PAGE NO.
DATE: / /

Assignment - 1

Section - A

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

Section - B

(1)

Model	Loss function	Regularizer
SVM	Hinge Loss	$\|W\|^2$ (L2 norm)
LASSO	Squared Loss (L2 loss)	$\|W\|^2$ (L2 norm)
RIDGE	Squared Loss (L2 loss)	$\|W\|^2$ (L2 norm)

(2)
A: 0-1 Loss
B: Hinge Loss
C: Squared Loss
D: Absolute Loss (L1 Loss)
E: Logistic Loss

(3) (a) Squared Loss, Logistic Loss, Absolute Loss are optimizable using gradient descent because they are differentiable and their gradient exists.

PAGE NO.
DATE: / /

(6) Squared loss and logistic loss are optimizable using Newton's method because they are twice differentiable.

Section - C

(1) Under-fitting occurs when the model is ~~too~~ simple to capture the underlying pattern in the data.

(2) This means that model has high bias and is underfitting the data.

(3) Bagging reduces variance by averaging predictions from multiple models trained on different bootstraps samples, which cancels out individual model fluctuations.

(4) Boosting reduces bias significantly by focusing on hard-to-predict samples, but it can increase variance if overfitting occurs.

Section - E

(1) To reduce KNN computation time we should
- use efficient data structure (Ball-tree)
- reduce dimensionality
- Use approximate nearest neighbours
- reduce dataset size

PAGE NO.
DATE: / /

2 (a) No,
Squaring preserves distance ordering, so nearest neighbors remain the same.

2 (b) No,
Curse of Dimensionality remains unchanged since relative distances are unaffected.

(3) Distances become less discriminative and data becomes sparse, making nearest neighbors unreliable

(4) $k \downarrow$ bias $\downarrow$ variance $\uparrow$
OR, $k \uparrow$ bias $\uparrow$ variance $\downarrow$

(5) When the decision boundary is highly non-linear and sufficient data is available., KNN is preffered over linear SVM

Section E

(1) Say leaf n nodes . $y_1, y_2, \cdots y_n$

Squared loss :

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

$$\frac{\partial L}{\partial \hat{y}} = \sum 2(\hat{y} - y_i)$$

for minimum loss

$$\frac{\partial L}{\partial \hat{y}} = 0 \rightarrow \boxed{\hat{y} = \frac{1}{n} \sum_{i=1}^{n} y_i} = \text{mean value}$$

PAGE NO.
DATE: / /

(2) minimum = 0 (pure node)

$$\text{Maximum} = 1 - \sum_{i=1}^{3} \left(\frac{1}{3}\right)^2 = 2/3$$

(3) They make greedy, locally optimal splits without considering future splits.

(4) 1) Limit tree depth / minimum samples per leaf (pre-pruning)
2) Post-pruning using validation data

Section-F

(1) No,
Trees are trained on bootstrap samples while testing is done using out-of-bag (OOB) samples or separate test data.

(2)

Bagging	Boosting
Trains models independently to reduce variance	Trains model sequentially, focusing on previous errors to reduce bias