Fundamentos mátematicos de la IA

La Inteligencia Artificial (IA) y el Machine Learning se fundamentan en diversas ramas de las matemáticas. Estas herramientas permiten modelar datos, optimizar funciones y comprender patrones complejos. Los principales pilares matemáticos que se utilizan en IA son:

- **Probabilidad y Estadística:** Para modelar la incertidumbre, analizar datos y hacer inferencias.
- **Cálculo:** Fundamental en la optimización de modelos (por ejemplo, en el ajuste de parámetros mediante gradientes).
- **Álgebra Lineal:** Esencial para representar datos en forma de vectores y matrices, y para operaciones que se utilizan en algoritmos de aprendizaje.

A continuación, desarrollaremos cada uno de estos temas en detalle.

1. Probabilidad

La probabilidad es la rama de las matemáticas que estudia la incertidumbre y se utiliza para modelar eventos aleatorios.

1.1 Conceptos Básicos

Espacio Muestral y Eventos

- **Espacio muestral (S):** Es el conjunto de todos los posibles resultados de un experimento aleatorio.
  Ejemplo: Para el lanzamiento de un dado, $S = \{1, 2, 3, 4, 5, 6\}$.
- **Evento (A):** Es un subconjunto del espacio muestral.
  Ejemplo: Sacar un número par: $A = \{2, 4, 6\}$.

Definición de Probabilidad

La probabilidad de un evento A se define como:

$$P(A) = \frac{\text{Número de resultados favorables a } A}{\text{Número total de resultados en } S}$$

Ejemplo 1.1:

Para el dado:

- Número total de resultados: 6.
- Número de resultados favorables para obtener un número par: 3 (2, 4 y 6).

$$P(\text{par}) = \frac{3}{6} = 0.5$$

1.2 Variables Aleatorias y Funciones de Distribución

Variable Aleatoria

Una variable aleatoria es una función que asigna un número real a cada resultado del espacio muestral.

Ejemplo:

En el lanzamiento de un dado, definimos la variable aleatoria X que toma el valor del número obtenido.

Función de Probabilidad (para variables discretas)

La función de probabilidad $p(x)$ asigna la probabilidad de que X tome un valor x:

$$p(x) = P(X = x)$$

Para el dado, $p(x) = \frac{1}{6}$ para $x = 1, 2, \ldots, 6$.

Función de Distribución Acumulada (CDF)

La función de distribución acumulada $F(x)$ se define como:

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$$

1.3 Esperanza, Varianza y Desviación Estándar

Esperanza o Valor Esperado

La esperanza $E[X]$ de una variable aleatoria X es el promedio ponderado de todos los valores posibles, con sus probabilidades:

$$E[X] = \sum_i x_i \, p(x_i)$$

Ejemplo 1.2:

Para el dado, el valor esperado es:

$$E[X] = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

Varianza

La varianza $\mathrm{Var}(X)$ mide la dispersión de la variable respecto a su media:

$$\mathrm{Var}(X) = E[(X - E[X])^2] = \sum_i (x_i - E[X])^2 \, p(x_i)$$

Ejemplo 1.3:

Para el dado, usando $E[X] = 3.5$:

$$\mathrm{Var}(X) = \frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6}$$

$$= \frac{(2.5)^2 + (1.5)^2 + (0.5)^2 + (0.5)^2 + (1.5)^2 + (2.5)^2}{6} = \frac{6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25}{6} = \frac{17.5}{6} \approx 2.92$$

Desviación Estándar

La desviación estándar σ es la raíz cuadrada de la varianza:

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

$$\sigma \approx \sqrt{2.92} \approx 1.71$$

1.4 Distribuciones de Probabilidad Comunes

Distribución Binomial

Se utiliza para modelar el número de éxitos en n ensayos independientes, cada uno con probabilidad p de éxito.

La función de probabilidad es:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

donde:

- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ es el coeficiente binomial,
- n es el número total de ensayos,
- k es el número de éxitos.

Ejemplo 1.4:

Supongamos 10 lanzamientos de una moneda justa ($p = 0.5$), la probabilidad de obtener 6 caras:

$$P(X = 6) = \binom{10}{6} (0.5)^6 (0.5)^4 = \binom{10}{6} (0.5)^{10}$$

$$\binom{10}{6} = \frac{10!}{6!4!} = 210, \quad \text{entonces } P(X = 6) = 210 \times (0.5)^{10} = 210 \times \frac{1}{1024} \approx 0.205$$

2. Estadística

La estadística se ocupa de recolectar, analizar e interpretar datos. Es crucial en la IA para la toma de decisiones basadas en datos.

2.1 Medidas Descriptivas

Media

La media (o promedio) de un conjunto de datos $x_1, x_2, \ldots, x_n$ es:

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

Ejemplo 2.1:

Para los datos $4, 8, 6, 5, 3$:

$$\bar{x} = \frac{4+8+6+5+3}{5} = \frac{26}{5} = 5.2$$

Mediana

La mediana es el valor central de un conjunto de datos ordenado.

Ejemplo:

Para $\{3, 4, 5, 8, 9\}$, la mediana es 5.

Moda

La moda es el valor que ocurre con mayor frecuencia.

Ejemplo:

En $\{2, 3, 3, 5, 7\}$, la moda es 3.

2.2 Inferencia Estadística

Intervalos de Confianza

Un intervalo de confianza es un rango de valores que se utiliza para estimar un parámetro poblacional, con un cierto nivel de confianza (por ejemplo, 95%).

La fórmula para un intervalo de confianza para la media (cuando la varianza es conocida o el tamaño de la muestra es grande) es:

$$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

donde:

- $\bar{x}$ es la media muestral,
- $Z_{\alpha/2}$ es el valor crítico de la distribución normal (por ejemplo, para 95% es 1.96),
- σ es la desviación estándar poblacional,
- n es el tamaño de la muestra.

Ejemplo 2.2:

Supongamos que se mide la altura de 100 estudiantes y se obtiene una media de 170 cm con $\sigma = 10$ cm. El intervalo de confianza al 95% es:

$$170 \pm 1.96 \cdot \frac{10}{\sqrt{100}} = 170 \pm 1.96 \cdot 1 = 170 \pm 1.96$$

$$\text{Intervalo: } [168.04, 171.96]$$

Regresión Lineal Simple

La regresión lineal modela la relación entre una variable independiente x y una dependiente y usando una línea recta:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

donde:

- β_0 es la intersección (ordenada al origen),
- β_1 es la pendiente,
- ε es el término de error.

Para estimar β_0 y β_1, se usan las siguientes fórmulas (mínimos cuadrados):

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Ejemplo 2.3:

Considera los siguientes datos:

- $x : 1, 2, 3, 4$
- $y : 2, 3, 5, 4$

Calculamos:

- $\bar{x} = \frac{1+2+3+4}{4} = 2.5$
- $\bar{y} = \frac{2+3+5+4}{4} = 3.5$

Calcular β_1:

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = (1 - 2.5)(2 - 3.5) + (2 - 2.5)(3 - 3.5) + (3 - 2.5)(5 - 3.5) + (4 - 2.5)(4 - 3.5)$$

$$= (-1.5)(-1.5) + (-0.5)(-0.5) + (0.5)(1.5) + (1.5)(0.5) = 2.25 + 0.25 + 0.75 + 0.75 = 4$$

$$\sum (x_i - \bar{x})^2 = (-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2 = 2.25 + 0.25 + 0.25 + 2.25 = 5$$

$$\beta_1 = \frac{4}{5} = 0.8$$

$$\beta_0 = 3.5 - 0.8 \times 2.5 = 3.5 - 2 = 1.5$$

La ecuación de la recta es:

$$y = 1.5 + 0.8x$$

3. Cálculo

El cálculo es fundamental para la optimización en IA, ya que se utiliza para encontrar mínimos y máximos de funciones de pérdida y ajustar parámetros de modelos.

3.1 Derivadas y Gradiente

Derivada de una Función

La derivada de una función $f(x)$ es el límite que define la tasa de cambio instantánea:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Ejemplo 3.1:

Para $f(x) = x^2$:

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{2xh + h^2}{h} = \lim_{h \to 0} (2x + h) = 2x$$

Gradiente de una Función de Múltiples Variables

El gradiente es un vector de derivadas parciales y apunta en la dirección de mayor aumento de la función:

$$\nabla f(x_1, x_2, \ldots, x_n) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

Ejemplo 3.2:

Para $f(x, y) = x^2 + y^2$:

$$\frac{\partial f}{\partial x} = 2x, \quad \frac{\partial f}{\partial y} = 2y$$

Entonces,

$$\nabla f(x, y) = (2x, 2y)$$

Aplicación en Optimización: Descenso del Gradiente

El **descenso del gradiente** es un método para minimizar funciones (por ejemplo, la función de pérdida en un modelo de IA). La regla de actualización para cada parámetro θ es:

$$\theta := \theta - \alpha \nabla f(\theta)$$

donde:

- α es la tasa de aprendizaje (learning rate),
- $\nabla f(\theta)$ es el gradiente en θ.

Ejemplo 3.3:

Supongamos que queremos minimizar $f(x) = (x - 3)^2$.

- La derivada es $f'(x) = 2(x - 3)$.
- Con un $\alpha = 0.1$ y $x_0 = 0$, la actualización es:

$$x_1 = x_0 - 0.1 \times 2(0 - 3) = 0 - 0.1 \times (-6) = 0 + 0.6 = 0.6$$

- Repetir el proceso:

$$x_2 = 0.6 - 0.1 \times 2(0.6 - 3) = 0.6 - 0.1 \times (-4.8) = 0.6 + 0.48 = 1.08$$

El proceso se repite hasta que x se aproxime a 3, que es el mínimo de la función.

4. Álgebra Lineal

El álgebra lineal es esencial para el procesamiento de datos en IA, ya que los datos se representan en forma de vectores y matrices.

4.1 Vectores y Operaciones

Un vector es una lista ordenada de números. Por ejemplo, un vector en $\mathbb{R}^3$ se puede escribir como:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

Suma y Producto por Escalar

- **Suma de vectores:**

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{pmatrix}$$

- **Producto por un escalar c:**

$$c\mathbf{v} = \begin{pmatrix} c \cdot v_1 \\ c \cdot v_2 \\ c \cdot v_3 \end{pmatrix}$$

Ejemplo 4.1:

Sean $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ y $\mathbf{v} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$.

- Suma:

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} 1+4 \\ 2+5 \\ 3+6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix}$$

- Producto por escalar (por 2):

$$2\mathbf{u} = \begin{pmatrix} 2 \cdot 1 \\ 2 \cdot 2 \\ 2 \cdot 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$$

4.2 Matrices y Multiplicación

Una matriz es una colección de números organizados en filas y columnas. Por ejemplo, una matriz A de dimensión $m \times n$ es:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Multiplicación de Matrices

Si A es de dimensión $m \times n$ y B es $n \times p$, el producto $C = A \cdot B$ es una matriz $m \times p$ definida como:

$$c_{ij} = \sum_{k=1}^{n} a_{ik}\, b_{kj}$$

Ejemplo 4.2:

Sean

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

El producto $C = A \cdot B$ es:

$$c_{11} = 1 \cdot 5 + 2 \cdot 7 = 5 + 14 = 19$$

$$c_{12} = 1 \cdot 6 + 2 \cdot 8 = 6 + 16 = 22$$

$$c_{21} = 3 \cdot 5 + 4 \cdot 7 = 15 + 28 = 43$$

$$c_{22} = 3 \cdot 6 + 4 \cdot 8 = 18 + 32 = 50$$

Entonces,

$$C = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

4.3 Eigenvalores y Eigenvectores

En muchos algoritmos de IA (por ejemplo, en reducción de dimensionalidad o análisis de componentes principales), se utilizan los eigenvalores y eigenvectores.

Definición

Para una matriz cuadrada A, un vector no nulo $\mathbf{v}$ es un eigenvector si existe un escalar λ (eigenvalor) tal que:

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

Para encontrar λ, se resuelve el **polinomio característico**:

$$\det(A - \lambda I) = 0$$

donde I es la matriz identidad.

Ejemplo 4.3:

Sea

$$A = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$$

El polinomio característico es:

$$\det\left(\begin{pmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix}\right) = (4-\lambda)(3-\lambda) - 2\cdot 1 = \lambda^2 - 7\lambda + 10 = 0$$

Resolviendo:

$$\lambda^2 - 7\lambda + 10 = 0 \quad \Rightarrow \quad (\lambda-5)(\lambda-2) = 0$$

Por lo tanto, los eigenvalores son $\lambda = 5$ y $\lambda = 2$.

Para $\lambda = 5$:

$$(A - 5I)\mathbf{v} = \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix} \mathbf{v} = \mathbf{0}$$

Una solución es $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ (después de normalizar si se requiere).

Recursos adicionales

1. **"Mathematics for Machine Learning"** por Marc Peter Deisenroth, A. Aldo Faisal, y Cheng Soon Ong.
   - Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.
   - Este libro es una excelente introducción a los conceptos matemáticos necesarios para el aprendizaje automático. Cubre álgebra lineal, cálculo, probabilidad, estadística y optimización, con un enfoque en cómo se aplican estos temas en la IA. Es muy completo y tiene muchos ejemplos. Es, en mi opinión, la mejor opción de las tres que te presento.
2. **"The Elements of Statistical Learning: Data Mining, Inference, and Prediction"** por Trevor Hastie, Robert Tibshirani, y Jerome Friedman.
   - Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
   - Aunque se centra más en el aprendizaje estadístico, este libro es una referencia fundamental. Explica conceptos clave como la regresión, la clasificación, los métodos de regularización, y los modelos lineales generalizados. Es un clásico, pero puede ser un poco más avanzado. El libro se puede descargar gratuitamente desde la página web de los autores.
3. **"Linear Algebra and Its Applications"** por Gilbert Strang.
   - Strang, G. (2016). *Linear algebra and its applications* (5th ed.). Wellesley-Cambridge Press.
   - Gilbert Strang es reconocido por su habilidad para explicar el álgebra lineal de manera clara y accesible. Este libro es un recurso excelente para dominar los conceptos de vectores, matrices, transformaciones lineales, valores propios y más. Incluye muchas aplicaciones prácticas.

Recursos en Línea Gratuitos:

1. **Khan Academy:**
   - https://www.khanacademy.org/
   - Khan Academy ofrece cursos gratuitos en línea sobre álgebra lineal, cálculo, probabilidad y estadística. Los videos son muy didácticos y hay muchos ejercicios para practicar. Es ideal para repasar conceptos o aprender desde cero. Los temas individuales son:

    - Álgebra lineal: https://www.khanacademy.org/math/linear-algebra
    - Cálculo: https://www.khanacademy.org/math/calculus-1 (y cursos de cálculo más avanzados)
    - Probabilidad y estadística: https://www.khanacademy.org/math/statistics-probability
2. **MIT OpenCourseWare:**
    - https://ocw.mit.edu/
    - El MIT ofrece acceso gratuito a los materiales de muchos de sus cursos, incluyendo cursos de matemáticas relevantes para la IA. Puedes encontrar clases magistrales en video, notas de clase y problemas resueltos. Algunos cursos relevantes:
        - 18.06 Linear Algebra (Álgebra Lineal): https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/ (impartido por Gilbert Strang)
        - 6.041 Probabilistic Systems Analysis and Applied Probability (Análisis de Sistemas Probabilísticos y Probabilidad Aplicada): https://ocw.mit.edu/courses/6-041-probabilistic-systems-analysis-and-applied-probability-fall-2010/
        - 18.065 Matrix Methods in Data Analysis, Signal Processing, and Machine Learning (Métodos Matriciales en Análisis de Datos, Procesamiento de Señales y Aprendizaje Automático): https://ocw.mit.edu/courses/18-065-matrix-methods-in-data-analysis-signal-processing-and-machine-learning-spring-2018/
3. **3Blue1Brown**
    - https://www.youtube.com/c/3blue1brown
    - Es un canal de youtube que realiza animaciones de muy alta calidad sobre temas complejos de matematicas, donde explica de manera sumamente visual, algebra lineal y calculo.