1 Basics

1.1 General questions

1. Q1: What is the difference between Classification and Regression?

   - classification: A function $f : \mathcal{R} \rightarrow 1, ...k$ that maps a sample to K categories, represented by a set of n-dimensional features, to a categorical output. The numeric code $y = f(x)$ is the category prediction for the input x. The output can be described on different ways, e.g. probabilistic.
   - regression: A function $f : \mathcal{R} \rightarrow \mathcal{R}$ that predicts a numerical value, the output is typically continuous.

2. Q2: What metrics are commonly used in classification and regression tasks?

   - classification: Accuracy, precision, recall, sensitivity, specificity, F1- score.
   - regression: Mean absolute error (MAE), root mean square Error(RMSE), MSE, p-norm

3. Q3: Consider a Machine Learning Model for the automatic detection of breast tumors. Will such a model rather optimize for Recall or Precision?

   - The consequences of missing a cancerous tumor are considered more severe than the consequences of performing unnecessary follow-up tests for false-positive cases.
   - If the model focuses on optimizing for recall, it will be designed to minimize false negatives, ensuring that it detects a high proportion of tumors correctly. This means that even if there is a higher number of false positives (cases where the model predicts the presence of a tumor incorrectly), it is acceptable because the primary objective is to minimize the chances of missing a cancerous tumor.
   - model will optimize for Recall.

4. Q4: Define supervised and unsupervised learning.

   - supervised: $D = \{\{x_1, y_1\}, ..., \{Px_n, y_n\}\}$, where we have access to pairs of input features $x_i \in \mathcal{R}$ and targets y_i , indexed by i
   - unsupervised: learn the data probability distribution $p(x)$ in implicit or explicit way.

5. Q5: What is overfitting? How can it be prevented?

   - Overfitting is a phenomenon that occurs when a machine learning model becomes overly complex and starts to memorize the training data instead of learning general patterns that can be applied to new, unseen data. In other words, the model becomes too specialized in fitting the training data and fails to generalize well to new data.

   (a) Regularization: In simple words, add a penalty during training, that encourages the model to learn simple representations of the input.

   (b) Data augmentation: Generate additional training examples by applying random transformations to the existing data, such as rotations, translations, or flips, to increase the diversity of the training data.

   (c) ...

6. Q6: Explain the Curse of Dimensionality.

   - The Curse of Dimensionality refers to the challenges and difficulties that arise when working with high-dimensional data. It describes various problems that occur as the number of features or dimensions in a dataset increases. When the dimensionality of a dataset is high, several issues arise:

   (a) Sparsity of Data: As the number of dimensions increases, the available data becomes increasingly sparse. In high-dimensional spaces, the amount of data required to adequately cover the space grows exponentially. This sparsity makes it difficult to obtain reliable statistical estimates, as the data points become more dispersed.

   (b) Overfitting: High-dimensional data increases the risk of overfitting, where a model becomes overly complex and fits the noise or random variations present in the data. With a large number of features, the model has more degrees of freedom to fit the training data perfectly, but it may fail to generalize well to new data. This can lead to poor predictive performance.

   (c) Curse of Sample Size: As the dimensionality increases, the number of samples required to obtain reliable statistical estimates also increases exponentially. The available training data may not be sufficient to capture the underlying patterns and relationships accurately. This lack of data can lead to unstable and unreliable models.

1.2 Linear algebra

1. Q1: What is the trace of a matrix A? How is it defined?

   - The trace of a square matrix A is the sum of the elements on its main diagonal. The main diagonal of a square matrix consists of the elements that are positioned from the top left to the bottom right of the matrix.

   - Mathematically, if A is an $n \times n$ matrix, then the trace of A, denoted as $\mathrm{tr}(A)$, is given by:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}$$

   where A_{ii} represents the i-th diagonal element of matrix A. In other words, the trace is the sum of the entries A_{11}, A_{22}, A_{33}, and so on, up to A_{nn}.

   - The trace of a matrix has several important properties. Some of these properties include:

   (a) For any two square matrices A and B of the same size, $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$.

   (b) For any square matrix A and scalar c, $\mathrm{tr}(cA) = c \cdot \mathrm{tr}(A)$.

   (c) For any square matrices A and B such that the product AB is defined, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. This property is known as the cyclic property of the trace.

   - The trace of a matrix is often used in various areas of mathematics, including linear algebra, matrix theory, and differential equations. It has several applications and interpretations, such as calculating matrix similarity, determining eigenvalues, and defining matrix norms.

2. Q2: What is the determinant of a matrix A? How is it defined?

- The determinant of a square matrix A is a scalar value that is computed from the elements of the matrix. It provides important information about the matrix and its linear transformation properties. The determinant is denoted as $\det(A)$ or $|A|$.

  For a square matrix A of size $n \times n$, the determinant is defined recursively. The determinant of a 2×2 matrix is computed as follows:

$$\det(A) = \begin{vmatrix} a & b \, c & d \end{vmatrix} = ad - bc$$

- For matrices larger than 2×2, the determinant is calculated using expansion by minors or cofactor expansion along any row or column. Let's consider the expansion along the first row as an example:

$$\det(A) = a_{11}C_{11} - a_{12}C_{12} + a_{13}C_{13} - \ldots + (-1)^{1+n}a_{1n}C_{1n}$$

  where a_{ij} represents the element of matrix A in the i-th row and j-th column, and C_{ij} represents the cofactor of a_{ij}. The cofactor C_{ij} is defined as the determinant of the submatrix obtained by deleting the i-th row and j-th column of A, multiplied by $(-1)^{i+j}$.

- It is important to note that the determinant exists only for square matrices. The determinant can be used to determine if a matrix is invertible. If the determinant is nonzero, the matrix is invertible; otherwise, it is singular.
- The determinant has several properties, some of which include:

(a) For a square matrix A, if we interchange any two rows or columns of A, the sign of the determinant changes.

(b) If we multiply any row or column of A by a scalar c, the determinant is multiplied by c. If A has two identical rows or columns, the determinant is zero. If A is an upper triangular or lower triangular matrix, the determinant is the product of the diagonal elements.

3. Q3: Let $A = \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix}$. Compute the determinant, the trace and the rank of A.

- Determinant:
  Using the cofactor expansion along the first row, we have:

$$\begin{aligned}
\det(A) &= \begin{vmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{vmatrix} \\
&= 8\begin{vmatrix} 5 & 7 \\ 9 & 2 \end{vmatrix} - 1\begin{vmatrix} 3 & 7 \\ 4 & 2 \end{vmatrix} + 6\begin{vmatrix} 3 & 5 \\ 4 & 9 \end{vmatrix} \\
&= 8(10 - 63) - (6 - 28) + 6(27 - 20) \\
&= 8(-53) - (-22) + 6(7) \\
&= -424 + 22 + 42 \\
&= -360.
\end{aligned}$$

- Trace: The trace of matrix A is the sum of the diagonal elements. Therefore, $[\mathrm{tr}(A) = 8 + 5 + 2 = 15]$
- Rank: The rank of a matrix is the maximum number of linearly independent rows or columns. We can find the rank by performing row reduction via Gaussian Elimination. The rank is the number of lines which are non-zeros.

$$\begin{pmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{pmatrix} \times \left(\frac{-3}{8}\right) \quad \underset{\mathrm{II}-\left(\frac{3}{8}\right)\cdot I \to \mathrm{II}}{\sim} \begin{pmatrix} 8 & 1 & 6 \\ 0 & \frac{37}{8} & \frac{19}{4} \\ 4 & 9 & 2 \end{pmatrix} \times \left(\frac{-1}{2}\right) \quad \underset{\mathrm{III}-\left(\frac{1}{2}\right)\cdot I \to \mathrm{III}}{\sim} \begin{pmatrix} 8 & 1 & 6 \\ 0 & \frac{37}{8} & \frac{19}{4} \\ 0 & \frac{17}{2} & -1 \end{pmatrix} \times \left(\frac{-68}{37}\right) \quad \underset{\mathrm{III}-\left(\frac{68}{37}\right)\cdot \mathrm{II} \to \mathrm{III}}{\sim} \begin{pmatrix} 8 & 1 & 6 \\ 0 & \frac{37}{8} & \frac{19}{4} \\ 0 & 0 & \frac{-360}{37} \end{pmatrix}$$

$$\mathrm{rang}\begin{pmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{pmatrix} = \mathrm{rang}\begin{pmatrix} 8 & 1 & 6 \\ 0 & \frac{37}{8} & \frac{19}{4} \\ 0 & 0 & \frac{-360}{37} \end{pmatrix} = 3$$

4. Q4: How are Eigenvectors and Eigenvalues defined?

   - Given a square matrix A, an eigenvector of A is a non-zero vector $\mathbf{u}$ that, when multiplied by A, results in a vector that is parallel to $\mathbf{u}$. In other words, applying the linear transformation represented by A to an eigenvector simply scales the eigenvector by a scalar value, known as the eigenvalue.
   - Mathematically, for a square matrix A and a non-zero vector $\mathbf{u}$, we have:

   $$A\mathbf{u} = \lambda\mathbf{u}$$

   Here, λ is the eigenvalue corresponding to the eigenvector $\mathbf{u}$. The eigenvalue represents the factor by which the eigenvector is scaled when multiplied by A.
   - To find the eigenvectors and eigenvalues of a matrix, we solve the equation $(A - \lambda I)\mathbf{u} = \mathbf{0}$, where I is the identity matrix and $\mathbf{0}$ is the zero vector. This equation is equivalent to finding the nullspace (kernel) of the matrix $(A - \lambda I)$.
   - To find the eigenvalues of a general matrix A, you need to solve the characteristic polynom $\det(A - \lambda I) = 0$, where λ is the eigenvalue and I is the identity matrix.
   - Once you have found the eigenvalues, you can further find the corresponding eigenvectors by solving the equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$, where $\mathbf{v}$ is the eigenvector associated with the eigenvalue λ. The solution to this equation will give you the eigenvectors.

5. Q5: Find the Eigenvalues of the matrix $B = \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix}$.

   - The characteristic polynomial for B is

   $$\det(B - tI) = \begin{bmatrix} -2 - t & -1 \\ 5 & 2 - t \end{bmatrix} = t^2 + 1.$$

   - The eigenvalues are the solutions of the characteristic polynomial. Thus solving $t^2 + 1 = 0$, we obtain eigenvalues $\pm i$, where $i = \sqrt{-1}$. Thus the eigenvalue of a real matrix B is pure imaginary numbers $\pm i$.

6. Q6: Find all the Eigenvalues and Eigenvectors of the matrix $A = \begin{bmatrix} 3 & -2 \\ 6 & -4 \end{bmatrix}$.

- To determine eigenvalues of A, we compute the determinant of $A - \lambda I$. We have

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 3-\lambda & -2 \\ 6 & -4-\lambda \end{vmatrix} \\ &= (3-\lambda)(-4-\lambda) + 12 \\ &= \lambda^2 + \lambda = \lambda(\lambda+1). \end{aligned}$$

- The eigenvalues are solutions of $\det(A - \lambda I) = 0$, hence eigenvalues of A are $0, -1$. Next, we find the eigenvector corresponding to the eigenvalue $\lambda = 0$. Eigenvectors $\mathbf{x}$ are nonzero solutions of $(A - 0I)\mathbf{x} = \mathbf{0}$. Thus, we solve $A\mathbf{x} = \mathbf{0}$. The augmented matrix of the system is

$$\left[\begin{array}{cc|c} 3 & -2 & 0 \\ 6 & -4 & 0 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c} 3 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right] \xrightarrow{\frac{1}{3}R_1} \left[\begin{array}{cc|c} 1 & -2/3 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

  Thus, if $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is a solution, then $x_1 = \frac{2}{3}x_2$, hence

$$\mathbf{x} = x_2 \begin{bmatrix} 2/3 \\ 1 \end{bmatrix}$$

  is an eigenvector corresponding to $\lambda = 0$ for any nonzero scalar x_2.

- Finally, we find the eigenvectors corresponding to the eigenvalue $\lambda = -1$. In this case we need to solve $(A - (-1)I)\mathbf{x} = \mathbf{0}$. The augmented matrix is

$$\left[\begin{array}{cc|c} 4 & -2 & 0 \\ 6 & -3 & 0 \end{array}\right] \xrightarrow[\frac{1}{3}R_2]{\frac{1}{2}R_1} \left[\begin{array}{cc|c} 2 & -1 & 0 \\ 2 & -1 & 0 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right] \xrightarrow{\frac{1}{2}R_1} \left[\begin{array}{cc|c} 1 & -1/2 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

  It follows that the eigenvectors associated to $\lambda = -1$ are

$$\mathbf{x} = x_2 \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$$

  for any nonzero scalar x_2.