



FACULTAD DE INGENIERIA
ASIGNATURA: MACHINE LEARNING

UNIDAD DE APRENDIZAJE 1. INTRODUCCIÓN AL MACHINE LEARNING

Contenido

1. Fundamentos del Machine Learning.....	1
2. Aprendizaje Supervisado.....	4
3. Aprendizaje No Supervisado	6
4. Regresión Lineal como modelo base.....	8
5. Función de pérdida / costo / error	11
6. Descenso del Gradiente (Gradient Descent)	14
7. Tasa de Aprendizaje (Learning Rate).....	16

1. Fundamentos del Machine Learning

¿Qué es el aprendizaje automático (Machine Learning)?

El **Machine Learning (ML)** funciona de manera similar. Es una rama de la Inteligencia Artificial que permite a los sistemas informáticos "aprender" de los datos sin ser programados explícitamente para cada tarea. En lugar de escribir líneas de código para cada posible escenario, alimentamos un algoritmo con grandes cantidades de datos. El algoritmo encuentra patrones y relaciones dentro de esos datos, lo que le permite hacer predicciones o tomar decisiones sobre datos nuevos.

Analogía 1: Enseñar a una máquina con datos es como enseñar a un niño a reconocer frutas mostrándole muchas fotos de manzanas, bananas y naranjas. Con el tiempo, el niño aprende a identificarlas por su forma y color.

Analogía 2: Imagina que quieres enseñarle a un niño a reconocer perros. No le das una lista de reglas exactas ("si tiene cuatro patas y cola, es un perro"). En cambio, le muestras muchas fotos de perros y no perros, y con el tiempo, el niño aprende a identificar las características comunes de un perro.



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Diferencia entre programación tradicional y aprendizaje automático

Característica	Programación Tradicional	Aprendizaje Automático (ML)
Enfoque	Reglas explícitas, lógicas y algoritmos definidos por el humano.	Aprende patrones de los datos, genera sus propias "reglas".
Entrada	Datos y reglas.	Datos y respuestas esperadas (en algunos casos).
Salida	Respuestas.	Reglas o un modelo que puede generar respuestas.
Modificación	Requiere cambios manuales en el código para nuevas reglas.	Mejora con más datos o ajustes al modelo.
Complejidad	Maneja problemas bien definidos con lógica clara.	Ideal para problemas complejos donde las reglas son difíciles de definir.

Ejemplo:

Programación Tradicional: Si quieres calcular el impuesto sobre las ventas, escribes una función: $\text{Impuesto} = \text{Precio} * \text{TasaDelImpuesto}$. Las reglas son fijas y explícitas.

Machine Learning: Si quieres predecir si un correo electrónico es spam, sería casi imposible escribir todas las reglas que definen el spam. En ML, alimentas un algoritmo con miles de correos electrónicos (algunos marcados como spam, otros no), y el algoritmo aprende a distinguir el spam por sí mismo, basándose en patrones en el texto, el remitente, etc.

Enfoque	Entrada	Salida
Programación tradicional	Datos + Reglas	Resultado
Machine Learning	Datos + Resultados esperados	Reglas (modelo aprendido)

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Importancia y aplicaciones actuales del Machine Learning

El ML es fundamental en nuestra vida diaria, a menudo sin que nos demos cuenta. Su capacidad para procesar vastas cantidades de datos y descubrir insights lo hace invaluable en casi todos los sectores.

- **Salud:**
 - **Diagnóstico de enfermedades:** Analiza imágenes médicas (radiografías, resonancias) para detectar tumores o anomalías.
 - **Descubrimiento de fármacos:** Predice la efectividad de nuevas moléculas.
 - **Medicina personalizada:** Adapta tratamientos según el perfil genético del paciente.
- **Industria:**
 - **Mantenimiento predictivo:** Prevé fallos en maquinaria para realizar mantenimiento antes de que ocurran.
 - **Control de calidad:** Inspecciona productos para detectar defectos automáticamente.
 - **Optimización de la cadena de suministro:** Predice la demanda para gestionar inventarios.
- **Educación:**
 - **Sistemas de recomendación de contenido:** Sugiere cursos o materiales de estudio personalizados.
 - **Evaluación automática:** Califica ensayos o exámenes.
 - **Detección de plagio:** Identifica similitudes entre textos.
- **Finanzas:**
 - **Detección de fraude:** Identifica transacciones sospechosas con tarjetas de crédito.
 - **Análisis de riesgo crediticio:** Evalúa la probabilidad de que un cliente pague un préstamo.
 - **Trading algorítmico:** Realiza operaciones bursátiles automáticamente basándose en predicciones de mercado.
- **Marketing y Ventas:**
 - **Sistemas de recomendación:** Sugiere productos en línea (Amazon, Netflix).

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

- **Segmentación de clientes:** Agrupa clientes con comportamientos similares para campañas dirigidas.
- **Transporte:**
 - **Vehículos autónomos:** Perciben el entorno, toman decisiones de navegación y evitan obstáculos.
 - **Optimización de rutas:** Planifica las rutas más eficientes para el reparto.

Tipos de Aprendizaje en ML: Supervisado, No Supervisado y por Refuerzo

Existen tres categorías principales de aprendizaje en Machine Learning, diferenciadas por el tipo de datos que usan y el objetivo del aprendizaje:

- **Aprendizaje Supervisado:** El modelo aprende de datos que ya tienen la respuesta correcta (etiquetas). Piensa en un maestro que te da ejemplos con las soluciones.
- **Aprendizaje No Supervisado:** El modelo explora datos sin ninguna etiqueta o respuesta predefinida. Su objetivo es encontrar patrones ocultos, estructuras o relaciones por sí mismo. Aquí no hay "maestro", el modelo aprende por descubrimiento.
- **Aprendizaje por Refuerzo:** (Mención breve) Es un tipo de aprendizaje donde un agente aprende a tomar decisiones interactuando con un entorno. Recibe recompensas o penalizaciones por sus acciones, y su objetivo es maximizar la recompensa acumulada. Es como enseñar a un perro trucos dándole premios.

2. Aprendizaje Supervisado

Definición y características principales

El **Aprendizaje Supervisado** es el tipo más común de Machine Learning. Su nombre "supervisado" viene de la idea de que el proceso de aprendizaje es guiado o "supervisado" por un conjunto de datos que ya contiene las **respuestas correctas** (o etiquetas).

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Imagina que eres un estudiante y te dan un examen con las preguntas y las respuestas correctas. Tu tarea es aprender la relación entre las preguntas y las respuestas para poder responder preguntas nuevas correctamente. En ML supervisado, el modelo hace exactamente eso: aprende un mapeo entre un conjunto de **variables de entrada** (las "preguntas") y una **variable de salida** (las "respuestas").

Características clave:

- **Requiere datos etiquetados:** Cada ejemplo en el conjunto de datos de entrenamiento debe tener su correspondiente valor de salida conocido.
- El objetivo es **predecir** una variable de salida basándose en las variables de entrada.
- Se divide en dos categorías principales: **regresión** y **clasificación**.

Variables de Entrada (Features) y Variable de Salida (Target)

- **Features (X):** características del problema (ej: número de habitaciones).
- **Target (y):** resultado esperado (ej: precio de una casa).

Ejemplos comunes

- **Regresión:** predecir un valor continuo (precio, temperatura).
- **Clasificación:** predecir una clase (aprobado vs. reprobado).

Flujo básico del aprendizaje supervisado

El proceso de construir un modelo de aprendizaje supervisado generalmente sigue estos pasos:

Entrenamiento:

- **Preparación de datos:** Recopilar y limpiar un conjunto de datos con features y targets.
- **División de datos:** Dividir el conjunto de datos en entrenamiento (la mayoría de los datos, para que el modelo aprenda) y prueba (una porción más pequeña, para evaluar qué tan bien aprendió).

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- **Aprendizaje del modelo:** El algoritmo de ML procesa los datos de entrenamiento para encontrar patrones y construir un "modelo". Es como el estudiante que estudia los ejemplos resueltos.

Evaluación:

- Una vez entrenado, el modelo se prueba con el conjunto de datos de prueba, que son datos que nunca ha visto antes.
- Se comparan las predicciones del modelo con las respuestas correctas reales en el conjunto de prueba.
- Se utilizan métricas de evaluación (como la precisión, el error, etc.) para determinar qué tan bien se desempeña el modelo. Esto es crucial para saber si el modelo aprendió bien.

Predicción:

- Si el modelo se evalúa bien, se considera "listo" para usar.
- Ahora, puedes alimentarle nuevos datos no vistos (sin etiquetas) y el modelo usará lo que aprendió para hacer predicciones. Por ejemplo, le das los detalles de una casa nueva y te predice su precio.

Analogía: Piénsalo como un chef (el modelo) aprendiendo a cocinar (entrenamiento). Le das muchas recetas con ingredientes (features) y el plato final (target). Él prueba las recetas y las ajusta hasta que el plato sabe bien. Luego, lo pruebas con un plato nuevo (evaluación). Si pasa la prueba, está listo para cocinar platos para clientes (predicción).

3. Aprendizaje No Supervisado

Definición y características principales

El **Aprendizaje No Supervisado** es una rama del Machine Learning donde el modelo trabaja con datos que **no tienen etiquetas o respuestas predefinidas**. A diferencia del aprendizaje supervisado, aquí no hay un "maestro" que le diga al modelo cuál es la respuesta correcta.



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Imagina que te entregan una caja llena de juguetes mezclados: bloques, carros, muñecas, peluches. No te dicen qué es cada uno, pero tu tarea es organizarlos de alguna manera. Podrías empezar a agrupar los juguetes por color, por forma, o por tipo de material. El aprendizaje no supervisado hace algo similar: descubre **patrones, estructuras o agrupaciones ocultas** dentro de los datos por sí mismo.

Características clave:

- Trabaja con **datos no etiquetados**.
- El objetivo principal es **descubrir la estructura subyacente** en los datos.
- No hay una variable de salida o "target" definido.
- Útil para exploración de datos, compresión de datos y para preparar datos para el aprendizaje supervisado.

Objetivos comunes:

- **Agrupar datos similares (clustering).**
- **Reducir la cantidad de variables (dimensionalidad).**

Ejemplos Comunes: Clustering (Agrupamiento), Reducción de Dimensionalidad

Clustering (Agrupamiento): Es la tarea de agrupar puntos de datos similares en "clusters" o grupos. Los puntos de datos dentro de un mismo grupo son más similares entre sí que con los puntos de otros grupos.

- **Ejemplos:**
 - **Segmentación de clientes:** Agrupar a los clientes de una empresa en diferentes segmentos (ej. "clientes de alto valor", "clientes sensibles al precio") basándose en su historial de compras y comportamiento.
 - **Agrupamiento de noticias:** Organizar artículos de noticias por temas (ej. política, deportes, tecnología) sin que alguien los haya etiquetado previamente.
 - **Detección de comunidades en redes sociales:** Identificar grupos de usuarios con intereses o conexiones similares.
- **Algoritmos comunes:** K-Means, DBSCAN, Agrupamiento Jerárquico.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Reducción de Dimensionalidad: Implica reducir el número de variables (features) en un conjunto de datos manteniendo la mayor cantidad de información posible. Esto es útil cuando tienes muchos features y algunos pueden ser redundantes o no informativos, lo que puede dificultar el análisis o el rendimiento del modelo.

- **Ejemplos:**
 - **Compresión de imágenes:** Reducir el tamaño de un archivo de imagen eliminando información menos relevante sin afectar drásticamente la calidad visual.
 - **Visualización de datos:** Cuando tienes datos con muchas dimensiones, es difícil visualizarlos. La reducción de dimensionalidad permite proyectar los datos en 2 o 3 dimensiones para poder graficarlos y entender su estructura.
 - **Preprocesamiento de datos:** Eliminar ruido o multicolinealidad antes de usar un algoritmo de aprendizaje supervisado, lo que puede mejorar su rendimiento.
- **Algoritmos comunes:** Análisis de Componentes Principales (PCA), t-SNE.

Analogía: Piensa en un bibliotecario (el algoritmo) al que se le entregan miles de libros nuevos (datos) sin ninguna etiqueta ni categorización. En lugar de saber de antemano si un libro es "ciencia ficción" o "biografía", el bibliotecario examina los libros (los datos) por su cuenta, busca patrones en el lenguaje, los temas y las palabras clave, y luego decide cómo agruparlos para que los usuarios puedan encontrarlos fácilmente.

4. Regresión Lineal como modelo base

La **Regresión Lineal** es uno de los algoritmos más fundamentales y ampliamente utilizados en Machine Learning supervisado, específicamente para problemas de **regresión**. Su objetivo es modelar la relación entre una variable dependiente (la que queremos predecir, el **target**) y una o más variables independientes (los **features**) ajustando una **línea recta** (o un plano, o un hiperplano en más dimensiones) a los datos.

En esencia, la regresión lineal busca encontrar la línea que mejor se ajusta a los puntos de datos, de modo que pueda predecir el valor del target para nuevos datos.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

¿Para qué se usa?

- **Predicción continua:** Predice valores numéricos, como precios de viviendas, calificaciones de estudiantes, ventas de productos, etc.
- **Identificación de relaciones:** Ayuda a entender cómo el cambio en una o más variables de entrada afecta la variable de salida. Por ejemplo, ¿cómo influye el tamaño de una casa en su precio?
- **Análisis explicativo:** Puede usarse para determinar qué features son más significativos para predecir el target.

Ecuación General de una Regresión Lineal

La regresión lineal simple (con una sola variable de entrada) se expresa con una ecuación familiar:

$$y = mx + b$$

Donde:

- y: Es la variable que queremos predecir (el **target**).
- m: Es la **pendiente** de la línea. Nos dice cuánto cambia y por cada unidad de cambio en x. En ML, a menudo se le llama **peso** o **coeficiente** (w_1).
- x: Es la variable de entrada (el **feature**).
- b: Es la **intersección con el eje y** (donde la línea cruza el eje y cuando x es 0). En ML, se le conoce como **sesgo** (bias) o intercepto (w_0).

Para múltiples variables de entrada (Regresión Lineal Múltiple), la ecuación se generaliza a:

$$y = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$$

Donde:

- y: La variable de salida predicha.
- w_0 : El **sesgo (bias)**, o la intercepción con el eje y. Es el valor de y cuando todos los x_i son cero.
- w_1, w_2, \dots, w_n : Son los **pesos (coeficientes)** para cada variable de entrada (x_1, x_2, \dots, x_n). Cada w_i representa la importancia y dirección de la relación de la variable x_i con y.
- x_1, x_2, \dots, x_n : Son las múltiples **variables de entrada (features)**.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Analogía: Piensa en la relación entre el tiempo que estudias para un examen (x) y tu calificación (y). Una regresión lineal intentaría dibujar una línea que represente esa relación. La pendiente (m) sería cuánto mejora tu calificación por cada hora extra de estudio, y el intercepto (b) sería la calificación que obtendrías si no estudiaras nada.

Supuestos del Modelo Lineal

Para que una regresión lineal funcione de manera óptima y sus resultados sean confiables, se asumen ciertas condiciones en los datos. Si estas no se cumplen, el modelo aún puede funcionar, pero sus predicciones y la interpretación de sus coeficientes podrían no ser tan precisas o válidas.

- **Linealidad:** La relación entre las variables de entrada y la variable de salida debe ser lineal. Es decir, los puntos de datos deben agruparse aproximadamente alrededor de una línea recta.
- **Independencia de los residuos:** Los errores (la diferencia entre la predicción y el valor real) deben ser independientes entre sí. No debe haber un patrón en los errores.
- **Homocedasticidad:** La varianza de los residuos debe ser constante en todos los niveles de las variables de entrada. Gráficamente, los errores no deben "abrirse" o "cerrarse" como un embudo.
- **Normalidad de los residuos:** Los errores deben seguir una distribución normal.
- **No multicolinealidad:** Las variables de entrada no deben estar altamente correlacionadas entre sí. Si dos features son casi idénticos, puede dificultar que el modelo determine la contribución individual de cada uno.

Interpretación de Coeficientes y el Sesgo (Bias)

- **Coeficientes (w_i o m):**
 - Un coeficiente positivo significa que a medida que esa variable de entrada aumenta, la variable de salida también tiende a aumentar (manteniendo las otras variables constantes).
 - Un coeficiente negativo significa que a medida que esa variable de entrada aumenta, la variable de salida tiende a disminuir.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

- La **magnitud** del coeficiente indica la fuerza de la relación. Un coeficiente grande (en valor absoluto) sugiere una influencia más fuerte.
- **Ejemplo:** En un modelo de precio de vivienda, si el coeficiente para "número de habitaciones" es \$50,000, significa que, por cada habitación adicional, el precio de la casa se predice que aumentará en \$50,000 (asumiendo que los demás factores no cambian).
- **Sesgo (w_0 o b - Intercepto):**
 - Representa el valor predicho de la variable de salida cuando **todas las variables de entrada son cero**.
 - A menudo, no tiene una interpretación práctica directa si el cero para las variables de entrada no tiene sentido en el contexto real (ej. ¿cuál es el precio de una casa con 0 habitaciones y 0 metros cuadrados?).
 - Sin embargo, es crucial para ajustar la línea de regresión para que pase por la ubicación correcta en el espacio de datos.

5. Función de pérdida / costo / error

¿Por qué Necesitamos una Función de Pérdida?

Cuando entrenamos un modelo de Machine Learning, especialmente en el aprendizaje supervisado, el objetivo es que el modelo haga predicciones lo más cercanas posible a los valores reales. Pero ¿cómo sabemos si el modelo está haciendo un buen trabajo? ¿Cómo cuantificamos qué tan "mal" se equivoca? Aquí es donde entra la **Función de Pérdida** (también conocida como **Función de Costo** o **Función de Error**). Es una medida numérica de la diferencia entre el valor que nuestro modelo **predijo** y el **valor real** que debería haber predicho.

Piensa en ella como un "sistema de puntuación" para la equivocación del modelo. Una puntuación alta significa que el modelo se está equivocando mucho, mientras que una puntuación baja (cercana a cero) indica que el modelo está haciendo predicciones muy precisas. Durante el entrenamiento, el objetivo del algoritmo es **minimizar esta función de pérdida**.





CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Diferencias entre Función de Pérdida y Función de Costo

Aunque a menudo se usan indistintamente, hay una pequeña distinción conceptual:

- **Función de Pérdida (Loss Function):** Mide la penalización para **una sola instancia** (un solo punto de dato). Es decir, calcula el error de la predicción de nuestro modelo para un *único ejemplo*.
 - **Ejemplo:** ¿Cuál es el error cuando predije \$100,000 para una casa que realmente valía \$110,000?
- **Función de Costo (Cost Function):** Mide la penalización promedio (o total) para **todo el conjunto de entrenamiento** (o un lote de datos). Es el promedio de las funciones de pérdida para todas las instancias. Esta es la función que realmente tratamos de minimizar durante el entrenamiento del modelo.
 - **Ejemplo:** ¿Cuál es el error promedio de mi modelo al predecir los precios de todas las casas en mi conjunto de datos de entrenamiento?

En la práctica, cuando hablamos de "minimizar la función de costo" o "función de pérdida", a menudo nos referimos a la función de costo que se calcula sobre todo el conjunto de datos para guiar la optimización.

Ejemplos: Error Cuadrático Medio (MSE), Error Absoluto Medio (MAE)

Aquí hay dos de las funciones de pérdida/costo más comunes para problemas de regresión:

1. Error Cuadrático Medio (MSE - Mean Squared Error):

- **Fórmula:** $MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$
 - N: Número total de instancias (ejemplos) en el conjunto de datos.
 - y_i : El valor real (target) para la instancia i.
 - \hat{y}_i : El valor predicho por el modelo para la instancia i.
- **Concepto:** Calcula la diferencia entre cada predicción y el valor real, eleva al cuadrado esa diferencia (para que los errores negativos no se cancelen con los positivos y para penalizar más los errores grandes), y luego promedia todos esos errores cuadrados.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

- **Ventajas:** Penaliza fuertemente los errores grandes, lo que puede ser útil en algunos contextos. Es diferenciable, lo que facilita el uso de algoritmos de optimización como el Descenso del Gradiente.
- **Desventajas:** Es sensible a los valores atípicos (outliers) debido al término cuadrado, lo que puede "engañar" al modelo para que se ajuste a estos valores extremos.

2. Error Absoluto Medio (MAE - Mean Absolute Error):

- **Fórmula:** $MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$
- **Concepto:** Calcula la diferencia absoluta entre cada predicción y el valor real, y luego promedia todos esos valores absolutos.
- **Ventajas:** Menos sensible a los valores atípicos que el MSE porque no eleva al cuadrado los errores. Es más interpretable porque el error está en la misma unidad que la variable de salida.
- **Desventajas:** No es diferenciable en el cero, lo que puede complicar un poco la optimización en ciertos escenarios (aunque los optimizadores modernos suelen manejarlo bien).

Visualización de la Función de Pérdida como Superficie

Imagina que quieres encontrar el punto más bajo en un valle. La **función de pérdida** es como la topografía de ese valle. Para un modelo simple como la regresión lineal con un solo feature, la función de costo (MSE, por ejemplo) es una parábola (una curva en forma de U) si graficamos el error contra el peso (m) o el sesgo (b) del modelo. Si tenemos dos parámetros (m y b), la función de costo es una superficie tridimensional con forma de cuenco.

El objetivo del entrenamiento es "descender" por esta superficie hasta encontrar el punto más bajo, que corresponde a los valores de los parámetros del modelo (pesos y sesgos) que minimizan el error y, por lo tanto, hacen que el modelo sea lo más preciso posible. El **Descenso del Gradiente** es el algoritmo que nos ayuda a encontrar ese punto más bajo.





6. Descenso del Gradiente (Gradient Descent)

Concepto General: Algoritmo de Optimización

El **Descenso del Gradiente (Gradient Descent)** es un algoritmo de optimización fundamental en Machine Learning, especialmente para entrenar modelos. Su propósito es encontrar el conjunto de parámetros (pesos y sesgos) de un modelo que **minimice la función de costo**.

Imagina que estás en la cima de una montaña muy nebulosa y quieres llegar al punto más bajo del valle (el mínimo de la función de costo). No puedes ver el valle completo, solo el terreno justo debajo de tus pies. ¿Qué haces? Das un pequeño paso en la dirección más inclinada hacia abajo. Repites este proceso: evalúas la pendiente, das otro pequeño paso hacia abajo, y así sucesivamente, hasta que ya no puedas bajar más.

El Descenso del Gradiente hace exactamente esto. Itera, ajustando los parámetros del modelo en la dirección opuesta al gradiente de la función de costo con respecto a esos parámetros.

¿Cómo Encuentra el Mínimo de la Función de Costo?

1. **Paso Inicial:** Comienza con valores aleatorios para los parámetros del modelo (pesos w y sesgo b). Esto es como estar en un punto aleatorio de la montaña.
2. **Cálculo del Gradiente:** En cada paso, el algoritmo calcula el **gradiente** de la función de costo con respecto a cada uno de los parámetros del modelo. El gradiente es un vector que apunta en la dirección de mayor aumento de la función.
3. **Ajuste de Parámetros:** Para minimizar la función de costo (es decir, "descender" en la montaña), los parámetros se actualizan moviéndose en la dirección **opuesta** al gradiente.
 - Cada parámetro se ajusta restando el gradiente multiplicado por un pequeño valor llamado **tasa de aprendizaje (learning rate)**.

$$\bullet \quad w_{nuevo} = w_{viejo} - \text{tasa_aprendizaje} \times \frac{\partial \text{Costo}}{\partial w_{viejo}}$$

$$\bullet \quad b_{nuevo} = b_{viejo} - \text{tasa_aprendizaje} \times \frac{\partial \text{Costo}}{\partial b_{viejo}}$$

○

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

4. **Iteración:** Los pasos 2 y 3 se repiten muchas veces (iteraciones o épocas) hasta que la función de costo deja de disminuir significativamente, lo que indica que se ha alcanzado un mínimo (local o global).

Derivada y Gradiente: Intuición Geométrica

- **Derivada:** Para una función de una sola variable, la derivada en un punto nos dice la **pendiente** de la curva en ese punto. Si la pendiente es positiva, la función está subiendo; si es negativa, está bajando. Queremos movernos en la dirección opuesta a la pendiente para encontrar un mínimo.
- **Gradiente:** Es la generalización de la derivada para funciones de múltiples variables. El gradiente es un vector que apunta en la dirección de la **mayor tasa de aumento** de la función. Por lo tanto, para minimizar la función, nos movemos en la dirección **opuesta al gradiente**.

Analogía: Si estás con los ojos vendados en la montaña y solo puedes sentir el terreno directamente debajo de tus pies:

- La **pendiente** (derivada) te dice si la tierra está inclinada hacia arriba o hacia abajo en la dirección en la que estás mirando.
- El **gradiente** te dice en qué dirección la pendiente es más pronunciada hacia arriba. Para bajar más rápido, debes moverte en la dirección exactamente opuesta.

Relación con el Entrenamiento del Modelo

El Descenso del Gradiente es el corazón del proceso de entrenamiento para muchos modelos de Machine Learning. Es la técnica que el modelo utiliza para **aprender los mejores valores para sus parámetros** (pesos y sesgos) a partir de los datos de entrenamiento. Al minimizar la función de costo, el modelo se vuelve más preciso en sus predicciones.

Sin un algoritmo de optimización como el Descenso del Gradiente, un modelo simplemente no sabría cómo ajustar sus parámetros para mejorar su rendimiento basándose en la información que le proporciona la función de costo.

Variantes del Descenso del Gradiente:

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

- **Batch Gradient Descent:** Calcula el gradiente usando *todos* los ejemplos del conjunto de entrenamiento en cada paso. Es lento para grandes datasets pero garantiza un descenso suave.
- **Stochastic Gradient Descent (SGD):** Calcula el gradiente y actualiza los parámetros usando *un solo ejemplo* de entrenamiento en cada paso. Es mucho más rápido y puede escapar de mínimos locales, pero el descenso es más ruidoso.
- **Mini-Batch Gradient Descent:** Un compromiso entre los dos anteriores, usando un *pequeño lote* de ejemplos de entrenamiento en cada paso. Es la más usada en la práctica.

7. Tasa de Aprendizaje (Learning Rate)

¿Qué es la Tasa de Aprendizaje?

La **Tasa de Aprendizaje (Learning Rate)** es uno de los hiperparámetros más críticos en el Descenso del Gradiente (y otros algoritmos de optimización). Es un número pequeño y positivo que determina el **tamaño del paso** que toma el algoritmo en cada iteración mientras desciende por la función de costo.

Volviendo a la analogía de la montaña con los ojos vendados: la tasa de aprendizaje es como la **longitud de cada paso** que das.

- Un valor de tasa de aprendizaje de 0.01 significa que darás un paso del 1% del tamaño del gradiente.
- Un valor de 0.1 significa un paso del 10%.

Efecto de una Tasa de Aprendizaje muy Alta vs. muy Baja

La elección de la tasa de aprendizaje tiene un impacto drástico en el proceso de entrenamiento y en la convergencia del modelo:

- **Tasa de Aprendizaje muy Alta:**
 - **Problema:** El algoritmo puede "saltar" el mínimo de la función de costo. Es como dar pasos tan grandes que saltas el punto más bajo del valle y terminas en el otro lado, o incluso subiendo por la ladera opuesta.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

- **Consecuencia:** La función de costo puede divergir (aumentar en lugar de disminuir) o oscilar salvajemente, lo que significa que el modelo nunca converge o nunca encuentra un buen conjunto de parámetros. El modelo podría volverse inestable.
- **Tasa de Aprendizaje muy Baja:**
 - **Problema:** El algoritmo toma pasos muy pequeños, lo que significa que le tomará **mucho tiempo** (muchas iteraciones) llegar al mínimo. Es como dar pasos de hormiga en la montaña.
 - **Consecuencia:** El entrenamiento es excesivamente lento. Aunque eventualmente podría converger a un buen mínimo, el proceso puede ser ineficiente y computacionalmente costoso, especialmente para grandes conjuntos de datos. También podría quedarse atrapado en un mínimo local antes de alcanzar el mínimo global si la superficie es compleja.

Visualización del Avance del Descenso del Gradiente con Diferentes Tasas

Imagina la función de costo como un cuenco.

- **Tasa Alta:** Tu bola rodaría demasiado rápido y saltaría el fondo del cuenco, rebotando de un lado a otro o saliéndose.
- **Tasa Baja:** Tu bola rodaría muy, muy lento hacia el fondo, tomando una cantidad excesiva de tiempo.
- **Tasa Óptima:** Tu bola rodaría de manera eficiente, haciendo progresos constantes y llegando al fondo del cuenco de forma razonable.

Elección de la Tasa de Aprendizaje como Hiperparámetro Clave

La tasa de aprendizaje es un **hiperparámetro**, lo que significa que no es un parámetro que el modelo aprende por sí mismo de los datos (como los pesos y sesgos), sino que es un valor que debemos **definir o "ajustar"** antes de que comience el entrenamiento.

Encontrar una buena tasa de aprendizaje es crucial para un entrenamiento efectivo y eficiente. Generalmente se hace mediante:

📍 Sede Quirinal: Calle 21 No. 6 - 01
📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

- **Experimentación:** Probar diferentes valores (ej. 0.1, 0.01, 0.001, 0.0001) y observar cómo se comporta la función de costo.
- **Grid Search o Random Search:** Técnicas automatizadas que prueban sistemática o aleatoriamente diferentes combinaciones de hiperparámetros.
- **Programas de tasa de aprendizaje:** Estrategias donde la tasa de aprendizaje se ajusta dinámicamente durante el entrenamiento (ej. disminuye con el tiempo) para permitir pasos grandes al inicio y pasos más finos a medida que se acerca al mínimo.

Una tasa de aprendizaje bien elegida permite que el Descenso del Gradiente converja de manera eficiente y estable hacia el mínimo de la función de costo, lo que resulta en un modelo bien entrenado y preciso.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"