Centro de Investigación en Ciencias

Diplomado en Ciencia de Datos con Python

Estadística y Algebra Lineal

Módulo 1

Mauricio Rosales-Rivera

Universidad Autónoma del Estado de Morelos

Algebra Lineal

Es la rama de las matemáticas que estudia las representaciones y operaciones de ecuaciones lineales, en particular:

Vectores.	Matrices.

Muchos problemas son representados usando vectores y matrices, por lo que es importante que esten familiarizado con los conceptos.

Vectores

Si n es un número entero positivo y $\mathbb{R}$ es el conjunto de números reales. Entonces $\mathbb{R}^n$ es el conjunto de todas las n-tuplas de números reales.

Un vector:

$$\vec{v} \in \mathbb{R}^n$$

es una n-tupla de número reales.

Por ejemplo:

$$\vec{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$$

Vectores

Operaciones con vectores (1)

Usemos los vectores $\vec{u} = (u_1, u_2, u_3)$ y $\vec{v} = (v_1, v_2, v_3)$ para ejemplificar las operaciones con vectores:

- suma:

$$\vec{u} + \vec{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$$

- resta:

$$\vec{u} - \vec{v} = (u_1 - v_1, u_2 - v_2, u_3 - v_3)$$

- norma:

$$||\vec{u}|| = \sqrt{u_1^2 + u_2^2 + u_3^2}$$

Vectores

- vector por escalar:

$$\lambda\vec{u} = (\lambda u_1, \lambda u_2, \lambda u_3)$$

- producto cruz:

$$\vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1)$$

Genera un nuevo vector. Solo para vectores en $\mathbb{R}^3$.

Operaciones con vectores (2)

Vectores

- producto punto o escalar (inner):

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + u_3 v_3$$

Genera un escalar.

- producto tensorial (outer):

$$\vec{u} \otimes \vec{v} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \end{bmatrix}$$

Genera una matriz.

Operaciones con vectores (3)

Matrices

Una matriz es un arreglo bidimensional de elementos numéricos. Una matriz $A \in \mathbb{R}^{m \times n}$ es un arreglo con dos dimensiones (m filas y n columnas) que contiene números reales.

Por ejemplo:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \in \mathbb{R}^{3 \times 2}$$

A se utiliza para referirse a la matriz entera y a_{ij} para referirse a sus entradas. i denotando a la fila y j a la columna.
La notación a_{*j} o $a_{*,j}$ es utilizada para referirse a la columna j de la matriz.
La notación a_{i*} o $a_{i,*}$ es utilizada para referirse a la fila i de la matriz.

Matrices

Operaciones de matrices (1)

- suma:

$$C = A + B \Leftrightarrow c_{ij} = aij + bij$$

Por ejemplo:

$$C = A + B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b \\ a_{21} + b_{21} & a_{22} + b \end{bmatrix}$$

- resta:

$$C = A - B \Leftrightarrow c_{ij} = aij - bij$$

Por ejemplo:

$$C = A - B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b \\ a_{21} - b_{21} & a_{22} - b \end{bmatrix}$$

Las dimensiones de las matrices deben ser iguales para poderse sumar restar.

Matrices

Operaciones de matrices (2)

- producto por un escalar:

$$C = \lambda A \Leftrightarrow c_{ij} = \lambda a_{ij}$$

Por ejemplo:

$$C = \lambda A = \lambda \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda a_{11} & \lambda a_{12} \\ \lambda a_{21} & \lambda a_{22} \\ \lambda a_{31} & \lambda a_{32} \end{bmatrix}$$

Matrices

- producto:
  Si $A \in \mathbb{R}^{m\times n}$ y $B \in \mathbb{R}^{n\times p}$ entonces:

$$C = AB \in \mathbb{R}^{m\times p} \Leftrightarrow c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

  Por ejemplo si:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

  entonces

$$C = AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} \in \mathbb{R}^{3\times 2}$$

El producto de dos matrices solo existe cuando el número de columnas en A es igual al número de filas en B.

Operaciones de matrices (3)

Matrices

- traspuesta:

$$C = A^T \Leftrightarrow c_{ij} = aji$$

Por ejemplo, si:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

entonces

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$$

La matriz traspuesta de A es la misma matriz con las filas y col intercambiadas.

Operaciones de matrices (4)

Matrices

Operaciones de matrices (5)

- traza:
  La traza de una matriz cuadrada $A \in \mathbb{R}^{n \times n}$ se denota por tr$A$ y es la suma de los elementos en su diagonal:

$$\mathrm{tr}A = \sum_{i=1}^{n} a_{ii}$$

- identidad:
  La matriz identidad se denota como $I \in \mathbb{R}^{n \times n}$, es una matriz cuadrada con unos en su diagonal y ceros en el resto de sus entradas.

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

  La matriz identidad tiene la propiedad que para toda matriz $A \in \mathbb{R}^{mxn}$:

$$AI = A = IA$$

Matrices

Operaciones de matrices (6)

- inversa:
  La matriz inversa de una matriz cuadrada $A \in \mathbb{R}^{n \times n}$ se denota como A^{-1}, y es una matriz única tal que:

  $$AA^{-1} = I = A^{-1}A$$

  No todas las matrices tienen inversa. Una matriz es invertible o no-singular si tiene inversa y no invertible o singular en el caso contrario.
- determinante:
  El determinante de una matriz cuadrada $A \in \mathbb{R}^{n \times n}$ se denota por $\det(A)$ o $|A|$ es una función $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$.
  Para una matriz 2x2:

  $$A = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \qquad |A| = a_{11}a_{22} - a_{12}a_{21}$$

Vectores como matrices

Es común representar a vectores n-dimensionales como matrices que contienen n filas y 1 columna, a esta representación se le conoce como **vector columna**. Por ejemplo, para un vector en $\mathbb{R}^n$:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Si se requiere expresar un vector como una fila (**vector fila**) de manera explícita usando matrices (1 fila y n columnas) usamos la notación *traspuesta* x^T. Por ejemplo:

$$x^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Representando vectores como matrices

Producto

Producto matriz-vector

Usando la representación matricial de los vectores, podemos definir el siguiente tipo de producto.
Dada una matriz $A \in \mathbb{R}^{m \times n}$ y un vector $x \in \mathbb{R}^n$ ($x \in \mathbb{R}^{n \times 1}$ como vector columna) su producto:

$$y = Ax \in \mathbb{R}^m$$

$$= \begin{bmatrix} a_{1,1} \dots a_{1,n} \\ \vdots \\ a_{m,1} \dots a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1,*} \cdot x \\ \vdots \\ a_{m,*} \cdot x \end{bmatrix}$$

La i-esima entrada del vector y es igual al producto punto de la fila i de la matriz A y el vector x.

Producto

En el segundo caso, es posible realizar el producto de un vector fila por una matriz.
Dada una matriz $A \in \mathbb{R}^{m\times n}$ y un vector $x \in \mathbb{R}^m$ ($x^T \in \mathbb{R}^{1\times m}$ como vector fila), el producto:

$$y^T = x^T A \in \mathbb{R}^n$$

$$=$$

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{1,1} \dots a_{1,n} \\ \vdots \\ a_{m,1} \dots a_{m,n} \end{bmatrix} = \begin{bmatrix} x^T \cdot a_{*,1} & \dots & x^T \cdot a_{*,n} \end{bmatrix}$$

La i-esima entrada del vector fila y^T es igual al producto punto del vector x y la columna i de la matriz A.

Estadística

Estadística

Estadística: disciplina **encargada de estudiar la recolección, organización, análisis, visualización e interpretación de datos con el objetivo de generar conocimiento.**

Las variables aleatorias son aquellas que toman **valores de forma aleatoria como resultado de un experimento.**

Discreta: **si el conjunto de valores que puede tomar se encuentra en un espacio discreto, el conjunto de valores se puede numerar.**
Ejemplo, el lanzamiento de un dado.

Continua: si **los valores que puede tomar** se encuentra **en un** espacio continuo, **el conjunto de valores no se puede numerar.**
Ejemplo, mediciones en un rango de números reales.

Probabilidad

Distribución de probabilidad

- Una función de probabilidad de una variable aleatoria, es aquella que asigna a cada valor una probabilidad para que suceda un experimento.

Función de distribución acumulada

La función de distribución acumulada de una variable aleatoria X, es una función que asigna a cada uno de sus valores x la probabilidad de que la variable X asuma un valor igual o menor a x. Matemáticamente:

$$F(x) = P(X \leq x)$$

Función de probabilidad

En muchos casos es más importante conocer la probabilidad de una variable aleatoria tome un valor particular. Para esto usamos las funciones de masa o densidad.

- Para una variable **discreta**. La función de masa de probabilidad para una variable discreta es la función que mapea la probabilidad de la variable X tome el valor x en algún experimento. Matemáticamente:

$$F(x) = P(X = x)$$

- Para una variable **continua**: La función de densidad de probabilidad para una variable continua es la función que mapea la probabilidad de que la variable X se encuentre en un intervalo [a, b]. Matemáticamente:

$$P(a \leq X \leq b)$$

Distribución Binomial

Una variable aleatoria binomial, cuenta el número de éxitos al repetir n experimentos con dos resultados posibles (éxito o fracaso), cada experimento con una probabilidad p de ser exitoso (y $1-p$ de fracasar).

La función de probabilidad de una variable aleatoria binomial:

$$B(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad 0 \leq p \leq 1$$

$$\text{donde} \quad x = \{1, 2, 3, \ldots, n\}$$

y siendo
$\binom{n}{x} = \frac{n!}{x!(n-x)!}$
n - número de experimentos.
p - probabilidad de experimento exitoso.

Distribución Uniforme

Los valores de una variable aleatoria uniforme tienen la misma probabilidad de presentarse en cualquier experimento.

Función de densidad de una variable aleatoria uniforme continua en un intervalo $[A, B]$:

$$U(x; A, B) = \begin{cases} \dfrac{1}{B - A} & A \leq x \leq B \\ 0 & \text{cualquier otro caso.} \end{cases}$$

Distribución Normal

Los valores de una variable aleatoria normal tienen mayor probabilidad de estar cerca de la media y se dispersan disminuyendo su probabilidad al alejarse de ésta de acuerdo con su desviación estándar.

Función de densidad de una variable aleatoria normal con media μ y varianza σ^2:

$$N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty$$

Teorema del Límite Central

El teorema del límite central indica que la suma o la media de n variables aleatorias independientes se aproxima a una distribución normal cuando el tamaño de n aumenta. Matemáticamente:
Sean $X_1, X_2 \ldots X_n$ variables aleatorias independientes con media u y varianza finita σ^2 y $\bar{X} = \sum_{i=1}^{n} X_i/n$, entonces:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

cuando n es lo suficientemente grande.

Este teorema es importante porque en la práctica muchos de los variables observadas pueden interpretarse como la suma de múltiples variables aleatorias independientes. Esto justifica que estimadores de una distribución normal puedan ser utilizados en muestras de gran tamaño.

Estimadores

Estimadores de tendencia central

- Media aritmética:

$$\mu = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

  Para la población (μ) o muestra ($\bar{x}$)

- Mediana:

$$x\left(\frac{n+1}{2}\right) \text{ si } n \text{ impar} \qquad \frac{x\left(\frac{n+1}{2}\right) + x\left(\frac{n}{2}\right)}{2} \text{ si } n \text{ par}$$

- Moda:
  Es el valor con más ocurrencias en la muestra o población.

Estimadores

Estimadores de dispersion

- Varianza:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{o} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Para la población σ^2 o una muestra s^2.

- Desviación estándar:

$$\sigma = \sqrt{\sigma^2} \quad \text{o} \quad s = \sqrt{s^2}$$

Para la población σ o una muestra s.