Consider a typical two class classification problem. We have a labelled set of 1000 samples. (say $N = 1000, d = 2$).

We train a classifier iterative (say using GD) and minimize a loss function (eg. a mean square error loss).

We use 80% data for training (i.e., 800) and rest 20% (i.e., 200) for testing.

(A) To train the model, We iterate until the loss on the training data becomes zero.

(B) If we allow to continue for enough time the the loss on training data will eventually become zero.

(C) If at any point of time, the loss on training data is zero, then the loss on test data will also be zero.

(D) If both training and test loss are zero, then we are sure that the algorithm has overfit.

(E) **[Ans]** None of the above.

Consider a two class classification problem. We have a labelled set of 1000 samples. (say $N = 1000, d = 2$).

We train a classifier iterative (say GD) and minimize a loss function (eg. a mean square error loss) by using all the samples.

In each iteration, we use a random 80% of the total data as training (i.e., 800) and rest 20% (i.e., 200) for testing.

(A) This is perfectly fine way of training the ML solution.

(B) This iterative algorithm will not converge.

(C) Since the test data is changed on a regular basis, the solution will generalize well.

(D) Since the training data is changing in every iteration (or regular basis), the loss will not come down.

(E) **[Ans]** None of the above

Consider a typical two class classification problem. We have a labelled set of 1000 samples. (say $N = 1000, d = 2$).

We train a classifier iterative (say using a GD) and minimize a loss function (eg. a mean square error loss).

We use 80% data for training (i.e., 800) and rest 20% (i.e., 200) for testing.

During the iteration k, loss on the training data and Test data are L^k_{Tr} and L^k_{Te}. Let the accuracy of the classifier at iteration i be η^i_{Tr} and η^i_{Te}.

(A) **[Ans]** If $L^k_{Tr} \gg L^l_{Tr}$, then $l > k$.

(B) If $L^k_{Tr} \gg L^l_{Tr}$, then $\eta^k_{Tr} > \eta^l_{Tr}$ (strictly greater).

(C) If $L^k_{Tr} > L^l_{Tr}$, then If $L^k_{Te} > L^l_{Te}$.

(D) If $L^k_{Tr} > L^l_{Tr}$, then If $L^k_{Te} < L^l_{Te}$.

(E) None of the above.

In the context of regularization:

(A) L_∞ regularization leads to sparse solution.

(B) L_2 regularization leads to sparse solution.

(C) **[Ans]** L_1 regularization leads to sparse solution.

(D) **[Ans]** L_0 regularization leads to sparse solution.

(E) Any regularization will lead to sparse solution.

In the context of supervised machine learning,

(A) **[Ans]** Overfitting is a modeling error that occurs when a function is too closely fit to a limited set of data points.

(B) **[Ans]** Occam's Razor says: Suppose there exist two explanations for an occurrence. In this case the one that requires the smallest number of assumptions is usually correct.

(C) Supervised learning is all about overfitting to the given data.

(D) **[Ans]** Regularization decrease the chance of overfitting.

(E) None of the above.