Week 6: Revision

Q2. We have a dataset with 5 observations in the form of (X, y):

$$D = \{(1,10), (2,6), (3,0), (5,7), (8,12)\}$$

Below is a table which shows 4 bootstrapped training sets (BTS) and the predictions from each BTS.

BTS	Prediction for X= 1	2	3	5	8
(1,10), (1,10), (1,10), (3,0), (8,12)	14	7	0	5	10
(3,0), (3,0), (5,7), (5,7), (5,7)	8	-1	1	7	13
(1,10), (2,6), (2,6), (5,7), (5,7)	13	4	4	10	16
(3,0), (5,7), (5,7), (8,12), (8,12)	12	7	-1	8	14

The OOB prediction for $X=1$ is $\frac{8+12}{2} = \frac{20}{2} = 10$

The OOB prediction for $X=2$ is $\frac{7-1+7}{3} = \frac{13}{3} = 4.\dot{3}$

The OOB prediction for $X=3$ is 4

The OOB prediction for $X=5$ is 5

The OOB prediction for $X=8$ is $\frac{13+16}{2} = 14.5$

The OOB MSE is $\frac{1}{n}\sum_{j=1}^{n}$ (label of D − OOB prediction)

The OOB MSE $= \frac{1}{n}\left[(10-10)^2 + (6-4.3)^2 + (0-4)^2 + (7-5)^2 + (12-14.5)^2\right]$

$$= \frac{1}{5}(0 + 2.89 + 16 + 4 + 6.25) = \frac{1}{5} \times 29.14 = 5.828$$

Q1. Which of the followings are possible outcomes of the R code.
sample (12, replace = TRUE)

b. 7 3 12 5 5 1 12 11 4 4 2 7
d. 4 4 4 4 4 4 4 4 4 4 4 4

Q3. - A dataset has p attributes. When applying ordinary recursive binary splitting to generate a decision tree, at each node, all the p attributes are considered.

- When applying random forest technique to generate a decision tree for a regression problem, we should consider P/3 random attributes at each node.

- When applying random forest technique to generate a decision tree for a classification problem, we should consider $\sqrt{P}$ (square root of P) random attributes at each node.

- Random forest method reduces variance

Q4. We applying boosting algorithm with learning rate 0.5.
An observation is in form of (x, y),
where; the attribute $x=5$ and the label $y=9$
In the first round of boosting algorithm, the prediction function $\hat{f}^1$ is generated.
It is found that $\hat{f}'(5) = 8$.
What is the updated residue of the observation after the first round? $r_i = y_i$

Observation	Prediction	$r_i - \eta.\hat{f}'(x)$	Update Observation
(x, y)	$\hat{f}^1(x) =$		
$(5, 9)$	$\hat{f}^1(5) = 8$	$9 - 0.5 \times 8 = 5$	$(5, 5)$

∴ 5

Q5. In both random forest and boosting, we generate B decision trees. when B keeps increasing, overfitting does not occur for random forest, and overfitting May occur for housting.

Q6. What is the printout of the following piece of R code? This is to test your understanding of OOP.

```
setClass ("complex", slots= list(real="numeric", imag="numeric"))
c1 <- new("complex", real=2, imag= -3)
c2 <- new ("complex", real=4, imag= -5)

multiply <- function (x,y) {
  r <- x@real * y@real - x@imag * y@imag
  i <- x@real * y@imag + y@real * x@imag
  return (new ("complex", real=r, imag=i))
}
radius.sq <- function (x) {
  return (x@real * x@real + x@imag * x@imag)
}
print (radius.sq( multiply (c1, c2)))
```

$r = 2 \times 4 - (-3) \times (-5) = 8 - 15 = -7 = \text{real}$
$i = 2 \times (-5) + 4 \times (-3) = -10 - 12 = -22 = \text{imag}$

$= -7 \times -7 + (-22) \times (-22) = 49 + 484 = 533$

Week 6: Quiz

Q1. Which of the followings are possible outcomes of the R Code sample(10, 5, replace = FALSE)?

d. 1 2 3 8 4   g. 1 2 5 4 3

f. 10 8 1 5 4

Q2. A dataset has 20 attributes. The labels are categorical/qualitative. When applying random forest technique to generate a decision tree using this dataset, ____ attributes should be sampled at each node.

$\sqrt{20} = 4.47 \approx 4$

Q4. In both random forest and boosting, we generate B decision trees.

- It is fine to choose arbitrarily large B when applying random forest method.
- It is not fine to choose arbitrarily large B when applying boosting method, because when B keeps increasing, overfitting may occur.

Q5. A dataset has 5 observations. When a bootstrapped training set is randomly generated, what is the probability that the 2^{nd} observation is sampled once or more? (Answer must be; a real number between 0 and 1 and to 3 d.p.)

Let $D = \{(1,10), \overset{2^{nd}}{(2,6)}, (3,0), (5,7), (8,12)\}$

Example: 1 bootstrapped training set (BTS).

(1,10), (1,10), (1,10), (3,0), (8,12)

p[once or more] =

Q3. We have a dataset with 6 observations in the form of (X, y):

$D = \{(0,3), (2,6), (3,10), (6,19), (8,20), (10,26)\}$

Below is a table which shows 4 bootstrapped training sets (BTS) and the predictions from each BTS.

	Predictions for $X =$					
BTS	0	2	3	6	8	10
(2,6), (2,6), (2,6), (3,10), (3,10), (8,20)	0	4	6	12	16	20
(3,10), (3,10), (3,10), (6,19), (6,19), (6,19)	4	7	12	27	20	24
(2,6), (2,6), (8,20), (8,20), (10,26), (10,26)	-3	4	9	18	17	30
(3,10), (6,19), (6,19), (6,19), (8,20), (10,26)	-1	7	10	18	14	25

The OOB prediction for $X=1$ is $= \dfrac{0+4+(-3)+(-1)}{4} = \dfrac{0}{4} = 0$

The OOB prediction for $X=2$ is $= \dfrac{7+7}{2} = \dfrac{14}{2} = 7$

The OOB prediction for $X=3 = \dfrac{9}{1} = 9$

The OOB prediction for $X=6 = \dfrac{12+18}{2} = \dfrac{30}{2} = 15$

The OOB prediction for $X=8$ is $\dfrac{20}{1} = 20$

The OOB prediction for $X=10$ is $\dfrac{20+24}{2} = \dfrac{44}{2} = 22$

The OOB MSE is $- \dfrac{1}{n}\sum_{j=1}^{n}$ (label of D $-$ OOB prediction)

$$\text{The OOB MSE} = \frac{1}{6}\left[(3-0)^2 + (6-7)^2 + (10-9)^2 + (19-15)^2 + (20-20)^2 + (26-22)^2\right]$$

$$= \frac{1}{6}[9+1+1+16+0+16) = \frac{1}{6} \times 43 = 7.166$$

$\therefore$ The OOB MSE $= 7.17$

Exam: Ensemble Methods

Q4: a) Why the predictive power of bagging is often better than that of its underlying algorithm?

- Bagging decreases variance, not bias, and solves over-fitting issues in a model.
- Bagging helps reducing variance

b) - Describe the OOB (out-of-bag) procedure to estimate the error of bagging.
- Give a pseudo code for both the regression and classification cases.

Bagging

- Given a large dataset, we sample a number of sub-datasets.
  How? Bootstrapping.
  - Generate B different bootstrapped training sets (BTS) independently.
  - To generate each BTS, we sample n observations with replacement.
- In each BTS, you expect:
  - some observations are sampled more than once.
  - some other observations are not sampled.
- The observations which are not sampled are called out-of-bag (OOB) observations.

c) - Consider a binary classification problem with a training set
- consisting of $n=3$ observations $\{O_1, O_2, O_3\}$ with labels $\begin{array}{ccc} y_1 & y_2 & y_3 \\ \hline 1 & 0 & 0 \end{array}$

- We carry out bagging, with $B=6$ decision trees based on the following bootstrapped training sets
- and giving the following predictions $\hat{y}_1, \hat{y}_2, \hat{y}_3$ for the 3 training observations O_1, O_2, O_3:

bootstrap sample	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$
$\{O_1, O_2, O_2\}$	1	0	0
$\{O_2, O_3, O_3\}$	0	1	0
$\{O_1, O_2, O_3\}$	1	0	0
$\{O_1, O_1, O_2\}$	1	1	0
$\{O_1, O_3, O_3\}$	1	0	0
$\{O_2, O_3, O_3\}$	0	0	1

Find the OOB error rate for this classification problem.

The prediction for O_1 is 0

The prediction for O_2 is 0

The prediction for O_3 is $\frac{0+0}{2} = 0$

Therefore,

The OOB MSE $= \frac{1}{3}\left[(1-0)^2 + (0-0)^2 + (0-0)^2\right] = \frac{1}{3}$