Section - A → True / false

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

Section - B → ERM & SVM

1) Fill the following table:

Model	Loss function	Regularizer
SVM	Hinge Loss $\max(0, 1 - y\hat{y})$	L2 Regularization
LASSO.	Squared Loss $(y-\hat{y})^2$	L1 Regularization
RIDGE	Squared Loss $(y-\hat{y})^2$	L2 Regularization

3) Short answer:

a) Loss functions that are differentiable can be optimized using gradient descent. Gradient descent uses first order derivative.
Ex. of such loss function - Mean squared loss, Logistic loss

b) Loss functions that are twice differentiable can be optimized by newton's method because newton's method require second-order derivatives.

Section - C → Bias & Variance

1) Underfitting of model occurs when model is too simple to capture the pattern in the data. When the model is not complex enough.

2) If both training & test error remain high, it implies that the data is underfitted, model cannot learn the data distribution properly

3) Bagging reduces variance by training multiple models on different random samples of the dataset & then averaging their prediction, which smoothes the errors.

4) Boosting helps reduces bias by repeatedly focusing on data points that were predicted incorrectly. By doing this, model learns more complex pattern in the data. If boosting is applied too many times, it may increase variance.

~~Section - D → KNN & Curse of dimensionality~~

Section - E → Decision Trees

1) Let a leaf node contains n data pt. with true target values: $y_1, y_2, y_3, \ldots y_n$

& prediction made at this leaf be c (const)

$$L(c) = \sum_{i=1}^{n} (y_i - c)^2$$

$$\frac{dL}{dc} = \sum_{i=1}^{n} -2(y_i - c)$$

$$\Rightarrow \frac{dL}{dc} = 0$$

$$\sum_{i=1}^{n} (c - y_i) = 0$$

$$nc = \sum_{i=1}^{n} y_i$$

$$\Rightarrow \boxed{c = \frac{1}{n}\sum_{i=1}^{n} y_i}$$

2) Gini Impurity : $G = 1 - \sum_{K} P_K^2$ → probability for each K

Minimum Gini impurity = 0
→ This occurs when all samples in the node belongs to the same class

Maximum gini impurity = 2/3
→ This occurs when all the three class are equally likely.

3) Decision trees are called myopic learners because they make decisions one step at a time. At each node, the tree chooses the best split based only on current situation & doesn't consider how it will ~~~~ affect future splits.

4) i) limiting the depth of the tree, which prevents the model from ~~~~ becoming too complex.
ii) Removing branches that do not improve performance on validation data.

Section-F → Boosting & Bagging

1) No, ~~~~ Random forest should not use the exact same data for training & testing if we want an unbiased & realistic measure of model's performance on new, unseen data. If we ~~~~ test on same data, model will achieves nearly 100% accuracy which will overfit. Even though each tree in random forest is trained on bootstrap, but testing must always be done on a seperate dataset to measure generalization performance

2) Bagging & Boosting are both ensemble methods.
In Bagging, models are trained independently (parallel) on different random samples of data & then their predictions are averaged. This mainly reduces variance.
In boosting, models are trained one after another & each new model focuses to reduces the earlier misclassified data points. It mainly reduces bias.