
Universidad
Tecmilenio®

Universidad Tecmilenio®

Matemáticas computacionales para inteligencia artificial

Cálculo vectorial

Introducción

La **retropropagación** es un método ideado años antes de los primeros modelos con redes neuronales. Basado en la intuición de atribuir una importancia reducida a cada evento ocurrido, a medida que se retrocede en la cadena que los enlaza.

Dicho método junto al algoritmo del gradiente descendente forman la columna vertebral y fuerza impulsora de las redes neuronales actuales. En este tema aprenderás sobre los fundamentos matemáticos en los que se soportan estas herramientas y conocerás algunos principios que están detrás de su funcionamiento.

Explicación

A medida que los sistemas se van complicando es necesario incluir nuevas herramientas que nos permitan realizar su análisis de una manera más sencilla.

Es el momento de incorporar las derivadas de orden superior, la matriz hessiana y el concepto de la derivada total.

Matriz hessiana:

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

Explicación

Por su parte, la derivada total se puede plantear de la siguiente forma:

$$\frac{df}{dt} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1(t)}{\partial t} \\ \frac{\partial x_2(t)}{\partial t} \end{bmatrix} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t}$$

Donde d es el gradiente y ∂ las derivadas parciales.

Explicación

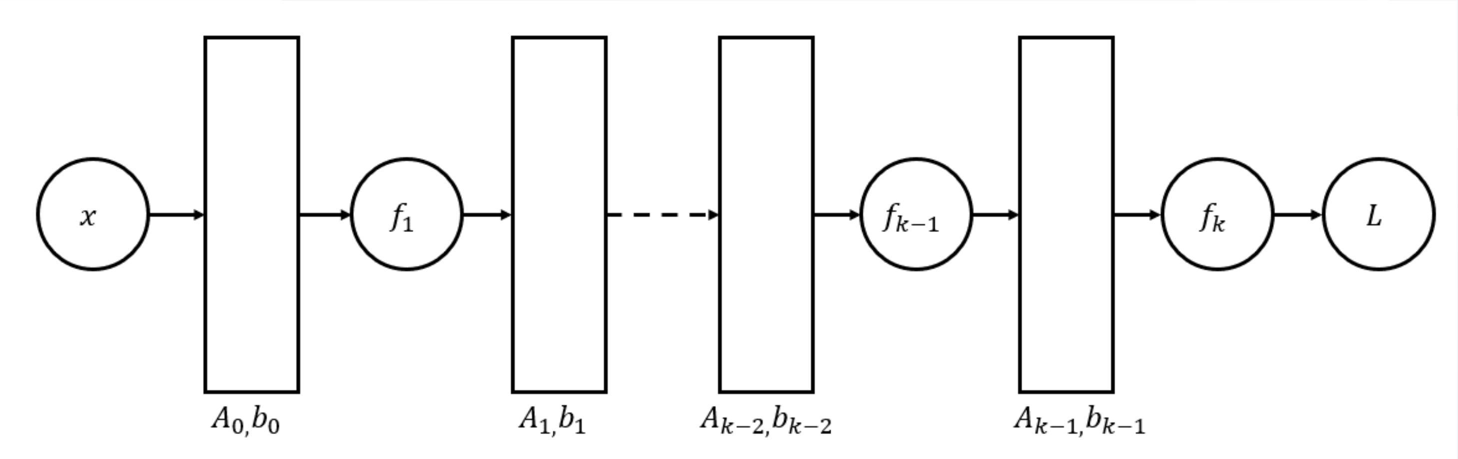
En algunas de las aplicaciones de aprendizaje automático podemos obtener los parámetros adecuados de un modelo aplicando el algoritmo del gradiente descendente sobre la función de aprendizaje con relación a los parámetros del modelo en cuestión.

La función principal se representa como una composición de funciones en subniveles inferiores:

$$\boldsymbol{y} = f_K\left(f_{K-1}\left(\dots\left(f_1\left(\boldsymbol{x}\right)\right)\dots\right)\right)$$

Donde $\boldsymbol{x}$ representa las entradas (por ejemplo: imágenes, textos, audios), $\boldsymbol{y}$ las observaciones (etiquetas) y cada función f_i , $i = 1, \dots, K$ posee sus propios parámetros. (Figura 2).

Explicación

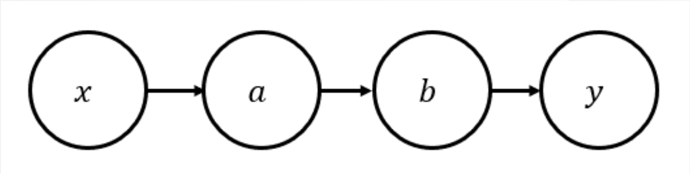


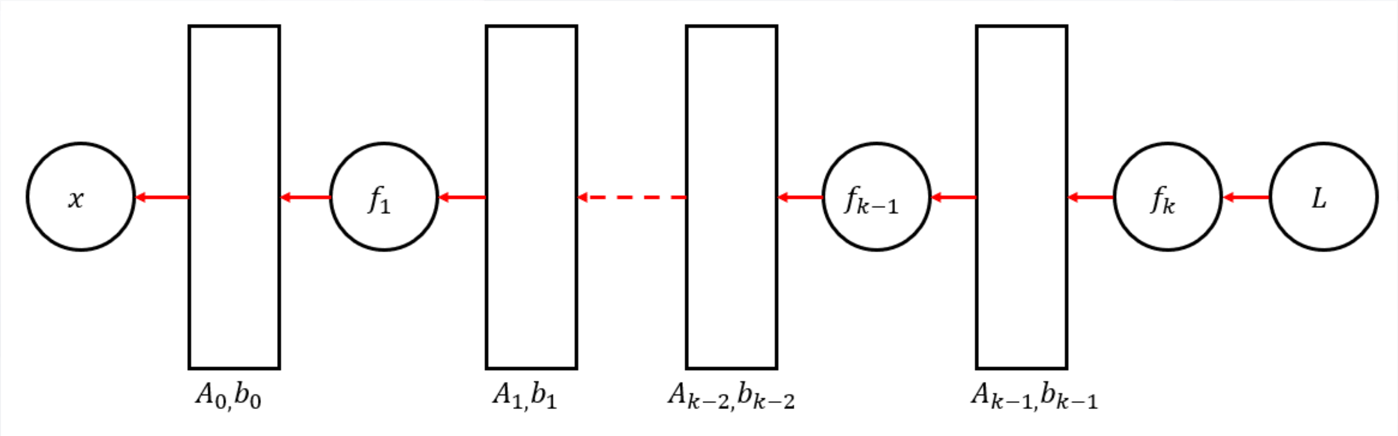
En una red neuronal con múltiples capas, encontramos la función $f_i(x_{i-1}) = \sigma(\boldsymbol{A}_{i-1}\boldsymbol{x}_{\boldsymbol{i}-1} + \boldsymbol{b}_{i-1})$ en cada i-ésima capa.

En tal caso x_{i-1} es la salida de la capa.

Explicación

Ambas figuras revelan el flujo de los datos dentro de la red y cómo ocurre el paso hacia atrás de los valores calculados del gradiente de la función de pérdida, es decir, la propagación inversa.





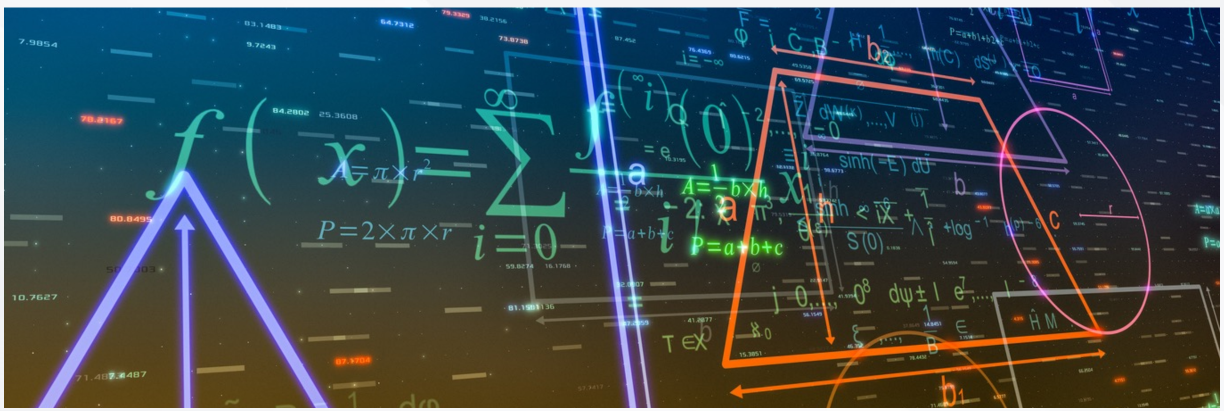
Actividad

Investiga sobre algún problema con imágenes, textos o audios que se resuelva aplicando el algoritmo de propagación inversa.

Cierre

El gradiente y la derivada son las herramientas que nos permiten representar las características de las funciones en puntos específicos de estas, y a partir de su análisis, se puede obtener un conjunto de información valiosa sobre el comportamiento del sistema que representan.

Las funciones pueden tener una o varias variables, pero, cuando se incrementa considerablemente el número de estas, es conveniente transformar las expresiones a su forma vectorial y trabajar en combinación con los recursos del álgebra lineal.



Universidad Tecmilenio

Matemáticas computacionales para inteligencia artificial

Distribuciones y probabilidades

Introducción

Las probabilidades, en la actualidad, pueden utilizarse para medir la ocurrencia de un suceso durante un experimento o para cuantificar la variabilidad de los datos, la calidad de los modelos de aprendizaje automático o la precisión de las predicciones realizadas por estos.

Asociado con la variable aleatoria, está la función que mide la probabilidad de ocurrencia de un posible resultado, dicha función normalmente se conoce como la **distribución de probabilidad**. Estas distribuciones probabilísticas son utilizadas para construir partes de otros conceptos, como el modelado de probabilidad, el modelado gráfico o la selección de un modelo específico para solucionar una problemática.

Explicación

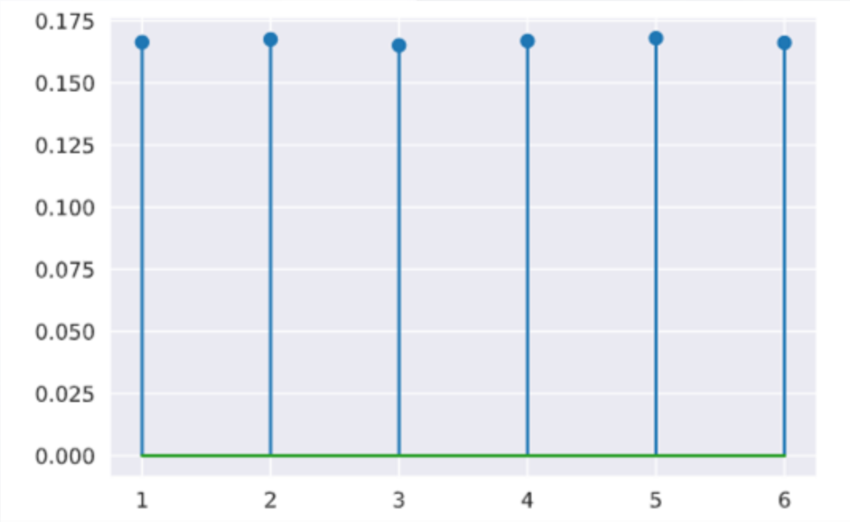
Espacio muestral (Ω): está conformado por el conjunto de posibles resultados aleatorios que puede tener el experimento.

Eventos aleatorios (A): es el conjunto de resultados potenciales que puede tener el experimento y constituye, a su vez, un subconjunto del espacio muestral.

Probabilidad (P): con cada evento A ∈ A se asocia un número P(A) que mide la probabilidad o el grado de certidumbre de que un evento pueda ocurrir. La notación P(A) es utilizada para expresar la probabilidad de A.

Explicación

La probabilidad continua incluye en su espacio objetivo el intervalo de valores reales aleatorios que puede tener una función continua.



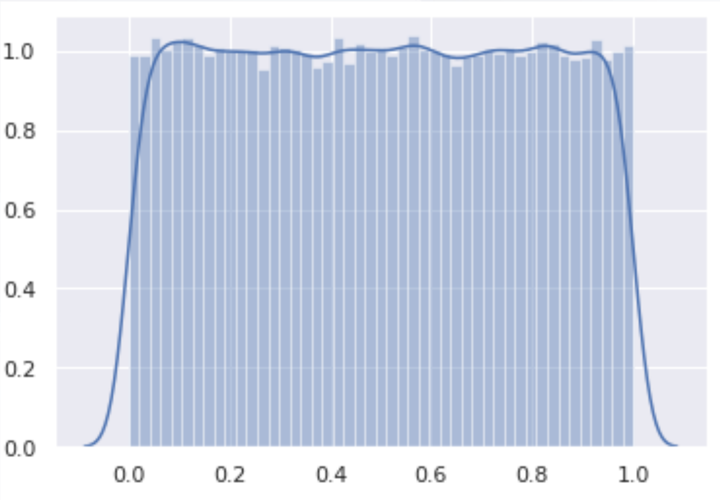
Una función de probabilidad de masa (***pmf***) puede representar eventos discretos, como el lanzamiento aleatorio de un dado un elevado número de veces. Además, es positiva y su integral vale 1.

Explicación

Por otra parte, una **función de distribución acumulativa** (cdf) de una variable aleatoria **X** con estados x ϵ R^D tiene una función de distribución acumulativa dada por:

$$F_X(\boldsymbol{x}) = P(X_1 \leq x_1, \ldots, X_D \leq x_D)$$

Donde $X = [X_1, \ldots X_D]^{\mathrm{T}}$, $\boldsymbol{x} = [x_1, \ldots, X_D]^{\mathrm{T}}$, y la parte derecha de la ecuación representa la probabilidad de que la variable aleatoria X_i, tome un valor menor o igual a x_i. La gráfica de un ejemplo de cdf (la distribución uniforme en este caso) se muestra en la figura.



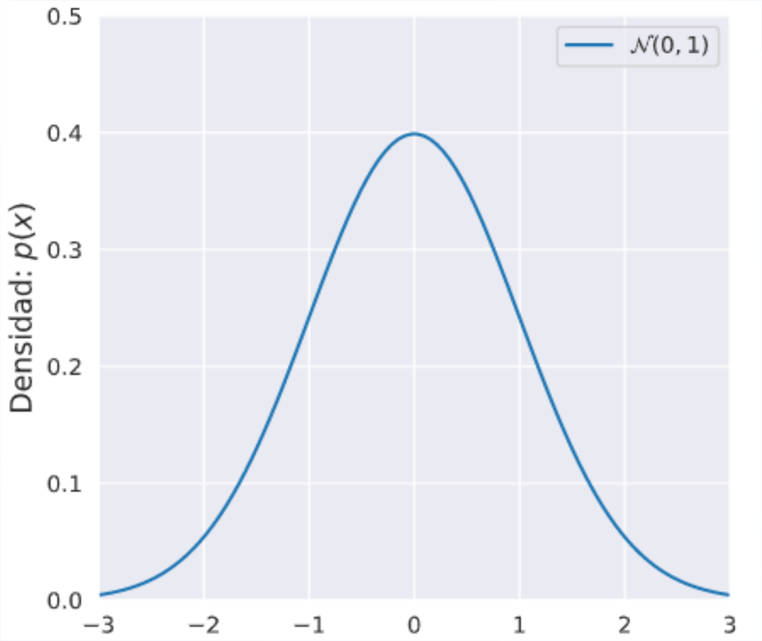
Explicación

A continuación se muestran las diferentes funciones y su aplicación en cada tipo de distribución probabilística.

Tipo	Probabilidad puntual	Intervalo de probabilidad
Discreta	$P(X = x)$ Función de probabilidad de masa	No aplica
Continua	$p(x)$ Función de densidad probabilística	Función de distribución acumulativa

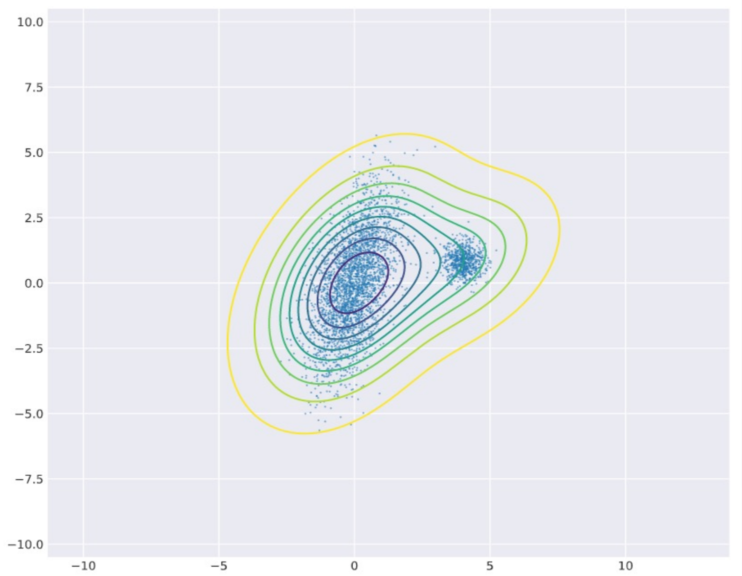
Explicación

La distribución gaussiana es una de las distribuciones más populares que se utilizan para describir las variables aleatorias continuas, también se conoce en la bibliografía como distribución normal.



Su importancia se origina gracias a que posee varias propiedades que, desde el punto de vista computacional, son muy convenientes. En el aprendizaje automático es muy utilizada para definir la verosimilitud a priori de la regresión lineal, para realizar los procesamientos gaussianos, la inferencia variacional o el aprendizaje reforzado.

También es posible visualizarla en su versión multidimensional (dos dimensiones):



Actividad

Piensa en qué problemas de tu campo podrían representarse con una distribución de probabilidad vista en este tema.

Cierre

En este tema solamente se han mencionado algunas de las distribuciones más relevantes, pero es importante decir que existen muchas otras como la distribución de Poisson o la geométrica.

Determinar exactamente cuál es la más adecuada para representar una problemática dependerá, sobre todo, de las características del fenómeno que se quiera modelar y de la experiencia acumulada durante años de investigación que nos sugiere cuáles son las distribuciones más efectivas para describir sucesos comunes previamente conocidos.