1. Given a symmetric matrix $A \in \mathbb{R}^{n \times n}$, show that all eigenvalues of A are real. How does this property affect PCA?
2. If $X \in \mathbb{R}^{m \times n}$ is a data matrix, explain why the covariance matrix $C = \frac{1}{m} X^T X$ is always positive semi-definite.
3. Prove that the determinant of a matrix is equal to the product of its eigenvalues. How does this relate to the condition number of the Hessian in optimization problems?
4. Compute the gradient of the following with respect to w:

$$L(w) = \frac{1}{2} \|Xw - y\|^2$$

   and explain how this is used in linear regression.
5. Let $f(x) = \sigma(w^T x)$ where $\sigma(z) = \frac{1}{1+e^{-z}}$. Derive both the gradient and Hessian of $f(x)$.

6. Suppose $X \sim \mathcal{N}(0, 1)$. Derive the probability that $|X| > 2$ using the cumulative distribution function (CDF).
7. You flip a biased coin with probability p of heads, n times. Derive the **expected number of heads** and the **variance**.
8. Let $X_1, X_2, \ldots, X_n$ be i.i.d samples from a distribution with mean μ and variance σ^2. Show that the sample mean is an **unbiased estimator** of μ.
9. A Gaussian Mixture Model (GMM) has two components $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ with mixing weights π_1, π_2. Write the log-likelihood function for data X, and explain how the EM algorithm maximizes it.
10. Prove that **KL Divergence** is always non-negative:

$$D_{KL}(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \geq 0$$

11. Derive the update rule for **gradient descent** in logistic regression starting from the likelihood function.
12. Show how **L1 regularization** leads to sparse solutions (feature selection), while **L2 regularization** leads to shrinkage but not sparsity. Use convex optimization arguments.
13. In Support Vector Machines (SVM), derive the dual optimization problem starting from the primal:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{s.t. } y_i(w^T x_i + b) \geq 1$$

14. Derive the **backpropagation rule** for a single hidden layer neural network with ReLU activation.
15. Consider the bias–variance decomposition:

$$\mathbb{E}[(y - \hat{f}(x))^2] = \text{Bias}^2 + \text{Variance} + \sigma^2$$

Derive this equation step by step.