

Introduction to Linear Algebra

ÁLGEBRA LINEAL

El **álgebra lineal** son las matemáticas de los datos. Se usa la aritmética en **vectores y matrices**. También es el estudio de líneas y planos, espacios vectoriales y mapeos para las **transformaciones lineales**. Con ayuda de ella, podemos resolver un sistema de ecuaciones con uno o más términos.

Por ejemplo, consideremos el sistema de ecuaciones:

$$\begin{aligned}y &= 0.1 x_1 + 0.4 x_2 \\y &= 0.3 x_1 + 0.9 x_2 \\y &= 0.2 x_1 + 0.3 x_2\end{aligned}$$

A la columna de valores **y** la podemos tomar como un **vector columna** lleno de outputs. Las dos columnas de valores enteros se pueden tomar como una **matriz A**. Los valores desconocidos, en este caso las **x_i** , son tomados junto con los coeficientes de las ecuaciones y juntos forman un vector **b** de **valores desconocidos** a resolver.

$$y = A \cdot b$$

ÁLGEBRA LINEAL NUMÉRICA

El **álgebra lineal numérica** es la aplicación del álgebra lineal en la computadora.

Algunas **librerías importantes** en Python para el ALN son:

- Linear Algebra Package
- Basic Linear Algebra Subprograms
- Automatically Tuned Linear Algebra Software

ÁLGEBRA LINEAL Y ESTADÍSTICA

En estadística, el álgebra lineal tiene las siguientes aplicaciones:

- Uso de notación **vectorial y matricial**, sobre todo en la Estadística multivariada.
- Solución de **mínimos cuadrados**, como en Regresión Lineal.
- Estimar la **media y varianza** de datos matriciales, así como también obtener la **covarianza** en distribuciones Gaussianas multinomiales.
- Análisis de Componentes Principales** para la disminución de datos.

APLICACIONES DEL ÁLGEBRA LINEAL

Matrices en ingeniería, gráficas y redes, matrices de Markov, programación lineal, series de Fourier, estadística y probabilidad y gráficas.

Linear Algebra and Machine Learning

Con el álgebra lineal podemos resolver **sistemas de ecuaciones** donde hay más ecuaciones que variables desconocidas. Pero, al no existir una **única solución**, se utiliza el método de **mínimos cuadrados lineales** para resolver este tipo de sistemas.

Los problemas con **mínimos cuadrados lineales** pueden resolverse usando operaciones matriciales, como la **factorización de matrices**. Ésta se usa para realizar operaciones más complejas, como para obtener la **inversa de una matriz**.

Examples of Linear Algebra in M.L.

DATASET Y DATOS

Lo ideal es que los datos estén en un **formato de matriz** donde cada fila represente una **observación** y cada columna una **característica** de dicha observación.

IMÁGENES Y FOTOGRAFÍAS

Cada imagen que vemos puede ser representada por una **tabla** pues ésta cuenta con **longitud y altura**, además de que cuenta con un **valor de pixel** en cada celda

para el caso de las imágenes en blanco y negro, o bien, **tres valores de pixel** para imágenes de colores.

ONE HOT ENCODING

Es un método utilizado para **codificar variables categóricas**. Se tiene una tabla que representa la variable con una columna por **categoría** y una fila para cada **ejemplo** en el dataset. Un valor de **1** se agrega en la columna correspondiente dada una fila y se agrega el valor **0** en cualquier otro caso, teniendo así un **vector binario**.

Por ejemplo, si tenemos la siguiente variable color con tres filas:

1	red
2	green
3	blue

Se codificaría como:

red	green	blue
1	0	0
0	1	0
0	0	1

REGRESIÓN LINEAL

La **regresión lineal** describe la relación entre variables y se ocupa en Machine Learning para **predecir valores numéricos**. En ella, se debe encontrar un **conjunto de coeficientes** que, cuando multipliquen a cada una de las variables de entrada y éstas se sumen, resulte en la **mejor predicción** de la variable de salida.

La mejor manera de resolver este problema es mediante el método de **Mínimos Cuadrados**, ocupando la **Descomposición LU** o la **Descomposición de Valores Singulares**.

REGULARIZACIÓN

Los **modelos simples** se caracterizan por modelos que tienen pocos coeficientes o variables, y, con ayuda de la **regularización**, se puede **minimizar el número de coeficientes o variables**. También existe la regularización **L^1** y **L^2** .

ANÁLISIS DE COMPONENTES PRINCIPALES

Para **reducir el número de columnas** a ocupar de un conjunto de datos, se utilizan métodos para **reducir la dimensión**. Uno de los métodos más populares es el **PCA**. Éste ocupa el **método de factorización de matrices**, aunque también puede ser ocupado el método de **eigendecomposición**.

DESCOMPOSICIÓN DE VALORES SINGULARES

También conocido como **SVD**. Es una **factorización de matrices** usado para la selección de características, reducción de ruido y más.

LATENT SEMANTIC ANALYSIS

En Machine Learning es común que, al trabajar con text data, se represente a los documentos como **grandes matrices** de word occurrences. Por ejemplo, las filas serían las **palabras conocidas** y las columnas las **oraciones, párrafos, páginas o documentos de texto**. Las celdas de la matriz indicarán la **frecuencia** de las veces en que apareció **x** palabra.

Al usar métodos de factorización para preparar este tipo de datos, como SVD, se le conoce como **Latent Semantic Analysis (LSA)**.

RECOMMENDER SYSTEM

Se usan para crear los **sistemas** que recomiendan productos. Un ejemplo es calcular la **similitud** entre el comportamiento de los clientes usando **vectores** y verificando la **distancia** que existe entre ellos.

DEEP LEARNING

Es el **resurgimiento** del uso de las **redes neuronales** con nuevos y más rápidos métodos que permiten el **desarrollo y entrenamiento** de redes más profundas para bases de datos más grandes. Trabaja con vectores, matrices e incluso tensores de entradas y coeficientes (un **tensor** es una matriz con más de **2** dimensiones: cada columna es una **coordenada** y cada fila, un **punto**).