🎯 Top 30 Must-Study Machine Learning Questions (University Exam Focus)

📘 Core Theory (10 Questions)

1. **Define Machine Learning.** Differentiate between Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning.
2. **Explain the assumptions of Linear Regression** and how to detect and address violations (Linearity, Homoscedasticity, Multicollinearity).
3. **Compare Batch Gradient Descent, Stochastic Gradient Descent, and Mini-Batch Gradient Descent** — with pros and cons.
4. **What is the bias-variance tradeoff?** Explain with a diagram and how regularization helps control it.
5. **Describe the ML life cycle** from data collection to deployment.
6. **Explain Ridge vs. Lasso vs. ElasticNet** regularization — when to use which.
7. **What is PCA and why is it used?** Explain its mathematical and geometric intuition.
8. **Discuss the concept of ROC-AUC, precision, recall, and F1-score.** How do they differ in imbalanced datasets?
9. **Explain the difference between hard-margin and soft-margin SVM.** When is each preferred?
10. **Compare Bagging and Boosting.** How do they reduce variance and bias respectively?

🧮 Important Derivations (10 Questions)

11. **Derive the Normal Equation for Linear Regression** from the least squares cost function.
12. **Prove that $R^2 = 1 - \frac{RSS}{TSS}$** and explain its meaning in regression analysis.
13. **Derive the gradient descent update rule** for a linear regression cost function.
14. **Derive the F-statistic** for testing overall model significance using ESS, TSS, and RSS.
15. **Derive the log-likelihood function in Logistic Regression** and how it's optimized using gradient descent.
16. **Derive the Ridge Regression cost function** and the modified normal equation.
17. **Derive PCA using variance maximization** and show the role of eigenvalues and eigenvectors.
18. **Show how SVM optimization problem is framed** and explain the concept of support vectors.
19. **Derive function update step for Gradient Boosting** using residual errors.
20. **Explain Taylor expansion and how it is used in XGBoost** to compute the objective function.

🔢 Numerical Problems You Should Practice (10 Questions)

21. **Given a small dataset, manually compute slope and intercept** for simple linear regression.
22. **Perform 2–3 iterations of Gradient Descent** by hand for a given θ and α.
23. **Calculate MAE, MSE, RMSE, and R^2** for predicted vs actual values.
24. **Calculate ANOVA F-value for feature selection.**
25. **Given a confusion matrix, compute accuracy, precision, recall, and F1-score.**
26. **Classify a new point using KNN** with $K = 3$ and given neighbors.
27. **Perform 2 iterations of KMeans** clustering on a small 2D dataset.
28. **Calculate eigenvalues and eigenvectors** of a covariance matrix and project data using PCA.
29. **Build a decision tree of depth 2 manually** using Gini Index or Information Gain.

30. **Compute similarity score and gain in XGBoost** for a single split using provided gradients and hessians.

📝 Tip for Revision:

- ⭐ Theory: Write 3–5 bullet point summaries for each theory question.
- 📐 Derivations: Practice by deriving **from scratch**, not memorizing.
- 🧮 Numerical: Use at least one **worked-out example** for each type.