^0$ as random initialization learning rate η as 0.1 and termination criteria as "if $||\mathbf{w}^{k+1} - \mathbf{w}^k||_2^2 < 10^{-6}$, terminate".

Consider a training set of 5 positive and 5 negative samples. They are linearly separable. We also have 10 test samples (5 each from both classes). They are also linearly separable.

(A) **[Ans]** Perceptron algorithm will converge.

(B) Perceptron algorithm will not converge.

(C) **[Ans]** Error (number of Mis-classification) on the training data is guaranteed to be zero.

(D) Error (number of Mis- classification) on the test data is guaranteed to be zero.

(E) none of the above

Consider a perceptron algorithm (batch mode) implementation with initialization $\mathbf{w}^0$ as random initialization learning rate η as 0.1 and termination criteria as "if $||\mathbf{w}^{k+1} - \mathbf{w}^k||_2^2 < 10^{-6}$, terminate".

Consider a training set of 5 positive and 5 negative samples. They are <u>not</u> linearly separable. We also have 10 test samples (5 each from both classes). They are linearly separable.

(A) This algorithm will converge.

(B) **[Ans]** This algorithm will not converge.

(C) Error (number of Mis-classification) on the training data is guaranteed to be zero.

(D) Error (number of Mis- classification) on the test data is guaranteed to be zero.

(E) none of the above

Consider a perceptron algorithm (batch mode) implementation with initialization $\mathbf{w}^0$ as random initialization learning rate η as 0.1 and termination criteria as "if $||\mathbf{w}^{k+1} - \mathbf{w}^k||_2^2 < 10^{-6}$, terminate".

In each iteration, we modify the learning rate as $\eta^{k+1} \leftarrow 0.8\eta^k$.

Consider a training set of 5 positive and 5 negative samples. They are <u>not</u> linearly separable. Then:

(A) **[Ans]** This algorithm will converge.

(B) This algorithm will not converge.

(C) This algorithm will oscillate.

(D) Error (number of Mis-classification) on the training data is guaranteed to be zero.

(E) none of the above

Consider a perceptron algorithm (batch mode) implementation with initialization $\mathbf{w}^0$ as random initialization learning rate η as 0.1 and termination criteria as "if $||\mathbf{w}^{k+1} - \mathbf{w}^k||_2^2 < 10^{-6}$, terminate".

Consider a training set of 5 positive and 5 negative samples. They are linearly separable.

We run this implementation 10 times as:

(A) **[Ans]** This algorithm will converge to a valid solution irrespective of the initialization. (assume learning rate fixed at 0.1)

(B) This algorithm will converge to the same solution irrespective of the initialization.(assume learning rate fixed at 0.1)

(C) **[Ans]** This algorithm will converge to a valid solution irrespective of the learning rate (say in the range 0.05 to 0.2). (assume the initialization is same in all cases.)

(D) This algorithm will converge to the same solution irrespective of the learning rate (say in the range 0.05 to 0.2). (assume the initialization is same in all cases.)

(E) **[Ans]** This algorithm will converge to the same solution irrespective of the relative ordering of the data. (i.e., data set was shuffled across runs) (assume initialization is same in all cases and η is fixed as 0.1)

Consider a perceptron algorithm (batch mode) implementation with initialization $\mathbf{w}^0$ as random initialization learning rate η as 0.1 and termination criteria as "if $||\mathbf{w}^{k+1} - \mathbf{w}^k||_2^2 < 10^{-6}$, terminate".

Consider a training set of 5 positive and 5 negative samples. They are <u>not</u> linearly separable.

(A) Since the problem is non-convex, if we can find a right initialization, we will get the best solution very fast.

(B) **[Ans]** Irrespective of the initalization, implementation will oscillate/cycle.

(C) **[Ans]** Assume the termination criteria was "if there is no change in the misclassification rate across two iterations, terminate". Then the implementation could have converged.

(D) Assume the termination criteria was "if there is no change in the mis-classification rate across two iterations, terminate". Even then the implementation will never converge.

(E) None of above.