__/__/____

SECTION A

True | False

1) True
2) True
3) True
4) True
5) True
6) False
7) True
8) True
9) True
10) True

SECTION - B

1)

Model	Loss function	Regularizer
SVM	hinge loss	L2 norm ($\|w\|^2$)
LASSO	Squared loss	L1 norm ($\|w\|_1$)
RIDGE	Squared loss	L2 norm ($\|w\|^2$)

2) A: Squared loss
B: Absolute loss
C: Hinge loss
D: Logistic loss.
E: Exponential loss.

3) a) Smooth & differentiable losses like Exponential, Logistic, Squared loss. so that gradient exists everywhere
b) Squared & Logistic loss bez twice differentiable

SECTION - C

1) Underfitting happens when model is too simple to learn the pattern in the data
like if we use linear model for very complex data it cannot capture the relationship properly.

__/__/__

2) If both training and testing error are high, it means model is not learning well at all.
This usually implies high bias or the data is very complex or noisy.

3) Bagging trains many models on different random samples of data then averages their output
Because of averaging, random errors of individual models cancel out. so variance reduces.

4) Boosting mainly reduces bias by focusing more on wrong ~~problems~~ predictions.
It can also reduce variance sometimes, but if we boost too much, it may even increases variance.

SECTION-D

1) We can reduce computation time by reducing the size of data or using some faster search methods.
For example, we can use K-D trees, reduce dimensions using PCA, or use approx nearest ~~exact~~ neighbours instead of exact ones.

2.(a) No, squared Euclidean distance does not change predictions because the nearest points remain the same since squaring is monotonic, order of distance does not change.

(b) No it does not affect the ~~the~~ curse of dimentionality conclusion. even if we square the distance, problem of distances becoming similar in high dimensions still remains

__/__/____

3) In high dimensions, data points become very sparse and far from each other.
So the nearest neighbours is not really "near", which makes distance-based methods like KNN unreliable.

4) When k is small, variance is high and bias is low because model is very sensitive to data
When k is large, bias increases and variance decreases since prediction become smooth

5) KNN is preffered when data is non-linear & small in size.
It is also useful when we do not want to assume any specific model form.

SECTION - E.

1) At a leaf, we want to predict one constant value ŷ for all data points
If we take squared loss and try minimise the total error, taking derivative and setting it to zero gives the average of target values.
So the best prediction at a leaf is the mean.

2) minimum Gini is 0, when all samples belong to one class.
Maximum Gini is around 0.67, when all 3 classes are equally likely.

__/__/____

3) Decision trees are called myopic because they make decisions greedily.
At each node, they choose the best split locally without thinking about future splits

4) One way is to limit the depth of the tree so it does not grow too complex.
Another method is pruning, where we remove branches that do not improve performance much.

SECTION-F

1) Random Forests usually should not use the exact same data for training and testing because it will give over-optimistic accuracy.
However, internally they use bootstrap sampling, so some data points are not used while training a tree and can be used for testing.

2) In bagging, models are trained independently on random samples of the data and then averaged.
In boosting, models are trained one after another and more importance is given to wrong prediction of previous model.