Date: / /

Assignment-1

Section-A

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

Section-B

1.

Model	Loss function	Regularizer
SVM	Hinge loss $[\max(0, 1-y_i f(x_i))]$	$\frac{1}{2}\|w\|^2$ [L2]
LASSO	Mean-squared error (MSE): $(y_i - \hat{y}_i)^2$	$\sum_i \|w_i\|$ [L1]
RIDGE	"	$\sum_i w_i^2$ [L1]

3. (a) Loss functions that are continuous and differentiable can be optimised using gradient descent. That is because gradient descent requires the function to be differentiable.

Date: / /

(b) Loss functions that are twice differentiable and continuous are optimized by newton's method. This is because newton's method requires the use of double differentiation.

Section-C

1. When the bias is high, the predicted data is in the form of a straight line. This causes the fitting of the actual data to be inaccurate. This causes underfitting of data.

2. It implies underfitting of the given data.

3. Firstly, multiple training sets are created by sampling with replacement from the original dataset. Each sample is slightly different. Then, a seperate model is trained on each bootstrap sample. Each model makes different errors. Then, we average out the predictions because that cancels out the positive and negative noise upon averaging leading to a better output. This reduces overall variance.

4. Boosting is used mainly to reduce bias but it can also reduce variance.

Date: / /

Section-E

1. Consider a leaf containing target values $y_1, y_2, \ldots, y_n$. The leaf predicts a constant value c. The squared loss at the leaf $\Rightarrow L(c) = \sum_{i=1}^{n} (y_i - c)^2$

To minimize the loss function;

$$L'(c) = \sum_{i=1}^{n} 2(c - y_i) = 0$$

$$cn - \sum_{i=1}^{n} y_i = 0 \qquad c = \frac{1}{n} \sum_{i=1}^{n} y_i$$

∴ The optimal prediction at a leaf under squared loss is the mean of target values.

2. $G_n = 1 - \sum_{k=1}^{3} p_k^2$

G_{min} is when either one of p_1 or p_2 or $p_3 = 1$ and the other two zero.

$G_{min} = 1 - 1 = 0$

G_{max} is when $p_1 = p_2 = p_3 = \frac{1}{3}$.

$G_{max} = 1 - 3 \times \frac{1}{9} = \frac{2}{3}$

3. Decision trees are myopic because they make locally optimum split decisions without considering the long-term effect on the overall tree.

Date: / /

4. Methods to avoid overfitting:
a) limit number of features per split
b) using ensemble methods (random forests)

Section-F

1. Yes. In random forests, each tree is trained on a bootstrap sample of the data, so some data points are left out of training for that tree. These left out points are used to test the tree. Hence, the same dataset is used for both training and testing without testing a tree on the data it was trained on.

2.

Bagging	Boosting
1. Training of data is done parallely.	Training of data is done sequentially.
2. Sampling is done by uniform bootstrap sampling	Sampling is done by reweighted sampling
3. All points are equally focused	Main focus is given on hard misclassified points.
4. Its main effect is to reduce variance	Main effect is to reduce bias
5. It is resistant to overfitting	Overfitting is possible.