Brief Reasons for Each Question

Each correct option is equally weighted. For example, if 2 out of 3 correct options were marked, +2 marks will be awarded for that question.

1. **Question 1:** Correct options: (a), (d)
   (a) *Both X^TX and XX^T have the same eigenvalues.*
   This follows from the fact that X^TX and XX^T share the same non-zero eigenvalues (they have the same singular values).
   (d) *X contains the eigenvalues of X^TX on its diagonal.*
   In the SVD $X = UDV^T$, the diagonal entries of D are the singular values, which are the square roots of the eigenvalues of X^TX.

2. **Question 2:** Correct options: (a), (d) *or* (a), (c), (d)
   (a) Often states that the rank of X is related to the non-zero singular values.
   (c) If included, typically addresses that X has a particular form/behavior (e.g., full column rank) that connects B or D to rank properties.
   (d) Can involve diagonalization arguments or how the nullity and rank interplay with X and its factors.

   The ambiguity in (d) due to —A— not being defined anywhere is handled by awarding (a),(d) or (a),(c),(d)

3. **Question 3:** Correct options: (a), (b), (c), (d)

(a)–(d) This question concerns a set of 2D points where all points share the same second coordinate $x_2 = 5$, leading to dependencies among the coordinates.

- The covariance matrix is computed as:

$$\Sigma = \begin{bmatrix} \mathrm{Var}(X_1) & 0 \\ 0 & 0 \end{bmatrix}$$

  where $\mathrm{Var}(X_1)$ is the variance of the first coordinate values $\{1, 2, 3, 4, 5\}$.

- The variance of X_1 is given by:

$$\mathrm{Var}(X_1) = \frac{1}{5}\sum_{i=1}^{5}(x_i - \bar{x})^2 = 2.$$

- Thus, the covariance matrix simplifies to:

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

- The eigenvalues of Σ are the diagonal entries:

$$\lambda_1 = 2, \quad \lambda_2 = 0.$$

- Since $\lambda_1 > \lambda_2$, statement (a) holds.

4. **Question 4:** Correct options: (a), (c), (d)
   (a) Increasing training data helps reduce overfitting.
   (c) Reducing model complexity (e.g. fewer parameters/features) also combats overfitting.
   (d) Proper regularization can reduce overfitting by penalizing large weights or complex models.

5. **Question 5:** Correct options: (a), (b)
   - If A is an $n \times n$ matrix with orthonormal columns, then $A^T A = I$ and $\|a_i\| = 1$.

   (a) Typically $A^T = A^{-1}$ for an orthonormal (orthogonal) matrix.
   (b) Each column is unit norm and orthogonal to the others, matching the given conditions.

6. **Question 6:** Correct options: (a), (c), (d)
   - Overfitting can be reduced by:
     (a) Getting more data,
     (c) Increasing regularization or decreasing model complexity,
     (d) Early stopping or other gradient-based methods that prevent overfitting.

7. **Question 7:** Correct options: (a), (d), (g), (h)
   - As per definitions

8. **Question 8:** Correct option: (a)
   - If the set of points $\{x_i\}$ lies on a single line, then the rank of the sum $\sum x_i x_i^T$ is 1. Hence only (a) is correct.

9. **Question 9:** Correct options: (a), (c), (d)
   (a) Gradient descent requires the loss to be differentiable for a direct gradient-based update.
   (c) It is used for optimizing continuous parameters in many ML problems.
   (d) It can be extended for non-convex problems, though convergence is not guaranteed to the global optimum.

10. **Question 10:** Correct options: (a), (d)
    (a) k-Nearest Neighbors is indeed a "lazy" learning algorithm.
    (d) L_2 regularization is a common choice (e.g. ridge regression) that is popular because it is differentiable and helps control model complexity.

Answer to Q2.3: Relationship between the eigenvalues of $\mathbf{A}^T\mathbf{A}$ and the singular values of $\mathbf{A}$

Given: The Singular Value Decomposition (SVD) of a matrix $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T,$$

where:

- $\mathbf{U}$ is an orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}$).
- $\mathbf{V}$ is an orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{I}$).
- $\mathbf{D}$ is a diagonal matrix whose diagonal entries are the singular values $\sigma_i \geq 0$.

Derivation:

$$\mathbf{A}^T\mathbf{A} \;=\; \left(\mathbf{U}\mathbf{D}\mathbf{V}^T\right)^T\left(\mathbf{U}\mathbf{D}\mathbf{V}^T\right) \;=\; \mathbf{V}\mathbf{D}^T\mathbf{U}^T\,\mathbf{U}\,\mathbf{D}\mathbf{V}^T \;=\; \mathbf{V}\mathbf{D}^T\mathbf{D}\mathbf{V}^T \;=\; \mathbf{V}\mathbf{D}^2\mathbf{V}^T,$$

since $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{D}^T = \mathbf{D}$ (diagonal).

- The columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{A}^T\mathbf{A}$.
- The diagonal entries of $\mathbf{D}^2$ (i.e. σ_i^2) are the corresponding eigenvalues of $\mathbf{A}^T\mathbf{A}$.

Hence, if σ_i are the singular values of $\mathbf{A}$, and λ_i are the eigenvalues of $\mathbf{A}^T\mathbf{A}$, then:

$$\lambda_i \;=\; \sigma_i^2.$$

Marking Scheme (6 marks)

1. Writing $\mathbf{A}^T\mathbf{A} = (\mathbf{U}\mathbf{D}\mathbf{V}^T)^T(\mathbf{U}\mathbf{D}\mathbf{V}^T)$ correctly ....1 mark
2. Simplifying to $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{D}^T\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T$ ....1 mark
3. Recognizing $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ due to orthogonality ....1 mark
4. Further simplifying to $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{D}^T\mathbf{D}\mathbf{V}^T$ ....1 mark
5. Recognizing $\mathbf{D}^T = \mathbf{D}$ (since $\mathbf{D}$ is diagonal) and getting $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{D}^2\mathbf{V}^T$ .....1 mark
6. Stating the final relationship: $\lambda_i = \sigma_i^2$ (the eigenvalues of $\mathbf{A}^T\mathbf{A}$ are the squares of the singular values of $\mathbf{A}$) ....1 mark

Note: Partial marks can be awarded for correct steps with minor errors or alternative valid derivations.