Machine Learning Important Questions PDF

The most important Machine Learning questions are the types of learning, the bias-variance trade-off, overfitting and regularisation, linear and logistic regression with gradient descent, decision trees using entropy and information gain, Naive Bayes numericals, k-nearest neighbours, support vector machines and kernels, k-means and hierarchical clustering, PCA, the confusion matrix with precision, recall and F1 score, cross-validation, and bagging, boosting and random forests.

Key takeaways

- Decision tree (information gain), Naive Bayes and k-means numericals are the most common long questions.
- Learn the formulas for precision, recall, F1 score, entropy and Gini index.
- Explain every algorithm with its idea, steps, a small example, and its pros and cons.
- These are practice questions, not past papers; check your own syllabus and official previous papers.

About these important questions

This page lists the most important Machine Learning questions for B.Tech CSE and IT exams, arranged unit by unit. The units follow the standard syllabus that most Indian technical universities share, so the order may differ slightly from yours.

Please note: these are practice questions based on common syllabus topics. They are not verified past papers, and we do not claim that any question will appear in your exam. Check your own university syllabus first, then solve the official previous year papers. AKTU students can download them from the AKTU old question paper archive.

Unit 1: Introduction to Machine Learning

Covers definitions, types of learning, the ML workflow and core concepts.

Short questions (2 marks)

- Define machine learning (Tom Mitchell's definition).
- Differentiate between supervised, unsupervised and reinforcement learning.
- What is overfitting?
- What is the curse of dimensionality?

Long questions

1. Explain the steps in designing a learning system, using a checkers or spam-filter example.
2. Explain the bias-variance trade-off with a diagram of underfitting and overfitting.
3. Explain the machine learning workflow: data collection, preprocessing, feature engineering, training, evaluation and deployment.
4. Explain the applications of machine learning and the main challenges.
5. Explain hypothesis space, inductive bias and the concept learning task (Find-S and candidate elimination).

Unit 2: Regression and Model Evaluation

Covers linear and logistic regression, gradient descent, regularisation and evaluation metrics.

Short questions (2 marks)

- Differentiate between regression and classification.
- What is a cost function?
- Define precision and recall.

Long questions

1. Explain simple linear regression, and fit a line to a small data set using the least squares method.
2. Explain gradient descent and the effect of the learning rate.
3. Explain logistic regression and the sigmoid function.
4. Explain L1 (lasso) and L2 (ridge) regularisation.
5. From a given confusion matrix, calculate accuracy, precision, recall, F1 score and specificity.
6. Explain k-fold cross-validation and the ROC curve.

Unit 3: Classification Algorithms

Covers decision trees, Naive Bayes, k-NN and SVM. This is the most numerical unit.

Short questions (2 marks)

- Define entropy and information gain.
- What is the "naive" assumption in Naive Bayes?
- What is a support vector?

Long questions

1. Build a decision tree for the "play tennis" data set using ID3, showing entropy and information gain for each attribute.
2. Explain the CART algorithm and the Gini index, and the need for pruning.
3. Classify a new record using Naive Bayes for a given training table.
4. Explain the k-nearest neighbour algorithm and classify a point for k = 3 with Euclidean distance.
5. Explain SVM, the maximum margin hyperplane and the kernel trick.

Unit 4: Unsupervised Learning

Covers clustering, association rules and dimensionality reduction.

Short questions (2 marks)

- Differentiate between classification and clustering.
- How do you choose k in k-means?
- What is PCA used for?

Long questions

1. Apply the k-means algorithm to a small set of points for two iterations.
2. Explain hierarchical clustering (agglomerative) and draw a dendrogram using single or complete linkage.
3. Explain DBSCAN and its advantages over k-means.
4. Explain principal component analysis step by step, using covariance and eigenvectors.
5. Explain association rule mining with support, confidence and the Apriori algorithm.

Unit 5: Ensembles, Neural Networks and Reinforcement Learning

Covers ensemble methods, an introduction to neural networks and reinforcement learning. The exact topics vary by university.

Short questions (2 marks)

- Differentiate between bagging and boosting.
- What is an activation function?
- Define reward and policy in reinforcement learning.

Long questions

1. Explain bagging and random forests.
2. Explain boosting with AdaBoost and gradient boosting.
3. Explain the perceptron learning algorithm and its limitation with XOR.
4. Explain a multilayer perceptron and backpropagation in brief.
5. Explain reinforcement learning, the Markov decision process and Q-learning.

Sample answer outline: Decision tree with information gain (ID3)

For a training table such as "play tennis", answer like this:

1. **Write the formulas.** Entropy(S) = -p log2 p - n log2 n, where p and n are the fractions of positive and negative examples. Gain(S, A) = Entropy(S) - sum over values v of (|Sv| / |S|) × Entropy(Sv).
2. **Find the entropy of the whole set.** For 9 yes and 5 no, Entropy(S) = -(9/14) log2(9/14) - (5/14) log2(5/14), which is about 0.940.
3. **Compute gain for every attribute.** Make a small table for each attribute (Outlook, Temperature, Humidity, Wind) with the yes and no counts per value, then compute the gain.
4. **Pick the root.** Choose the attribute with the highest gain. In the standard play tennis data this is Outlook.

5. **Repeat for each branch.** A branch where all examples have the same label becomes a leaf. Otherwise, repeat the gain calculation on that subset with the remaining attributes.
6. **Draw the final tree** and write one or two rules from it, such as "if Outlook = Overcast then Play = Yes".

Mention that ID3 can overfit and that pruning or a minimum leaf size helps.

Marks weightage and how to write answers

A common pattern, used by AKTU and many other universities, has a compulsory short-answer section that covers every unit, followed by long-answer questions with a choice inside each unit. Marks per question change between universities and years, so confirm the scheme on a recent official paper.

- **Short answers (about 2 marks):** a crisp definition, one example or diagram, and nothing more. Three to five lines is enough.
- **Long answers (7 to 10 marks):** start with a definition, add a labelled diagram or table, explain the steps, and finish with an example, advantages or a comparison.
- Write every formula before you substitute numbers. Formula marks are given even if arithmetic slips.
- For algorithm questions, give the steps as a numbered list, then a small worked example.

A 7-day revision plan for Machine Learning

1. **Days 1 to 5:** one unit a day. Read your notes, then write the long answers for that unit from memory and check them against your notes.
2. **Day 6:** redo one numerical each of linear regression, confusion matrix metrics, ID3, Naive Bayes, k-NN and k-means.
3. **Day 7:** solve one full official previous year paper in three hours, then check which units cost you time.

Keep a single sheet of formulas, definitions and diagrams for the last hour before the exam.

Frequently asked questions

Which Machine Learning topics are most important?

Types of learning, bias and variance, linear and logistic regression, decision trees with information gain, Naive Bayes, k-NN, SVM, k-means, PCA, evaluation metrics and ensemble methods are asked most often.

Do I need to write Python code in a Machine Learning exam?

Usually not. Most papers ask for algorithms, derivations and numericals. Some practical exams ask for scikit-learn code, so check your scheme.

How is this subject useful after B.Tech?

It is the base for data science, data analyst and ML engineer roles. Add Python, statistics and project work to make the theory useful in practice.

Are these Machine Learning questions from previous year papers?

No. They are practice questions built from topics that appear in most B.Tech Machine Learning syllabi. Use them with your own university's official previous papers, for example the AKTU archive at https://aktu.ac.in/old-question-paper.html.

Is this PDF free?

Yes. Fill in the short form on this page and the PDF downloads right away.

Sources and further reading

- AKTU: old question papers (official archive)
- scikit-learn: User Guide
- NPTEL: online courses from IITs (search for Machine Learning)

About this PDF

These are practice questions for revision, prepared by the Uncodemy trainer team from common B.Tech syllabus topics. They are not past papers, a certificate or a promise of marks. Spot an error? Write to info@uncodemy.com and the team will review it.

More free resources and courses: uncodemy.com | info@uncodemy.com | +91 98183 66550