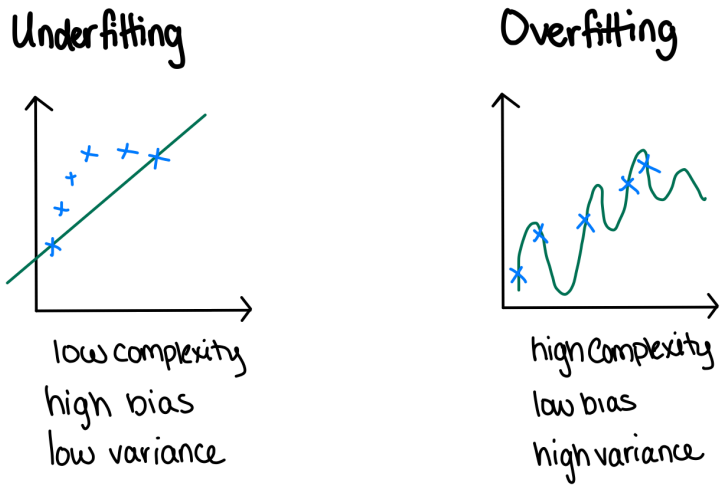
For Questions 1- 4, please submit a word file or a PDF file;
For Question 5 (programming question), please submit an .ipynb file.

Question 1: **[4 points] Explain what is the bias-variance trade-off? Describe few techniques to reduce bias and variance respectively.**

Bias-Variance trade-off is the trade-off in complexity of the model. The conflicting relationship between underfitting (bias) and overfitting (variance) when the model needs to minimize bias and variance in order to find the best solution is what creates this issue. Increasing the bias will decrease the variance and increasing the variance will decrease the bias; A low complexity model will have high bias and low variance while a high complexity model will have low bias and high variance. Therefore, Bias and variance have an inverse relationship.



Low complexity models are prone to underfitting (Accurate but initially incorrect predictions, low error) , high complexity models are prone to overfitting (Accurate, large error).

Techniques to reduce bias error;

1. Change model methods (reduce high bias)
2. Train the data using multiple models (reduce high variance)
3. Feed the model more data (reduce high variance)
4. Ensemble learning methods
5. Using diverse training data

Question 2: [6 points] Assume the following confusion matrix of a classifier. Please compute its
1) precision,
2) recall, and
3) F_1-score.

Predicted results

Actual values		Class 1	Class 2
	Class 1	50	30
	Class 2	40	60

Predicted results

Actual values		Class 1 (Predicted positive)	Class 2 (negative)
	Class 1 (Positive)	50	30
	Class 2 (Negative)	40	60

$$PREC = \frac{TP}{TP + FP} = \frac{50}{50+40} = 0.556$$

(correct positive predictions / Total number of positive predictions)

$$SN\ (recall) = \frac{TP}{TP+FN} = \frac{TP}{P} = \frac{50}{50+30} = 0.63$$

(correct positive predictions / Total number of positives)

* Best sensitivity is 1.0, worst sensitivity = 0.0

$$F_\beta = \frac{(1+\beta^2)(PREC \cdot SN)}{\beta^2 \cdot PREC + REC}$$

$$F_1 = \frac{2 \cdot PREC \cdot SN}{PREC + REC} = \frac{2 \cdot 0.556 \cdot 0.63}{0.556 + 0.63} = \frac{0.701}{1.186} = 0.591$$

Question 3: [10 points] Build a decision tree using the following training instances (using information gain approach):

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Entropy

$$-\sum_{i=1}^{c} P(x_i)\log_b P(x_i)$$

case of yes = 6/10 case of no = 4/10

$-(6/10 \cdot \log_2(6/10) + (4/10 \cdot \log_2(4/10)) = 0.97$

Information Gain

original entropy

$IG(T,A) = \text{entropy}(T) - \sum_{v \in A} \frac{|T_v|}{T} \cdot \text{entropy}(T_v)$

T = Target = play Tennis column

A = variable (column) being tested

v = each value in A

Outlook entropies

Sunny = $-(1/4 \cdot \log_2(1/4)) = 0.5$

Overcast = $-(2/2 \cdot \log_2(2/2)) = 0$

Rain = $-(3/4 \cdot \log_2(3/4)) = 0.31$

OUTLOOK IG

$0.97 - (4/10 \cdot 0.6 + 3/10 \cdot 0 + 4/10 \cdot 0.31) = 0.605$

Temperature entropies

Hot = $-(1/3 \cdot \log_2(1/3)) = 0.53$

mild = $-(2/3 \cdot \log_2(2/3)) = 0.39$

cool = $-(3/4 \cdot \log_2(3/4)) = 0.31$

Temperature IG

$0.97 - (3/10 \cdot 0.53 + 3/10 \cdot 0.39 + 4/10 \cdot 0.31)$

$= 0.57$

Humidity entropies

High = $-(2/5 \cdot \log_2(2/5)) = 0.53$

Normal = $-(4/5 \cdot \log_2(4/5)) = 0.26$

Humidity IG

$0.97 - (5/10 \cdot 0.53 + 5/10 \cdot 0.26)$

$= 0.58$

Wind entropies

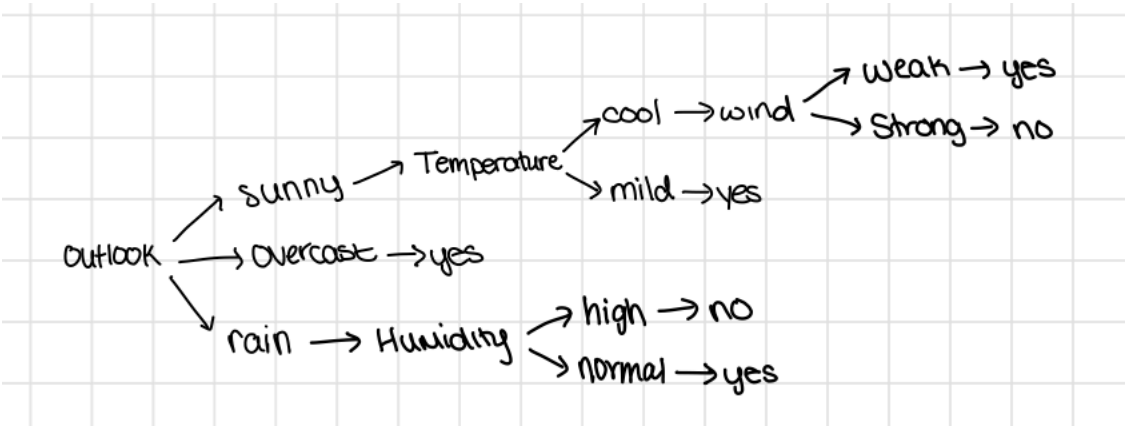
weak = $-(5/7 \cdot \log_2(5/7)) = 0.35$

Strong = $-(1/3 \cdot \log_2(1/3)) = 0.53$

Wind IG

$0.97 - (7/10 \cdot 0.35 + 3/10 \cdot 0.53)$

$= 0.57$



Question 4. [10 points] The naïve Bayes method is an ensemble method as we learned in Module 5. Assuming we have 3 classifiers, and their predicted results are given in the table 1. The confusion matrix of each classifier is given in table 2. Please give the final decision using the Naïve Bayes method:

Table 1 Predicted results of each classifier

Sample x	Result
Classifier 1	Class 1
Classifier 2	Class 1
Classifier 3	Class 2

Table 2 Confusion matrix of each classifier

i) Classifier 1

	Class1	Class2
Class1	40	10
Class2	30	20

ii) Classifier 2

	Class1	Class2
Class1	20	30
Class2	20	30

iii) Classifier 3

	Class1	Class2
Class1	50	0
Class2	40	10

$P(w_1 | d_{1,1}(x)) = 1) = \frac{40}{70}$, $P(w_2 | d_{1,1}(x) = 1) = 30/70$

$P(w_1 | d_{2,1}(x) = 1) = \frac{20}{40}$, $P(w_2 | d_{2,1}(x) = 1) = \frac{20}{40}$

$P(w_1 | d_{3,2}(x) = 1) = 0/70$, $P(w_2 | d_{3,2}(x) = 10/10$

Probability of the two classes

class 1: $40/70 \cdot 20/40 \cdot 0/70 = 0$

class 2: $30/70 \cdot 20/40 \cdot 10/10 = 0.214$

Predicted result is class 2