UNIT–1 : Introduction to Machine Learning (PYQ Solution)

AKTU · Machine Learning Techniques (BCDS062) · Solutions by **Chirag Jain**, Ycotes

SECTION A Answers (≈70 Words Each)

Q1. What is Machine Learning? Define it. (May 2026, Jan 2026)

Machine Learning (ML) is the branch of AI in which a computer program improves its performance on a task by learning from data instead of being explicitly programmed. Tom Mitchell's definition: a program learns from experience E with respect to task T and performance measure P, if its performance at T, measured by P, improves with E. Examples are spam filtering and recommendation systems.

Q2. What is a well-defined learning problem? Give an example. (May 2026, Jan 2026, Jan 2025)

A well-defined learning problem specifies three things: the task T, the performance measure P and the training experience E. Without them the learning goal is vague. Example, handwriting recognition: T = classify handwritten words, P = percentage of words classified correctly, E = a database of labelled handwritten words. Checkers learning and spam detection are other examples.

Q3. How does machine learning differ from data science, and how do they complement each other? (Jan 2025)

Data science is the broad field of extracting insights from data using statistics, visualisation, data engineering and ML. Machine learning is one tool inside it that builds predictive models which learn from data. Data science prepares, cleans and interprets data; ML builds the models on it. Together they turn raw data into automated predictions and decisions.

Q4. Discuss the important objectives of Machine Learning. (Jan 2024)

The main objectives of ML are:

- **To let machines learn from data and improve with experience without explicit programming.**
- **To generalise to unseen data, not just memorise the training set.**
- **To predict outcomes (classification, regression) and discover patterns (clustering).**
- **To automate decision making and adapt to changing environments.**
- **To handle large, complex data efficiently and with high accuracy.**

Q5. Compare purely analytical learning and purely inductive learning. (Jan 2024)

Inductive learning generalises from many training examples using statistics (e.g. decision trees, neural networks); it needs a lot of data and has no prior theory. Analytical learning uses prior domain knowledge to explain a few examples and generalise from the explanation (e.g. explanation-based learning). Inductive learning can find new patterns but may be wrong with little data; analytical learning needs few examples but depends on correct domain theory.

SECTION B Answers (≈250 Words Each)

Q1. Explain the types of learning in Machine Learning (supervised, unsupervised, reinforcement) with examples. Compare supervised and unsupervised learning. (May 2026, Jan 2026, Jan 2024)

Machine learning is classified by the kind of feedback available during training.

1. Supervised learning: the training data has **labelled** input-output pairs (x, y). The model learns a mapping f: x → y and predicts labels for new inputs.

- Classification: spam or not spam, disease diagnosis.
- Regression: house price, temperature prediction.
- Algorithms: linear/logistic regression, decision tree, SVM, k-NN.

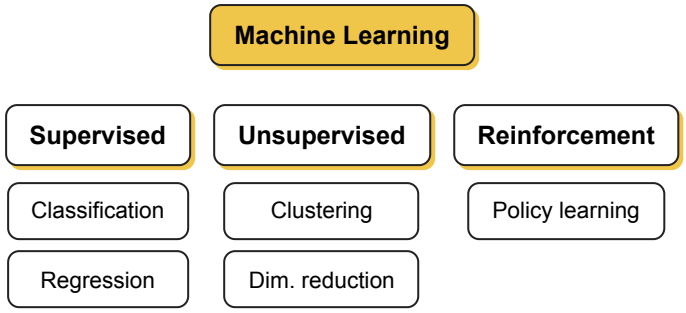
2. Unsupervised learning: the data is **unlabelled**; the model finds hidden structure.

- Clustering: customer segmentation (k-means).
- Dimensionality reduction (PCA), association rules (market basket).

3. Reinforcement learning: an **agent** interacts with an environment, takes actions and receives **rewards or penalties**. It learns a policy that maximises total reward. Examples: game playing (chess, Go), robot navigation, self-driving.

(Semi-supervised learning uses a few labelled and many unlabelled examples.)

Basis	Supervised	Unsupervised
Data	Labelled	Unlabelled
Goal	Predict output	Discover structure
Feedback	Direct (error vs label)	None
Tasks	Classification, regression	Clustering, association
Examples	Spam filter, price prediction	Customer segmentation
Evaluation	Accuracy, RMSE	Harder; silhouette score



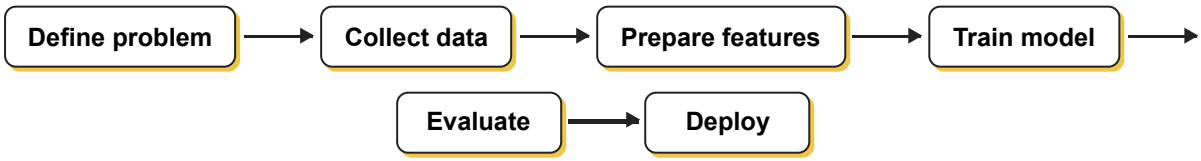
Types of machine learning

Conclusion: the choice depends on whether labels exist and whether learning is by examples or by interaction.

Q2. Describe the steps involved in designing a machine learning system. Illustrate with a real-world example. (Jan 2025)

Designing an ML system is a sequence of steps; the running example is a **spam e-mail filter**.

1. **Define the problem and success measure:** classify mails as spam/not spam; measure = accuracy and false-positive rate.
2. **Collect data:** gather thousands of labelled mails.
3. **Prepare data:** clean, remove duplicates, handle missing values, convert text to features (word counts, TF-IDF), split into train/validation/test (e.g. 70/15/15).
4. **Choose the model:** e.g. Naive Bayes or logistic regression.
5. **Train the model:** fit parameters by minimising a loss function.
6. **Evaluate:** accuracy, precision, recall, F1 on the test set; check for over/underfitting.
7. **Tune hyperparameters:** cross-validation, regularisation.
8. **Deploy and monitor:** integrate into the mail server, monitor performance, retrain when spammers change tactics.



ML system design steps

Conclusion: the process is iterative; poor evaluation sends us back to data or model choice.

SECTION C Answers (≈500 Words Each)

Q1. Differentiate between Data Science and Machine Learning in terms of scope, goals and applications. (May 2026)

Data Science is an interdisciplinary field that extracts knowledge and insight from structured and unstructured data. **Machine Learning** is a subset of AI that builds algorithms learning patterns from data to predict or decide.

Basis	Data Science	Machine Learning
Scope	Entire data lifecycle: collection, cleaning, analysis, visualisation, modelling, communication	Model building and training stage only
Goal	Gain insight, support business decisions	Build accurate predictive/automated models
Skills	Statistics, programming, domain knowledge, visualisation	Algorithms, optimisation, maths
Data	Structured and unstructured	Mostly prepared, feature-engineered data
Output	Reports, dashboards, insights	Trained model, predictions
Tools	SQL, Excel, Tableau, Python	scikit-learn, TensorFlow, PyTorch
Applications	Business analytics, fraud trends, healthcare analysis	Image recognition, spam filter, recommendations, speech

Relationship: ML is a tool used within data science. A data scientist may use ML for the predictive part of a project, but data science also covers non-ML work such as reporting. ML can work on any ready dataset without the wider data-science pipeline.

Example: an e-commerce company uses data science to analyse sales trends and customer behaviour, and uses ML to build the recommendation engine.

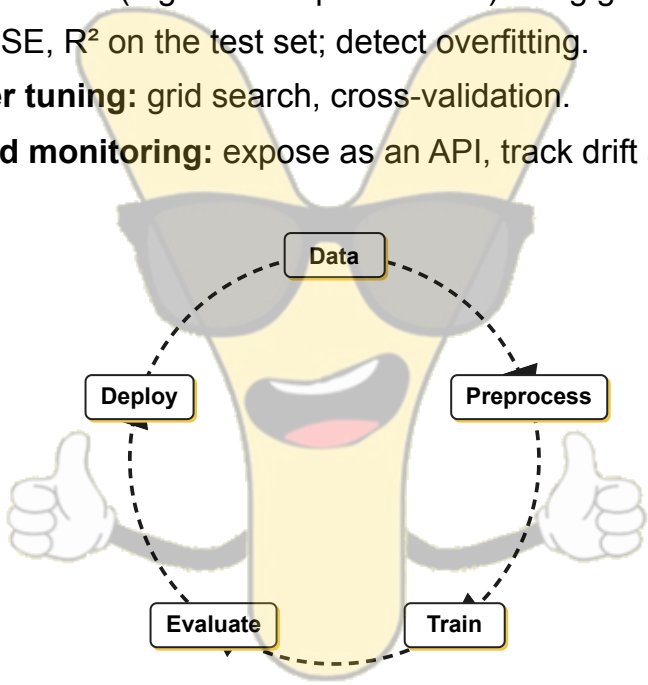
Conclusion: data science is the broader discipline focused on insight; ML is the technique focused on learning and prediction.

Q2. Describe the design process of a Machine Learning system with

suitable steps and examples. (May 2026)

The design of an ML system follows an iterative life cycle. Example used: **house price prediction**.

1. **Problem definition:** decide the task (regression) and metric (RMSE).
2. **Data collection:** area, location, bedrooms and sale price from listings.
3. **Data preprocessing:** remove duplicates, fill missing values, encode categorical features, scale numeric features.
4. **Feature engineering and selection:** create price-per-sqft, drop irrelevant features.
5. **Data split:** training, validation, test sets.
6. **Model selection:** linear regression, decision tree, random forest.
7. **Training:** minimise loss (e.g. mean squared error) using gradient descent.
8. **Evaluation:** RMSE, R^2 on the test set; detect overfitting.
9. **Hyperparameter tuning:** grid search, cross-validation.
10. **Deployment and monitoring:** expose as an API, track drift and retrain.



ML design life cycle

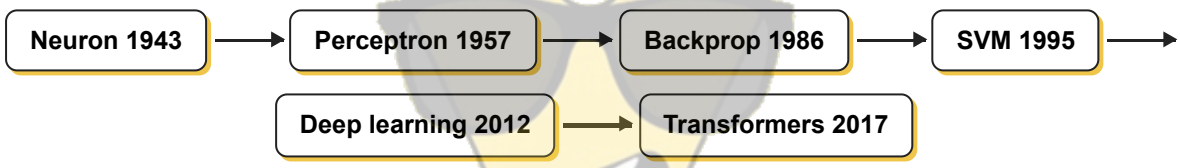
Key design choices (Mitchell): choose the training experience, the target function, its representation and a learning algorithm to approximate it.

Conclusion: careful design at every step determines the quality of the final system.

Q3. Discuss the history of machine learning. Highlight major milestones. (Jan 2025)

Machine learning developed over several decades.

- **1943:** McCulloch and Pitts proposed the first mathematical neuron model.
- **1950:** Alan Turing proposed the Turing Test ("Computing Machinery and Intelligence").
- **1952:** Arthur Samuel wrote a self-improving checkers program; he coined "machine learning" in 1959.
- **1957-58:** Rosenblatt invented the **Perceptron**.
- **1969:** Minsky and Papert showed perceptron limits (XOR), starting the first "AI winter".
- **1970s-80s:** Nearest neighbour, decision trees (ID3, 1986), and **backpropagation** (Rumelhart, Hinton, Williams, 1986) revived neural networks.
- **1990s:** Support Vector Machines (Cortes and Vapnik, 1995), ensemble methods, statistical learning; IBM Deep Blue beat Kasparov (1997).
- **2006:** Hinton introduced deep belief networks, starting **deep learning**.
- **2012:** AlexNet won ImageNet, showing the power of GPUs plus big data.
- **2014-17:** GANs, ResNet, AlphaGo (2016), and the Transformer (2017).
- **2018 onwards:** BERT, GPT and large language models, generative AI.



Milestones in ML

Conclusion: growth in data, computing power and algorithms moved ML from simple rules to today's deep and generative models.

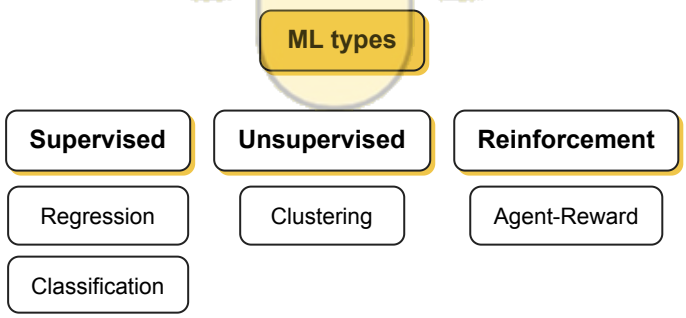
Q4. Compare and contrast supervised, unsupervised and reinforcement learning with examples. Also compare regression, classification and clustering with applications. (Jan 2025, Jan 2024)

Supervised learning learns from labelled data; **unsupervised** finds structure in unlabelled data; **reinforcement** learns through reward from interaction.

Basis	Supervised	Unsupervised	Reinforcement
Data	Labelled (x, y)	Unlabelled	No fixed dataset; experience from environment
Feedback	Correct answer given	None	Reward/penalty, often delayed
Goal	Predict outputs	Find clusters/patterns	Maximise cumulative reward
Algorithms	Regression, SVM, trees	k-means, PCA	Q-learning, policy gradient
Example	Spam detection	Customer segmentation	Game playing, robotics

Regression, classification, clustering:

Basis	Regression	Classification	Clustering
Type	Supervised	Supervised	Unsupervised
Output	Continuous value	Discrete class label	Group (cluster) of similar items
Algorithms	Linear regression	Logistic regression, SVM, k-NN	k-means, hierarchical
Application	House price, sales forecast	Spam, disease diagnosis	Market segmentation, image grouping



Learning paradigms

Conclusion: use supervised learning when labels exist, unsupervised for exploration, and reinforcement for sequential decision making.
