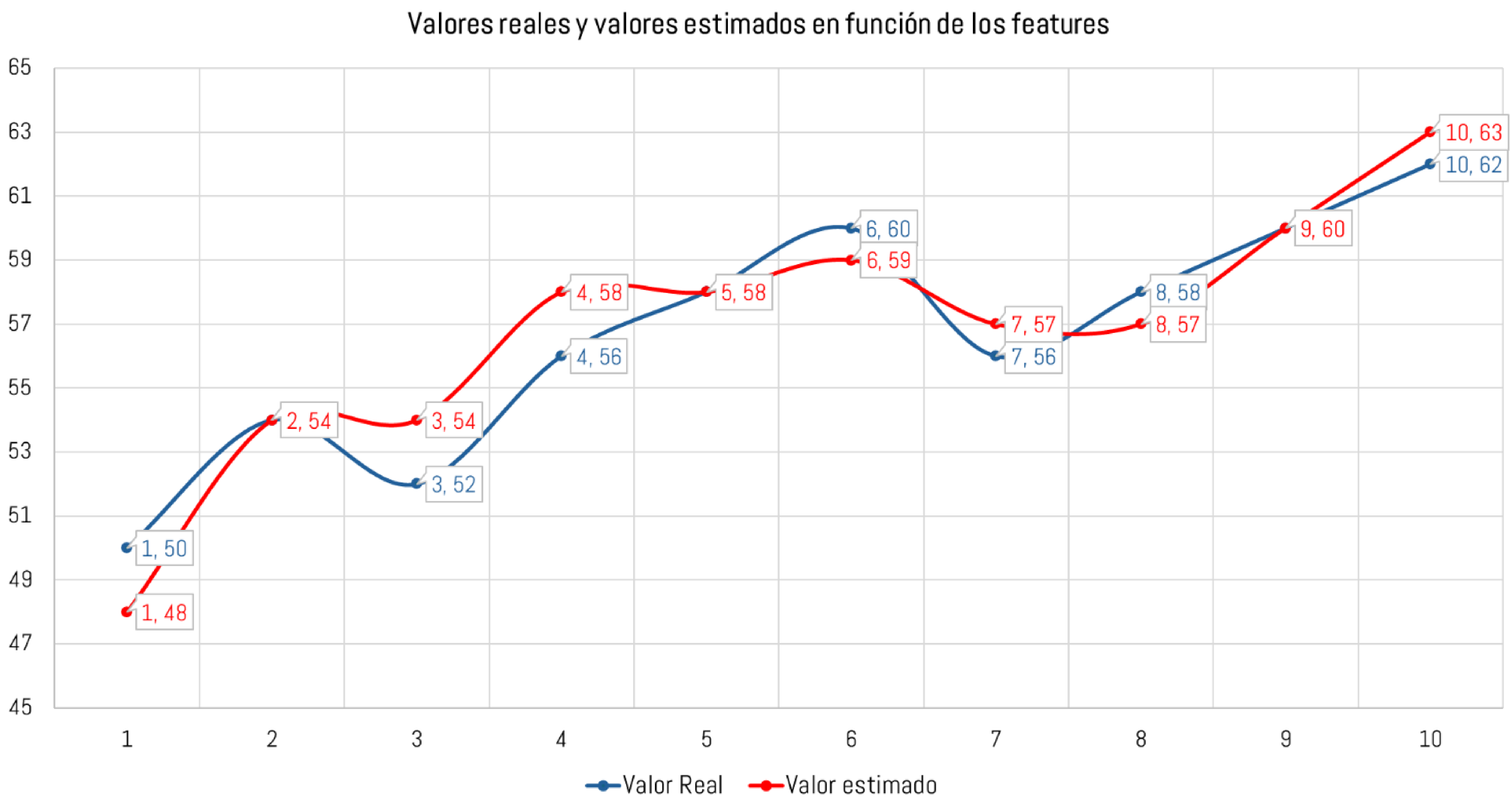
Universidad Tecnológica de Panamá
Facultad de Ingeniería Eléctrica
Maestría en Ingeniería Eléctrica

Semana 3

Introducción a Machine Learning

Métricas de evaluación para algoritmos de regresión

- En los problemas de regresión, **el objetivo es predecir un valor numérico continuo** a partir de un conjunto de variables independientes.
- Para evaluar el desempeño de los modelos de regresión, se utilizan diferentes métricas que permiten cuantificar el error entre los valores predichos por el modelo y los valores reales observados.
- La elección de la métrica adecuada depende del contexto del problema y de la sensibilidad que se quiera tener ante errores grandes o pequeños.



Valores reales y valores estimados en función de los features
65
63
61
59
57
55
53
51
49
47
45
1, 50
1, 48
2, 54
3, 54
3, 52
4, 58
4, 56
5, 58
6, 60
6, 59
7, 57
7, 56
8, 58
8, 57
9, 60
10, 63
10, 62
1
2
3
4
5
6
7
8
9
10
Valor Real
Valor estimado

Error Cuadrático Medio (MSE)

El **Error Cuadrático Medio** (*Mean Squared Error*, MSE) es una de las métricas más utilizadas para evaluar modelos de regresión. Mide el promedio de los errores al cuadrado entre los valores predichos por el modelo y los valores reales observados.

El **modelo matemático** del MSE es:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

donde y_i es el valor real, $\hat{y}_i$ es el valor predicho por el modelo y n es el número total de observaciones.

Ejemplo de cálculo de MSE

X	Valor Real (y)	Valor Estimado ($\hat{y}$)	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	50	48	2	4
2	54	54	0	0
3	52	54	-2	4
4	56	58	-2	4
5	58	58	0	0
6	60	59	1	1
7	56	57	-1	1
8	58	57	1	1
9	60	60	0	0
10	62	63	-1	1
			$\sum(y_i - \hat{y}_i)^2$	16
			n	10
			$\frac{1}{n}\sum(y_i - \hat{y}_i)^2$	1.6

Interpretación

Un MSE bajo indica que las predicciones del modelo están cerca de los valores reales; un MSE alto señala grandes errores.

Como penaliza más los errores grandes, es útil cuando se quiere evitar desviaciones importantes.

Sin embargo, su valor está en unidades al cuadrado, por lo que suele complementarse con métricas como RMSE o MAE para una interpretación más directa.

El MSE es útil para comparar modelos: el de menor MSE se ajusta mejor a los datos, considerando siempre el contexto del problema.

Raíz del Error Cuadrático Medio (RMSE)

La **Raíz del Error Cuadrático Medio** (*Root Mean Squared Error*, RMSE) es una métrica derivada del MSE que permite interpretar el error en las mismas unidades que la variable objetivo. Es especialmente útil para comprender la magnitud promedio de los errores cometidos por el modelo.

El **modelo matemático** del RMSE es:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

donde y_i es el valor real, $\hat{y}_i$ es el valor predicho por el modelo y n es el número total de observaciones.

Ejemplo de cálculo de RMSE

X	Valor Real (y)	Valor Estimado ($\hat{y}$)	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	50	48	2	4
2	54	54	0	0
3	52	54	-2	4
4	56	58	-2	4
5	58	58	0	0
6	60	59	1	1
7	56	57	-1	1
8	58	57	1	1
9	60	60	0	0
10	62	63	-1	1
			$\sum(y_i - \hat{y}_i)^2$	16
			n	10
			$\frac{1}{n}\sum(y_i - \hat{y}_i)^2$ (MSE)	1.6
			$\sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$ (RMSE)	1.26

Interpretación

El RMSE representa la desviación estándar de los residuos (errores de predicción).

Un valor bajo de RMSE indica que el modelo predice con precisión los valores reales, mientras que un valor alto señala errores significativos.

Al estar en las mismas unidades que la variable objetivo, facilita la interpretación directa del error promedio cometido por el modelo.

Error Absoluto Medio (MAE)

El **Error Absoluto Medio** (*Mean Absolute Error*, MAE) es una métrica que mide el promedio de las diferencias absolutas entre los valores reales y los valores predichos por el modelo. Es una de las métricas más intuitivas para evaluar el desempeño de modelos de regresión, ya que indica cuánto se equivoca el modelo, en promedio, en las mismas unidades que la variable objetivo.

El **modelo matemático** del MAE es:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

donde y_i es el valor real, $\hat{y}_i$ es el valor predicho por el modelo y n es el número total de observaciones.

Ejemplo del cálculo de MAE

X	Valor Real (y)	Valor Estimado ($\hat{y}$)	$y_i - \hat{y}_i$	$\lvert y_i - \hat{y}_i \rvert$
1	50	48	2	2
2	54	54	0	0
3	52	54	-2	2
4	56	58	-2	2
5	58	58	0	0
6	60	59	1	1
7	56	57	-1	1
8	58	57	1	1
9	60	60	0	0
10	62	63	-1	1
			$\sum \lvert y_i - \hat{y}_i \rvert$	10
			n	10
			$\mathrm{MAE} = \frac{1}{n} \sum \lvert y_i - \hat{y}_i \rvert$	1.0

Interpretación

El MAE indica el error promedio absoluto cometido por el modelo en sus predicciones.

Un valor bajo de MAE significa que, en promedio, las predicciones están cerca de los valores reales.

A diferencia del MSE y RMSE, el MAE no penaliza fuertemente los errores grandes, por lo que es menos sensible a valores atípicos (outliers).

Es útil cuando se desea una métrica robusta ante la presencia de errores extremos.

Coeficiente de determinación (R^2)

El **Coeficiente de Determinación**, conocido como R^2, es una métrica que indica la proporción de la varianza de la variable dependiente que es explicada por el modelo de regresión. Es ampliamente utilizado para evaluar el ajuste global de un modelo de regresión.

El **modelo matemático** del R^2 es:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

donde y_i es el valor real, $\hat{y}_i$ es el valor predicho por el modelo, $\bar{y}$ es el promedio de los valores reales y n es el número total de observaciones.

Ejemplo del cálculo de R^2

X	Valor Real (y)	Valor Estimado ($\hat{y}$)	$\bar{y}$	$(y_i - \hat{y}_i)^2$	$(y_i - \bar{y})^2$
1	50	48	56.6	4	43.56
2	54	54	56.6	0	6.76
3	52	54	56.6	4	21.16
4	56	58	56.6	4	0.36
5	58	58	56.6	0	1.96
6	60	59	56.6	1	11.56
7	56	57	56.6	1	0.36
8	58	57	56.6	1	1.96
9	60	60	56.6	0	11.56
10	62	63	56.6	1	29.16
				$\sum(y_i - \hat{y}_i)^2$ (SSE)	16
				$\sum(y_i - \bar{y})^2$ (SST)	128.4
				$\bar{y}$	56.6
				$R^2 = 1 - \frac{SSE}{SST}$	0.8754

Interpretación

El valor de R^2 varía entre $-\infty$ y 1.

- Un R^2 cercano a **1** indica que el modelo explica casi toda la variabilidad de la variable dependiente.
- Un valor cercano a **0** sugiere que el modelo no mejora la predicción con respecto a simplemente usar la media de los valores reales.
- Un R^2 **negativo** puede ocurrir si el modelo predice peor que una estimación constante basada en la media.

Aunque R^2 es útil para **comparar modelos**, **no debe ser la única métrica utilizada**, especialmente en contextos con datos **no lineales** o **valores atípicos**, donde su interpretación puede ser engañosa.

Nivel de ajuste	Rango de R^2	Interpretación
Excelente	$R^2 > 0.9$	Explica más del 90 % de la variabilidad
Bueno	$0.7 < R^2 \leq 0.9$	Explica entre 70 % y 90 % de la variabilidad
Regular	$0.5 < R^2 \leq 0.7$	Explica entre 50 % y 70 % de la variabilidad
Malo	$0 < R^2 \leq 0.5$	Explica menos del 50 % de la variabilidad
Muy malo	$R^2 \leq 0$	El modelo es peor que predecir simplemente con la media

Interpretación de R^2

Disciplina	Excelente	Bueno	Aceptable
Ciencias Físicas / Ingeniería	$R^2 > 0.95$	$R^2 > 0.90$	$R^2 > 0.80$
Ciencias Biológicas / Médicas	$R^2 > 0.80$	$R^2 > 0.60$	$R^2 > 0.40$
Ciencias Sociales / Economía	$R^2 > 0.70$	$R^2 > 0.50$	$R^2 > 0.30$
Marketing / Comportamiento	$R^2 > 0.60$	$R^2 > 0.40$	$R^2 > 0.20$
Finanzas / Mercados	–	$R^2 > 0.30$	$R^2 > 0.15$

> **Nota**: Un valor bajo de R^2 puede ser aceptable dependiendo de la disciplina y la variabilidad inherente de los datos. Por ejemplo, en estudios de comportamiento humano o mercados financieros, los fenómenos son altamente complejos y difíciles de predecir con alta precisión.

Valores válidos de R^2 según disciplina

Comparación entre Métricas

Las métricas de evaluación para regresión tienen características y aplicaciones distintas:

- **MSE (Error Cuadrático Medio)** : Se expresa en las unidades al cuadrado de la variable objetivo. Es muy sensible a los valores atípicos, ya que penaliza fuertemente los errores grandes. Su interpretación puede ser menos intuitiva debido a las unidades elevadas al cuadrado, pero es útil cuando se desea castigar los errores grandes.
- **RMSE (Raíz del Error Cuadrático Medio):** Se expresa en las mismas unidades que la variable objetivo, lo que facilita su interpretación. Al igual que el MSE, es sensible a los outliers y penaliza los errores grandes, pero permite entender el error promedio de manera directa.
- **MAE (Error Absoluto Medio):** También se expresa en las mismas unidades que la variable objetivo. Es menos sensible a los valores atípicos, ya que todos los errores se consideran por igual. Es ideal cuando se busca una métrica robusta ante outliers y se desea conocer el error absoluto promedio.
- **R^2 (Coeficiente de Determinación)** : Es una métrica adimensional que indica el porcentaje de la varianza explicada por el modelo. Permite comparar el ajuste de diferentes modelos y es especialmente útil para evaluar el desempeño global, aunque puede verse afectada por la presencia de outliers o relaciones no lineales.

Recomendaciones de Selección

Utiliza **MSE o RMSE** cuando los errores grandes sean especialmente críticos, como en aplicaciones médicas o financieras.

Prefiere **MAE** si existen valores atípicos o se requiere una métrica robusta ante errores extremos.

Emplea **R^2** para evaluar el ajuste general del modelo y comparar diferentes modelos sobre el mismo conjunto de datos.

Es recomendable combinar varias métricas para obtener una evaluación más completa y precisa del desempeño del modelo.

Introducción a modelos probabilísticos

Características principales

- **Incorporación de incertidumbre**: Reconocen que las predicciones tienen un grado de incertidumbre asociado
- **Distribuciones de probabilidad**: Utilizan distribuciones para describir el comportamiento de las variables
- **Predicciones con intervalos de confianza**: Proporcionan no solo una predicción puntual, sino también la confianza en esa predicción
- **Flexibilidad**: Pueden adaptarse a diferentes tipos de datos y problemas

Ventajas

- ✓ Proporcionan medidas de incertidumbre en las predicciones
- ✓ Permiten la incorporación de conocimiento previo (prior knowledge)
- ✓ Son especialmente útiles en situaciones con datos limitados o ruidosos
- ✓ Facilitan la toma de decisiones bajo incertidumbre

Aplicaciones comunes

- Diagnóstico médico
- Análisis de riesgo financiero
- Predicción meteorológica
- Procesamiento de lenguaje natural
- Sistemas de recomendación

Conceptos básicos de probabilidad

Probabilidad

- La probabilidad es un concepto fundamental en Machine Learning que nos permite modelar y cuantificar la incertidumbre inherente en los datos y procesos del mundo real.
- La probabilidad no solo nos ayuda a hacer predicciones, sino también a entender qué tan confiables son esas predicciones y cómo pueden cambiar cuando se dispone de nueva información.
- En esta sección exploraremos los conceptos básicos de probabilidad que constituyen la base teórica para desarrollar y comprender modelos probabilísticos efectivos en machine learning.

- **Experimento aleatorio**: Proceso cuyo resultado no se puede predecir con certeza. Ejemplos comunes incluyen lanzar una moneda, tirar un dado o seleccionar una muestra aleatoria de datos.
- **Espacio muestral** (S): Conjunto de todos los posibles resultados de un experimento aleatorio.
  Ejemplo: al lanzar un dado,

$$S = \{1, 2, 3, 4, 5, 6\}$$

- **Evento** (A): Subconjunto del espacio muestral. Representa un conjunto de resultados de interés.
  Ejemplo: $A = \{\text{obtener un número par}\} = \{2, 4, 6\}$
- **Probabilidad** ($P(A)$): Valor entre 0 y 1 que representa la posibilidad de que ocurra el evento A.
  Por ejemplo, si $A = \{2, 4, 6\}$, entonces

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

Definiciones clave

Propiedades de la probabilidad

1. **Rango de la probabilidad**:

$$0 \leq P(A) \leq 1$$

   Para cualquier evento A, su probabilidad siempre está entre 0 y 1.

2. **Probabilidad del espacio muestral**:

$$P(S) = 1$$

   La probabilidad de que ocurra algún resultado del espacio muestral completo es 1.

3. **Adición para eventos mutuamente excluyentes**:

   Si A y B son eventos **mutuamente excluyentes** (no pueden ocurrir al mismo tiempo), entonces:

$$P(A \cup B) = P(A) + P(B)$$

Tipos de probabilidad en machine learning

- **Probabilidad clásica:** Se usa cuando todos los resultados son igualmente probables, útil en ejemplos simples y teóricos.
- **Probabilidad frecuentista:** Se basa en la frecuencia relativa de un evento tras muchas repeticiones, común en la validación de modelos y experimentos.
- **Probabilidad bayesiana:** Interpreta la probabilidad como un grado de creencia, permitiendo actualizarla con nueva evidencia; es la base de muchos métodos modernos de machine learning.

Ejemplo aplicado

Supón que tienes un conjunto de datos de pacientes y quieres estimar la probabilidad de que un paciente tenga una enfermedad, dada cierta información.

Si de **100 pacientes**, **20** tienen la enfermedad, entonces la **probabilidad frecuentista** se estima como:

$$P(\text{enfermedad}) = \frac{20}{100} = 0.2$$

Esto significa que, según la frecuencia observada, hay un 20 % de probabilidad de que un paciente seleccionado al azar tenga la enfermedad.

Probabilidad Simple

La **probabilidad simple** es la forma más básica de calcular la posibilidad de que ocurra un evento específico. Se define como la razón entre el número de casos favorables y el número total de casos posibles, asumiendo que todos los resultados son igualmente probables.

Modelo matemático

$$P(A) = \frac{\text{Número de casos favorables}}{\text{Número total de casos posibles}}$$

Por ejemplo, al lanzar un dado justo, la probabilidad de obtener un número par es:

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

Características de la probabilidad simple

- Todos los resultados deben ser igualmente probables- Los eventos deben ser **mutuamente excluyentes** (no pueden ocurrir simultáneamente)
- Se aplica principalmente a experimentos con un **número finito de resultados.**

Ejemplos clásicos

Lanzamiento de una moneda:

- Espacio muestral:

$$S = \{\text{cara}, \text{cruz}\}$$

- Probabilidad de obtener cara:

$$P(\text{cara}) = \frac{1}{2} = 0.5$$

Ejemplos clásicos

Lanzamiento de un dado:

- Espacio muestral:

$$S = \{1, 2, 3, 4, 5, 6\}$$

- Probabilidad de obtener un número par:

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

- Probabilidad de obtener el número 5:

$$P(5) = \frac{1}{6} \approx 0.167$$

Probabilidad de Eventos Independientes

Independencia de Eventos

Dos eventos son **independientes** cuando la ocurrencia de uno **no afecta** la probabilidad de que ocurra el otro.
Este concepto es clave en *machine learning*, especialmente en algoritmos como **Naive Bayes**, que asume independencia condicional entre características.

Definición matemática

Dos eventos A y B son independientes si y solo si:

$$P(A \cap B) = P(A) \times P(B)$$

De forma equivalente, los eventos también son independientes si:

$$P(A \mid B) = P(A) \quad \text{y} \quad P(B \mid A) = P(B)$$

Esto significa que conocer la ocurrencia de uno no proporciona información sobre el otro.

Propiedades de eventos independientes

- La ocurrencia de un evento **no proporciona información** sobre el otro
- La probabilidad conjunta es el **producto de las probabilidades individuales**
- Son útiles para simplificar cálculos complejos en modelos probabilísticos

Ejemplos de independencia

1. Lanzamientos sucesivos de monedas

Dos lanzamientos de una moneda justa son eventos independientes, ya que el resultado del primero no afecta al segundo.

- Primer lanzamiento:

$$P(\text{cara}) = 0.5$$

- Segundo lanzamiento:

$$P(\text{cara}) = 0.5$$

- Probabilidad de obtener dos caras consecutivas:

$$P(\text{cara, cara}) = 0.5 \times 0.5 = 0.25$$

Ejemplos de independencia

2. Selección de características independientes en un dataset

Supón que en un conjunto de datos, la **altura** de una persona es independiente de su **preferencia musical**.

- Probabilidad de que una persona mida más de 1.70 m:

$$P(\text{altura} > 1.70\,\text{m}) = 0.4$$

- Probabilidad de que prefiera música clásica:

$$P(\text{música clásica}) = 0.2$$

- Probabilidad conjunta (si son independientes):

$$P(\text{altura} > 1.70\,\text{m} \cap \text{música clásica}) = 0.4 \times 0.2 = 0.08$$

Probabilidad de Eventos Dependientes

Dos eventos son dependientes cuando la ocurrencia de uno afecta la probabilidad de que ocurra el otro. En este caso, **el conocimiento sobre un evento proporciona información relevante sobre la probabilidad del segundo evento**. Este concepto es crucial en machine learning para modelar relaciones complejas entre variables.

Definición matemática

Dos eventos A y B son **dependientes** cuando:

$$P(A \cap B) \neq P(A) \times P(B)$$

Esto significa que la ocurrencia de uno **afecta** la probabilidad del otro.

En estos casos, usamos la **probabilidad condicional** para describir la relación entre los eventos.

Definición matemática

Modelo matemático de la probabilidad condicional

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Donde:

- $P(A \mid B)$: Probabilidad de que ocurra A dado que ocurrió B
- $P(A \cap B)$: Probabilidad de que ocurran ambos eventos A y B simultáneamente (intersección)
- $P(B)$: Probabilidad de que ocurra el evento B, con $P(B) > 0$

$$P(A \mid B) \quad \text{se lee como:} \quad \text{"la probabilidad de } A \text{ dado que ha ocurrido } B\text{"}$$

Características de eventos dependientes

- La ocurrencia de un evento **modifica la probabilidad** del otro
- Requieren el uso de **probabilidad condicional** para cálculos precisos
- Son más comunes en **datos del mundo real** donde las variables suelen estar correlacionadas
- La información sobre un evento es útil para predecir el otro

Ejemplo práctico

Uso del celular antes de dormir y calidad del sueño

Situación

- Se realizó una encuesta a 1,000 adolescentes para estudiar la relación entre el uso del celular antes de dormir y la calidad del sueño.

Objetivo

- Determinar si existe dependencia entre el uso del celular en la noche y reportar un sueño de buena calidad.

Datos obtenidos de la encuesta

- 700 adolescentes usan el celular antes de dormir.
- De los que usan el celular, 140 reportan tener un sueño de buena calidad.
- De los 300 que no usan el celular, 165 reportan un sueño de buena calidad.

Tabla de contingencia

	Sueño bueno	Sueño no bueno	Total
Usa celular	140	560	700
No usa celular	165	135	300
Total	305	695	1,000

Preguntas a responder

1. ¿Cuál es la probabilidad de que un adolescente use el celular antes de dormir?
2. ¿Cuál es la probabilidad de que un adolescente no use el celular antes de dormir?
3. ¿Cuál es la probabilidad de que un adolescente duerma bien?
4. ¿Cuál es la probabilidad de que un adolescente duerma bien *dado que* usa el celular?
5. ¿Cuál es la probabilidad de que un adolescente duerma bien *dado que* no usa el celular?
6. ¿Cuál es la probabilidad de que un adolescente use el celular y además duerma bien?

Pregunta 1

¿Cuál es la probabilidad de que un adolescente use el celular antes de dormir?

Datos relevantes:

Total de adolescentes encuestados: 1,000

Adolescentes que usan el celular antes de dormir: 700

Cálculo de la probabilidad:

$$P(\text{usa celular}) = \frac{700}{1000} = 0.70$$

Respuesta:

La probabilidad de que un adolescente use el celular antes de dormir es **0.70** o **70%**.

Pregunta 2

¿Cuál es la probabilidad de que un adolescente no use el celular antes de dormir?

Datos relevantes:

Total de adolescentes encuestados: 1,000

Adolescentes que no usan el celular antes de dormir: 300

Cálculo de la probabilidad:

$$P(\text{no usa celular}) = \frac{300}{1000} = 0.30$$

Respuesta:

La probabilidad de que un adolescente no use el celular antes de dormir es **0.30** o **30 %**.

Pregunta 3

¿Cuál es la probabilidad de que un adolescente duerma bien?

Datos relevantes:

Total de adolescentes encuestados: 1,000

Adolescentes que reportan sueño de buena calidad: 305

(140 que usan el celular + 165 que no lo usan)

Cálculo de la probabilidad:

$$P(\text{duerme bien}) = \frac{305}{1000} = 0.305$$

Respuesta:

La probabilidad de que un adolescente duerma bien es **0.305** o **30.5 %**.

Pregunta 4

¿Cuál es la probabilidad de que un adolescente duerma bien *dado que* usa el celular?

Datos relevantes:

Adolescentes que usan el celular antes de dormir: 700

De ellos, los que reportan sueño de buena calidad: 140

Cálculo de la probabilidad condicional:

$$P(\text{duerme bien} \mid \text{usa celular}) = \frac{140}{700} = 0.20$$

Respuesta:

La probabilidad de que un adolescente duerma bien *dado que* usa el celular antes de dormir es **0.20** o **20 %**.

Pregunta 5

¿Cuál es la probabilidad de que un adolescente duerma bien *dado que* no usa el celular?

Datos relevantes:

Adolescentes que no usan el celular antes de dormir: 300

De ellos, los que reportan sueño de buena calidad: 165

Cálculo de la probabilidad condicional:

$$P(\text{duerme bien} \mid \text{no usa celular}) = \frac{165}{300} = 0.55$$

Respuesta:

La probabilidad de que un adolescente duerma bien *dado que* no usa el celular antes de dormir es **0.55** o **55 %**.

Pregunta 6

¿Cuál es la probabilidad de que un adolescente use el celular y además duerma bien?

Datos relevantes:

Total de adolescentes encuestados: 1,000

Adolescentes que usan el celular y reportan sueño de buena calidad: 140

Cálculo de la probabilidad conjunta:

$$P(\text{usa celular} \cap \text{duerme bien}) = \frac{140}{1000} = 0.14$$

Respuesta:

La probabilidad de que un adolescente use el celular antes de dormir **y** duerma bien es **0.14** o **14 %**.

Diferencia entre la pregunta 4 y la pregunta 6

- **Pregunta 4: Probabilidad Condicional**

  Se pregunta: ¿Cuál es la probabilidad de que un adolescente **duerma bien** ***dado que*** **usa el celular** antes de dormir?

  Aquí se analiza **solo el subconjunto** de adolescentes que usan el celular, y se calcula cuántos de ellos duermen bien.

$$P(\text{duerme bien} \mid \text{usa celular}) = 0.20 \quad (20\%)$$

- **Pregunta 6: Probabilidad Conjunta**

  Se pregunta: ¿Cuál es la probabilidad de que un adolescente **use el celular y además duerma bien**?

  Aquí se considera **toda la población**, y se identifica a quienes cumplen ambas condiciones simultáneamente.

$$P(\text{usa celular} \cap \text{duerme bien}) = 0.14 \quad (14\%)$$

Hasta ahora hemos trabajado con:

- **Probabilidades simples:** como la proporción de adolescentes que usan el celular.
- **Probabilidades conjuntas:** como la probabilidad de que un adolescente use el celular y duerma mal.
- **Probabilidades condicionales directas:** como la probabilidad de dormir mal dado que se usó el celular.

Pero hay preguntas fundamentales que no se pueden responder solo con estos conceptos. Requieren una inversión en el sentido del razonamiento probabilístico, y eso solo es posible usando el Teorema de Bayes.