

Introduction to Machine Learning

– Prof. Balaraman Ravindran | IIT Madras

Problem Solving Session (Week-1)

Shreya Bansal

PMRF PhD Scholar
IIT Ropar

Week-1 Contents

— — —

1. Introduction to Machine Learning
2. Supervised Learning
3. Unsupervised Learning
4. Reinforcement Learning
5. Statistical Decision Theory- Regression
6. Statistical Decision Theory- Classification
7. Bias-Variance

What is Machine Learning?

— — —

Definition by Tom Mitchell (1997):

An agent is said to learn from **experience (E)** with respect to some **class of tasks (T)** and **performance measure (P)** if its performance on tasks in T, as measured by P, improves with experience E.

Why Define Tasks, Performance, and Experience?

— — —

- **Without clear tasks and performance measures, learning becomes vague or unmeasurable.**

Example: “Something seems to improve” is unclear.

- **Examples of Clear Performance Measures:**

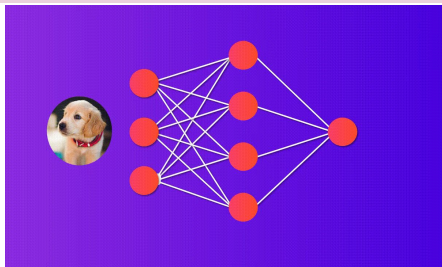
Exam marks for test-taking.

- **Think of learning as training for a marathon.**

Clear tasks (running), measurable performance (time), and experience (practice runs) ensure improvement.

Machine Learning Paradigms

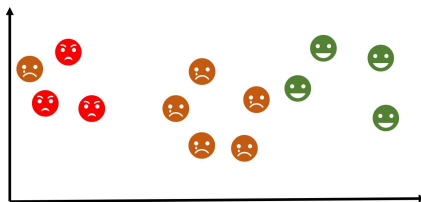
Supervised Learning



[Google Images](#)

**Classification
Regression**

Unsupervised Learning



[Google Images](#)

**Clustering
Association**

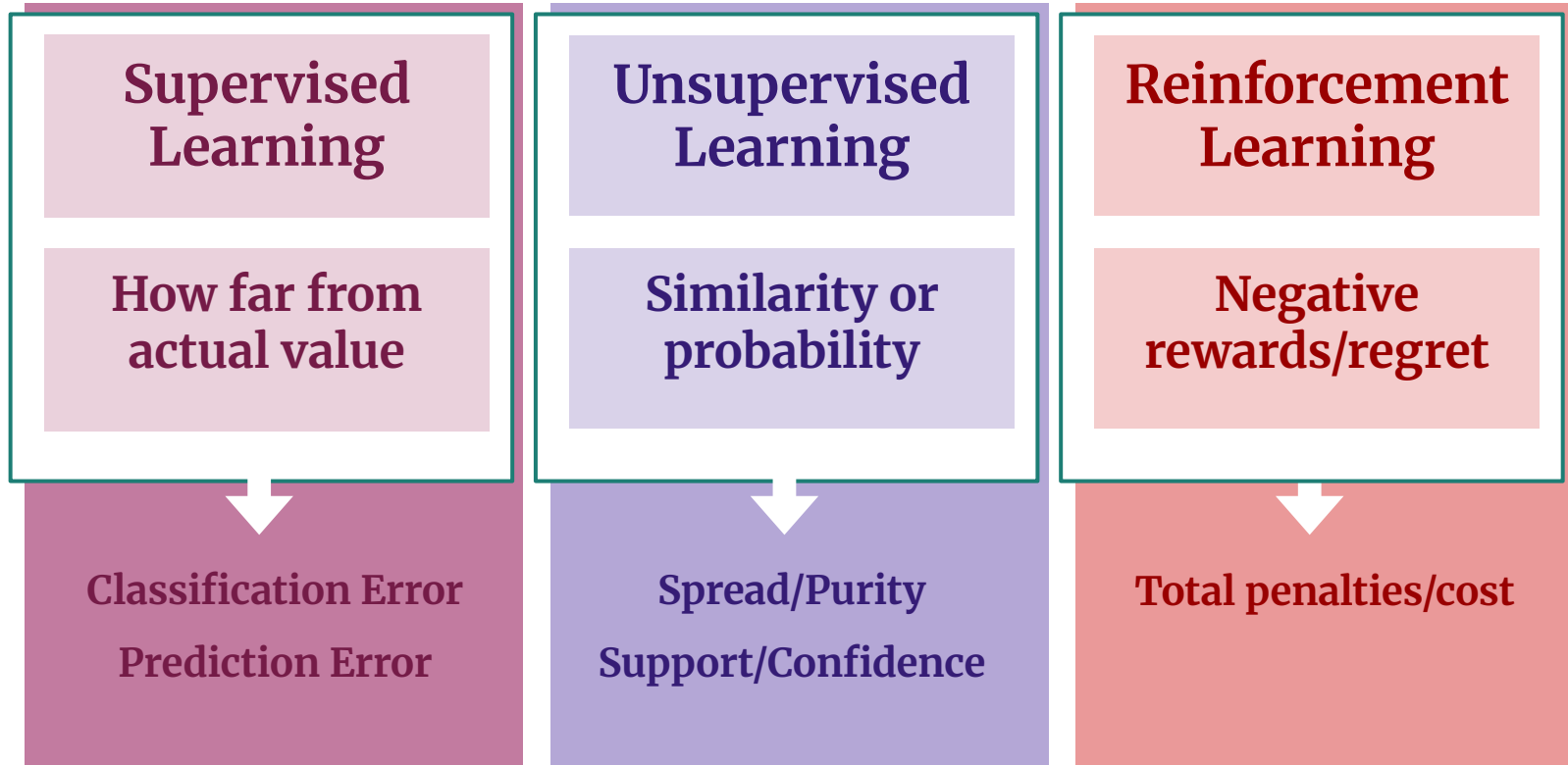
Reinforcement Learning



[Google Images](#)

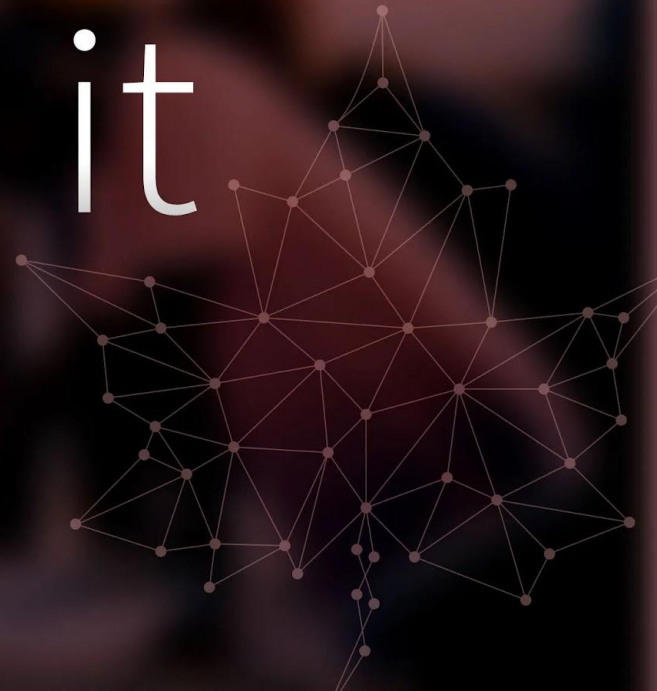
**Rewards
Feedback**

Performance Measures in Machine Learning



Assignment-1 (Cs-46- 2025) (Week-1)

Let's ^{SOLVE} = it



Question-1

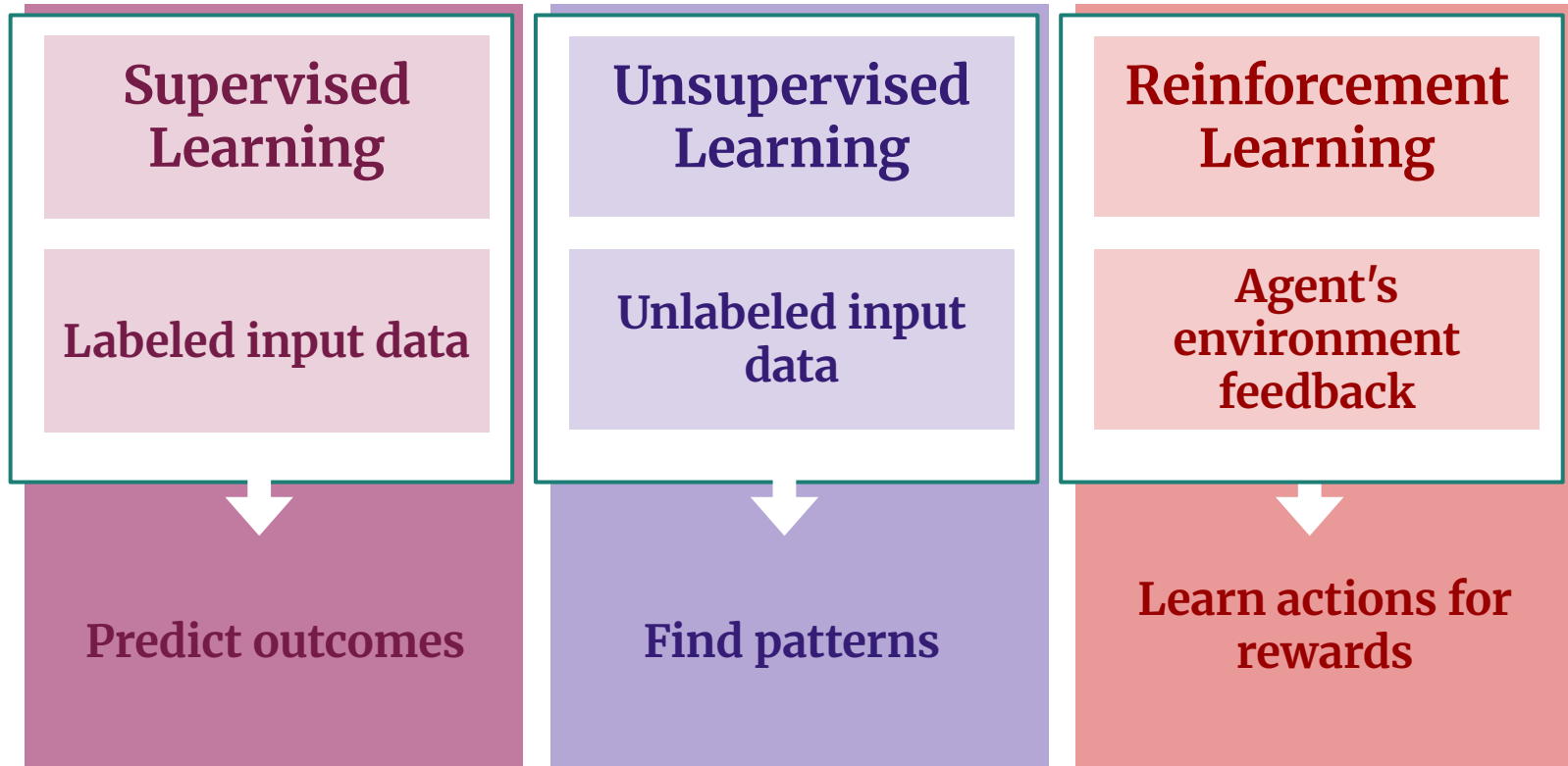
— — —

01:00

Which of the following is/are unsupervised learning problem(s)?

- a) Sorting a set of news articles into four categories based on their titles
- b) Forecasting the stock price of a given company based on historical data
- c) Predicting the type of interaction (positive/negative) between a new drug and a set of human proteins
- d) Identifying close-knit communities of people in a social network
- e) Learning to generate artificial human faces using the faces from a facial recognition dataset

Question-1: Explanation



Question-1- Correct answer

— — —

Which of the following is/are unsupervised learning problem(s)?

- a) Sorting a set of news articles into four categories based on their titles
- b) Forecasting the stock price of a given company based on historical data
- c) Predicting the type of interaction (positive/negative) between a new drug and a set of human proteins
- d) Identifying close-knit communities of people in a social network
- e) Learning to generate artificial human faces using the faces from a facial recognition dataset

Correct options: (d) (e)

Question-2

— — —

01:00

Which of the following statement(s) about Reinforcement Learning (RL) is/are true?

- a) While learning a policy, the goal is to maximize the reward for the current time step
- b) During training, the agent is explicitly provided the most optimal action to be taken in each state.
- c) The actions taken by an agent do not affect the environment in any way.
- d) RL agents used for playing turn based games like chess can be trained by playing the agent against itself (self play).
- e) RL can be used in an autonomous driving system.

Question-2-Explanation

— — —

Which of the following statement(s) about Reinforcement Learning (RL) is/are true?

- a) While learning a policy, the goal is to maximize the reward for the **current time step**
- b) During training, the agent is explicitly provided the most optimal action to be taken in each state. —> **supervision**
- c) The actions taken by an agent do not affect the environment in any way.
- d) RL agents used for playing turn based games like chess can be trained by playing the agent against itself (self play).
- e) RL can be used in an autonomous driving system.

Question-2 - Correct answer

— — —

Which of the following statement(s) about Reinforcement Learning (RL) is/are true?

- a) While learning a policy, the goal is to maximize the reward for the current time step
- b) During training, the agent is explicitly provided the most optimal action to be taken in each state.
- c) The actions taken by an agent do no affect the environment in any way.
- d) RL agents used for playing turn based games like chess can be trained by playing the agent against itself (self play).
- e) RL can be used in a autonomous driving system.

Correct options: (d) (e)

Question-3

01:00

Which of the following is/are regression tasks(s)?

- a) Predicting whether an email is spam or not spam
- b) Predicting the number of new CoVID cases in a given time period
- c) Predicting the total number of goals a given football team scores in an year
- d) Identifying the language used in a given text document

Question-3- Correct answer

— — —

Which of the following is/are regression tasks(s)?

- a) Predicting whether an email is spam or not spam
- b) Predicting the number of new CoVID cases in a given time period
- c) Predicting the total number of goals a given football team scores in an year
- d) Identifying the language used in a given text document

Correct options: (b) (c)

Question-4

— — —

01:00

Which of the following is/are classification task(s)?

- a) Predicting whether or not a customer will repay a loan based on their credit history
- b) Forecasting the weather (temperature, humidity, rainfall etc.) at a given place for the following 24 hours
- c) Predict the price of a house 10 years after it is constructed.
- d) Predict if a house will be standing 50 years after it is constructed.

Question-4- Correct answer

— — —

Which of the following is/are classification task(s)?

- a) Predicting whether or not a customer will repay a loan based on their credit history
- b) Forecasting the weather (temperature, humidity, rainfall etc.) at a given place for the following 24 hours
- c) Predict the price of a house 10 years after it is constructed.
- d) Predict if a house will be standing 50 years after it is constructed.

Correct options: (a) (d)

Question-6

03:00

Given the following dataset. Using a k-nearest neighbour (k-NN) regression model with $k = 3$, predict the value of y at $(x_1, x_2) = (1.0, 0.5)$. Use the Euclidean distance to find the nearest neighbours.

- a) -1.766
- b) -1.166
- c) 1.133
- d) 1.733

X1	X2	Y
1.0	0	2.65
-1.0	0.5	-2.05
2.0	1.0	1.95
-2.0	-1.5	0.90
1.0	1.0	0.60
-1.0	1.0	1.45

Question-6

Given the following dataset. Using a k-nearest neighbour (k-NN) regression model with $k = 3$, predict the value of y at $(x_1, x_2) = (1.0, 0.5)$. Use the Euclidean distance to find the nearest neighbours.

X1	X2	Y	$A=(X_1-1)^2$	$B=(X_2-0.5)^2$	$D= A+B$	Nearest (1-6)
1.0	0	2.65	0	0.25	0.25	1
-1.0	0.5	-2.05	4	0	4	5
2.0	1.0	1.95	1	0.25	1.25	3
-2.0	-1.5	0.90	9	4	13	6
1.0	1.0	0.60	0	0.25	0.25	2
-1.0	1.0	1.45	4	0.25	1.25	4

Question-6- Correct answer

Given the following dataset. Using a k-nearest neighbour (k-NN) regression model with $k = 3$, predict the value of y at $(x_1, x_2) = (1.0, 0.5)$. Use the Euclidean distance to find the nearest neighbours.

- a) -1.766
- b) -1.166
- c) 1.133
- d) 1.733

Correct options: (d)

Question-7

03:00

Consider the following dataset with three classes: 0, 1 and 2. x_1 and x_2 are the independent variables whereas y is the class label. Using a k -NN classifier with $k = 5$, predict the class label at the point $(x_1, x_2) = (1.0, 1.0)$. Use the Euclidean distance to find the nearest neighbours.

- a) 0
- b) 1
- c) 2
- d) Cannot be predicted

X1	X2	Y
-1.0	1.0	0
-1.0	0.0	0
-2.0	-1.0	0
0.0	0.0	1
2.0	1.0	1
1.0	2.0	1
2.0	-1.0	2
2.0	0.0	2

Question-7

Consider the following dataset with three classes: 0, 1 and 2. x_1 and x_2 are the independent variables whereas y is the class label. Using a k -NN classifier with $k = 5$, predict the class label at the point $(x_1, x_2) = (1.0, 1.0)$. Use the Euclidean distance to find the nearest neighbours.

X1	X2	Y	$A=(X_1-1)^2$	$B=(X_2-1)^2$	$D= A+B$	Nearest (1-6)
-1.0	1.0	0	4	0	4	5
-1.0	0.0	0	4	1	5	6
-2.0	-1.0	0	9	4	13	8
0.0	0.0	1	1	1	2	4
2.0	1.0	1	1	0	1	1
1.0	2.0	1	0	1	1	2
2.0	-1.0	2	1	4	5	7
2.0	0.0	2	1	1	2	3

Question-7- Correct answer

Consider the following dataset with three classes: 0, 1 and 2. x_1 and x_2 are the independent variables whereas y is the class label. Using a k -NN classifier with $k = 5$, predict the class label at the point $(x_1, x_2) = (1.0, 1.0)$. Use the Euclidean distance to find the nearest neighbours.

- a) 0
- b) 1
- c) 2
- d) Cannot be predicted

Correct options: (b)

Question-8

— — —

01:00

Consider the following statements regarding linear regression and k-NN regression models. Select the true statements.

- a) A linear regressor requires the training data points during inference.
- b) A k-NN regressor requires the training data points during inference.
- c) A k-NN regressor with a higher value of k is less prone to overfitting.
- d) A linear regressor partitions the input space into multiple regions such that the prediction over a given region is constant.

Question-8 - Correct answer

— — —

Consider the following statements regarding linear regression and k-NN regression models. Select the true statements.

- a) A linear regressor requires the training data points during inference.
- b) A k-NN regressor requires the training data points during inference.
- c) A k-NN regressor with a higher value of k is less prone to overfitting.
- d) A linear regressor partitions the input space into multiple regions such that the prediction over a given region is constant.

Correct options: (b) (c)

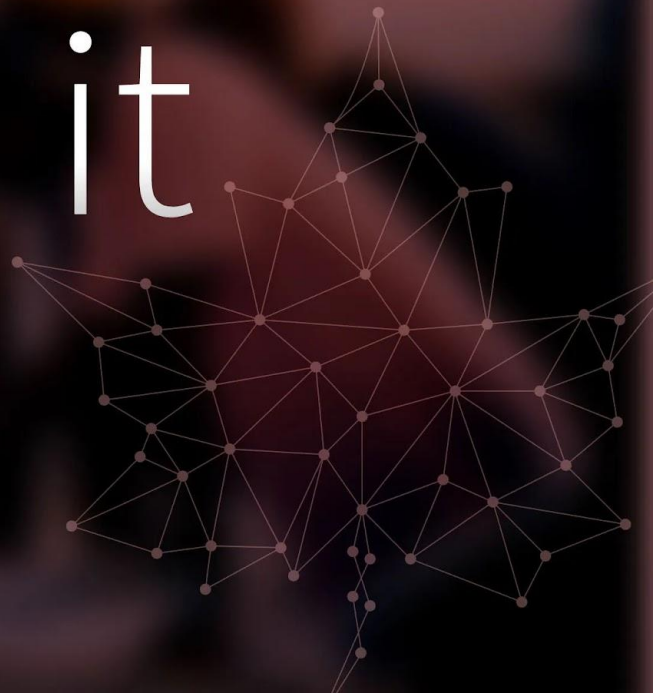
Question-9 : Explanation

— — —

	The Bias-Variance Trade-off
Bias	<p>Measures the error due to the difference between the true function and the expected prediction of the model</p> <p>How far the model from expected prediction</p> $f(x) - E(f^{\wedge}(x))$
Variance	<p>Measures the variability of the model's predictions around its expected value</p> <p>Variance of predicted values</p> $E[(f^{\wedge}(x) - E(f^{\wedge}(x)))^2]$

Assignment-1 (Cs-101- 2024) (Week-1)

Let's ^{SOLVE} = it



Question-1

— — —

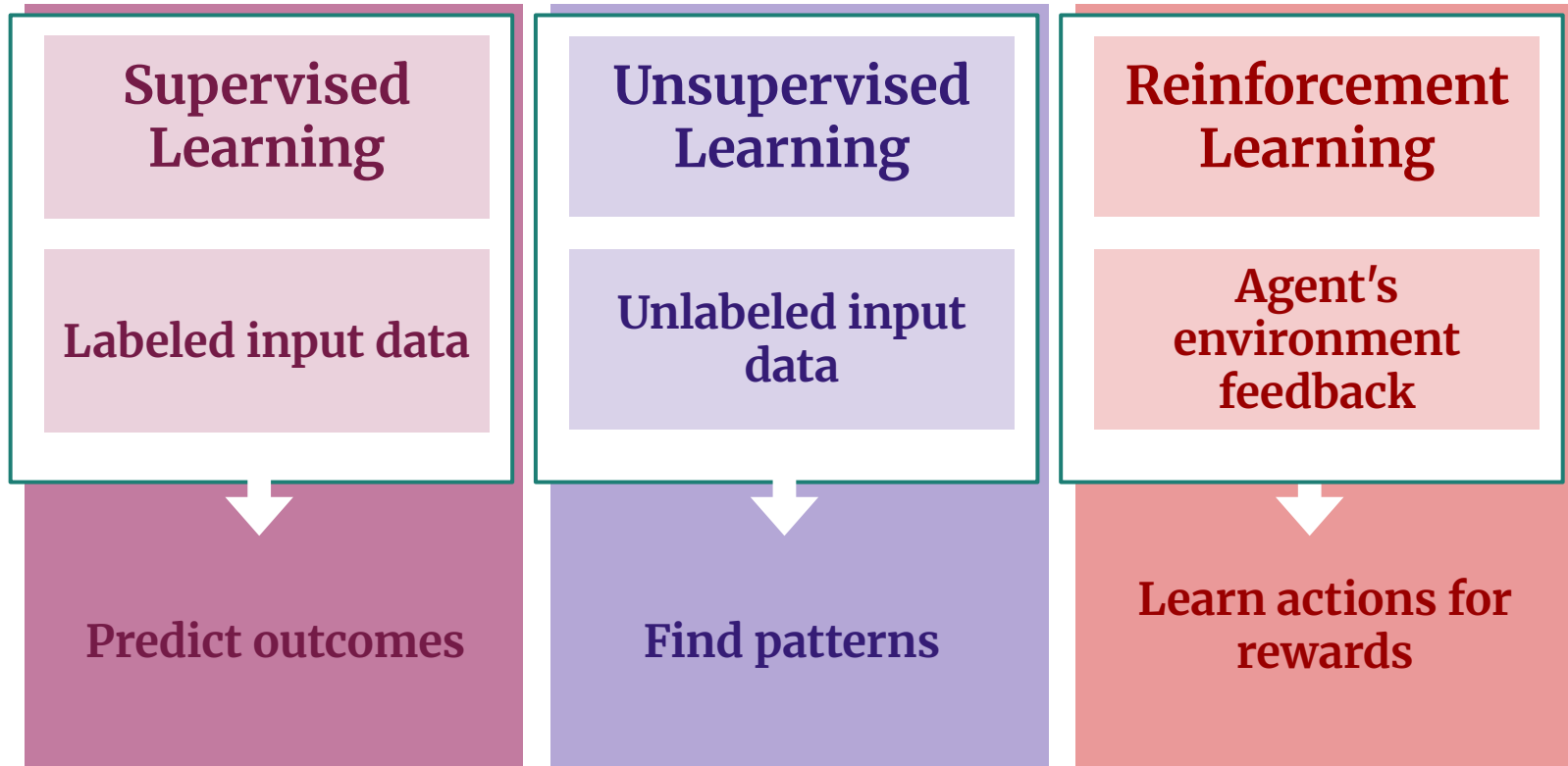
01:00

Which of the following are supervised learning problems
Correct)?

(Multiple

- a) Clustering Spotify users based on their listening history
- b) Weather forecast using data collected by a satellite
- c) Predicting tuberculosis using patient's chest X-Ray
- d) Training a humanoid to walk using a reward system

Question-1: Explanation



Question-1- Correct answer

Which of the following are supervised learning problems (Multiple Correct)?

- a) Clustering Spotify users based on their listening history - Unsupervised
- b) Weather forecast using data collected by a satellite - Supervised
- c) Predicting tuberculosis using patient's chest X-Ray- Supervised
- d) Training a humanoid to walk using a reward system- Reinforcement

Correct options: (b) (c)

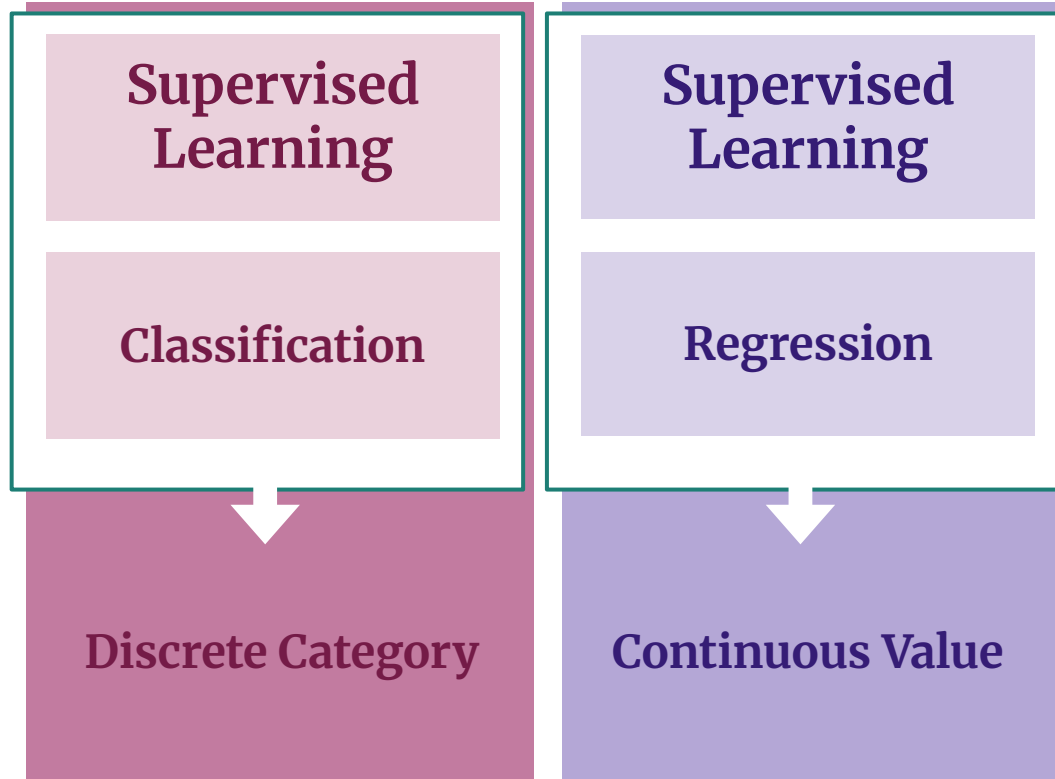
Question-2

01:00

Which of the following are regression tasks (Multiple Correct)?

- a) Predicting the outcome of an election
- b) Predicting the weight of a giraffe based on its weight
- c) Predicting the emotion conveyed by a sentence
- d) Identifying abnormal data points

Question-2: Explanation



Question-2: Explanation

— — —

a)	Predicting the outcome of an election	Supervised—-> whether win or not —-> Classification
b)	Predicting the weight of a giraffe based on its weight	Supervised—-> weight in kgs (unit of weights) —-> Regression
c)	Predicting the emotion conveyed by a sentence	Supervised—-> whether happy, sad, angry, distressed etc—-> Classification
d)	Identifying abnormal data points	Unsupervised—-> make patterns —> if distance > threshold —--> anomaly

Question-2 - Correct answer

Which of the following are regression tasks (Multiple Correct)?

- a) Predicting the outcome of an election - Classification
- b) Predicting the weight of a giraffe based on its weight- Regression
- c) Predicting the emotion conveyed by a sentence- Classification
- d) Identifying abnormal data points- Unsupervised

Correct options: (b)

Question-3

— — —

01:00

Which of the following are classification tasks (Multiple Correct)?

- a) Predicting the outcome of an election
- b) Predicting the weight of a giraffe based on its weight
- c) Predicting the emotion conveyed by a sentence
- d) Identifying abnormal data points

Question-3- Correct answer

— — —

Which of the following are regression tasks (Multiple Correct)?

- a) Predicting the outcome of an election - Classification
- b) Predicting the weight of a giraffe based on its weight- Regression
- c) Predicting the emotion conveyed by a sentence- Classification
- d) Identifying abnormal data points- Unsupervised

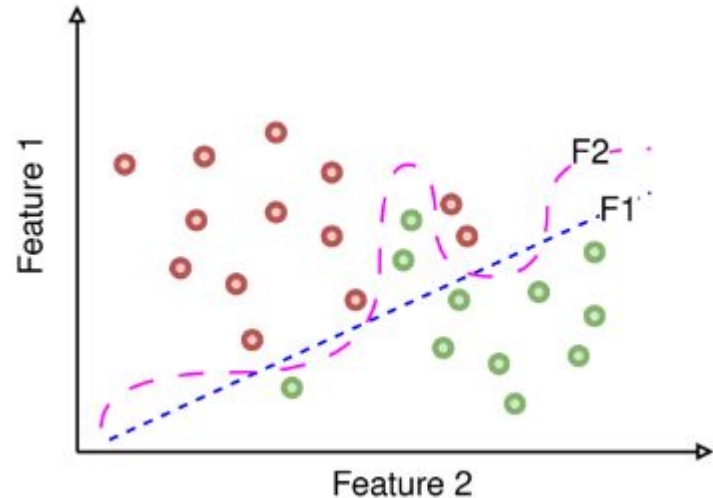
Correct options: (a) (c)

Question-4,5- Common data

Here is a 2-dimensional plot showing two functions that classify data points into two classes.

The red points belong to one class, and the green points belong to another.

The dotted blue line (F1) and dashed pink line (F2) represent the two trained functions.



Question-4

— — —

01:00

Which of the two functions overfit the training data?

- a) Both functions F1 & F2
- b) Function F1
- c) Function F2
- d) None

Question-4 : Explanation

Parameters	Under Fitting	Proper Fitting	Over Fitting
Model Complexity	Too low	Just right	Too high
Training Error	High	Low	Very low
Validation Error	High	Low	High
Generalization	Poor	Good	Poor

Question-4 - Correct answer

Which of the two functions overfit the training data?

- a) Both functions F1 & F2
- b) Function F1 - Not complex
- c) Function F2- Complex (overfit)
- d) None

Correct options: (c)

Question-5

— — —

01:00

Which of the following 2 functions will yield higher training error?

- a) Both functions F1 & F2 will have the same training error
- b) Function F1
- c) Function F2
- d) Can not be determined

Question-5 : Explanation

Parameters	Under Fitting	Proper Fitting	Over Fitting
Model Complexity	Too low	Just right	Too high
Training Error	High	Low	Very low
Validation Error	High	Low	High
Generalization	Poor	Good	Poor

Question-5 - Correct answer

Which of the two functions overfit the training data?

- a) Both functions F1 & F2 will have the same training error
- b) **Function F1**
- c) Function F2 - overfitted (low training error)
- d) Can not be determined

Correct options: (b)

Question-6

— — —

01:00

What does the term 'policy' refer to in reinforcement learning?

- a) A set of rules governing the environment
- b) The reward function
- c) The initial state of the environment
- d) The strategy the agent follows to choose actions

Question-6 : Explanation

Key components	Reinforcement Learning
Agent	Learns and takes actions.
Environment	Provides feedback in the form of rewards.
Reward Function	Guides the agent's learning by defining goals.
Policy	Guides the agent's actions at each state.
Value Function	Estimates the expected reward for a state or state-action pair.

Question-6 - Correct answer

What does the term 'policy' refer to in reinforcement learning?

- a) A set of rules governing the environment
- b) The reward function
- c) The initial state of the environment
- d) The strategy the agent follows to choose actions

Correct options: (d)

Question-7

03:00

Given the following dataset, for $k = 3$, use KNN regression to find the prediction for a new data-point (2,3)

(Use Euclidean distance measure for finding closest points)

- a) 2.0
- b) 2.6
- c) 2.8
- d) 3.2

X1	X2	Y
2	5	3.4
5	5	5
3	3	3
6	3	4.5
2	2	2
4	1	2.8

Question-7

Given the following dataset, for $k = 3$, use KNN regression to find the prediction for a new data-point (2,3)

(Use Euclidean distance measure for finding closest points)

X1	X2	Y	$A=(X1-2)^2$	$B=(X2-3)^2$	$D= A+B$	Nearest (1-6)
2	5	3.4	0	4	4	3
5	5	5	9	4	13	5
3	3	3	1	0	1	1
6	3	4.5	16	0	16	6
2	2	2	0	1	1	2
4	1	2.8	4	4	8	4

Question-7 - Correct answer

— — —

Given the following dataset, for $k = 3$, use KNN regression to find the prediction for a new data-point (2,3)

- a) 2.0
- b) 2.6
- c) 2.8
- d) 3.2

Correct options: (c)

Question-8

— — —

01:00

For any given dataset, comment on the bias of K -nearest classifiers upon increasing the value of K

- a) The bias of the classifier decreases
- b) The bias of the classifier increases
- c) The bias of the classifier does not change
- d) Can not be determined

Question-8 : Explanation

— — —

Bias	The error introduced by approximating a real-world problem with a simplified model.		
Parameters	Low K	Moderate K	High K
Decision boundary	Highly flexible	Balanced	Very Smooth
Fitting	Prone to overfitting	Proper fitting	Prone to underfitting
Bias	Low	Good tradeoff	High

Question-8 - Correct answer

For any given dataset, comment on the bias of K -nearest classifiers upon increasing the value of K

- a) The bias of the classifier decreases
- b) The bias of the classifier increases
- c) The bias of the classifier does not change
- d) Can not be determined

Correct options: (b)

Question-10

— — —

01:00

Which of the following statements are FALSE regarding bias and variance?

- a) Models which overfit have a low bias
- b) Models which overfit have a high bias
- c) Models which underfit have a high variance
- d) Models which underfit have a low variance

Question-10 : Explanation

— — —

	The Bias-Variance Trade-off		
Parameters	Under fitting	Proper fitting	Over fitting
Models	Too Simple	Balanced	Too complex
Bias	High	Good tradeoff	Low
Variance	Low	Good tradeoff	High

Question-10 – Correct answer

— — —

Which of the following statements are FALSE regarding bias and variance?

- a) Models which overfit have a low bias – True (high variance)
- b) Models which overfit have a high bias – False
- c) Models which underfit have a high variance – False
- d) Models which underfit have a low variance – True (high bias)

Correct options: (b) (c)



THANK YOU

Suggestions and Feedback



Next Session:

**Sunday: 3-Aug-2025
3:00 – 5:00 PM**