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

Semana 3

Introducción a Machine Learning

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.

- Las probabilidades conjuntas también se pueden calcular con el **Teorema de Bayes**.
- El teorema de Bayes es una herramienta fundamental en probabilidad y estadística que permite realizar este tipo de inferencias utilizando información previa y la probabilidad de observar esa consecuencia bajo diferentes causas.

El Teorema de Bayes

El teorema de Bayes establece que:

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

Donde:

- $P(A \mid B)$: Probabilidad de que ocurra el evento A dado que ocurrió el evento B (**probabilidad posterior**).
- $P(B \mid A)$: Probabilidad de observar el evento B si sabemos que ocurrió el evento A (**likelihood**).
- $P(A)$: Probabilidad de que ocurra el evento A (**probabilidad previa** o *prior*).
- $P(B)$: Probabilidad de que ocurra el evento B (**probabilidad marginal**).

En este contexto:

- La **probabilidad previa** (*prior*) representa nuestro conocimiento inicial sobre el evento A antes de observar B.
- La **probabilidad posterior** (*posterior*) es la probabilidad actualizada de A después de tener en cuenta la información de B.

Ejemplo

Si un adolescente durmió mal, ¿qué tan probable es que NO haya usado el celular?

El primer paso es identificar las 4 probabilidades requeridas para utilizar el Teorema de Bayes

Probabilidad	Expresión matemática	Definición en el contexto del problema
Prior	$P(\text{no usó el celular})$	Proporción total de adolescentes que no usaron el celular
Likelihood	$P(\text{durmió mal} \mid \text{no usó el celular})$	Probabilidad de que un adolescente que no usó el celular haya dormido mal
Marginal	$P(\text{durmió mal})$	Probabilidad total de que un adolescente haya dormido mal
Posterior (incógnita que queremos calcular)	$P(\text{no usó el celular} \mid \text{durmió mal})$	Probabilidad de que un adolescente que durmió mal no haya usado el celular

¿Cómo aplicar el Teorema de Bayes en este ejemplo?

Tenemos la probabilidad de que un adolescente **no haya usado el celular** $P(\text{no usó el celular})$ (*prior*).

Sin embargo, queremos conocer **cómo cambia esa probabilidad** una vez que sabemos que el adolescente **durmió mal** $P(\text{no usó el celular} \mid \text{durmió mal})$ (*posterior*).

Para calcular esta probabilidad actualizada, necesitamos además:

- La probabilidad total de que un adolescente **duerma mal** $P(\text{durmió mal})$ (*marginal*), y
- La probabilidad de que un adolescente **duerma mal dado que no usó el celular** $P(\text{durmió mal} \mid \text{no usó el celular})$ (*likelihood*), es decir, la probabilidad de observar esta evidencia si asumimos que no usó el celular.

Resolviendo

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

Probabilidad	Expresión matemática	Cálculo con la tabla	Valor
Prior	$P(\text{no usó el celular})$	$\frac{300}{1000}$	0.3
Likelihood	$P(\text{durmió mal} \mid \text{no usó el celular})$	$\frac{135}{300}$	0.45
Marginal	$P(\text{durmió mal})$	$\frac{695}{1000}$	0.695
Posterior	$P(\text{no usó el celular} \mid \text{durmió mal})$	$\frac{0.45 \times 0.3}{0.695}$	≈ 0.1942

Pregunta 4 (nuevamente)

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

$$P(\text{duerme bien} \mid \text{usa celular}) = \frac{P(\text{usa celular} \mid \text{duerme bien}) \cdot P(\text{duerme bien})}{P(\text{usa celular})}$$

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

Probabilidad	**Expresión matemática**	**Cálculo con la tabla**	**Valor**
Prior	$P(\text{duerme bien})$	$\frac{305}{1000}$	0.305
Likelihood	$P(\text{usa el celular} \mid \text{duerme bien})$	$\frac{140}{305}$	≈ 0.459
Marginal	$P(\text{usa el celular})$	$\frac{700}{1000}$	0.7
Posterior	$P(\text{duerme bien} \mid \text{usa el celular})$	$\frac{0.459 \times 0.305}{0.7}$	≈ 0.20

Pregunta 5 (nuevamente)

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

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

Probabilidad	Expresión matemática	Cálculo con la tabla	Valor
Prior	$P(\text{duerme bien})$	$\frac{305}{1000}$	0.305
Likelihood	$P(\text{no usa celular} \mid \text{duerme bien})$	$\frac{165}{305}$	≈ 0.541
Marginal	$P(\text{no usa celular})$	$\frac{300}{1000}$	0.3
Posterior	$P(\text{duerme bien} \mid \text{no usa celular})$	$\frac{0.541 \times 0.305}{0.3}$	≈ 0.55

Pregunta para resolver

Si un adolescente durmió mal, ¿cuál es la probabilidad de que haya usado el celular en la noche?

$$P(\text{usa celular} \mid \text{duerme mal}) = \frac{P(\text{duerme mal} \mid \text{usa celular}) \cdot P(\text{usa celular})}{P(\text{duerme mal})}$$

	Sueño bueno	Sueño no bueno (mal)	Total
Usa celular	140	560	700
No usa celular	165	135	300
Total	305	695	1000

Probabilidad	Expresión matemática	Cálculo con la tabla	Valor
Prior	$P(\text{duerme mal})$	$\frac{695}{1000}$	0.695
Likelihood	$P(\text{duerme mal} \mid \text{usa celular})$	$\frac{560}{700}$	0.8
Marginal	$P(\text{usa celular})$	$\frac{700}{1000}$	0.7
Posterior	$P(\text{usa celular} \mid \text{duerme mal})$	$\frac{0.8 \times 0.7}{0.695}$	≈ 0.806

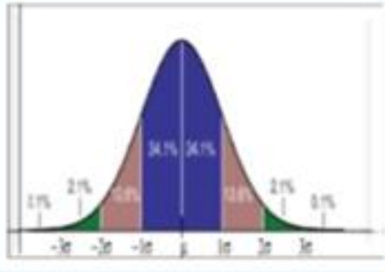
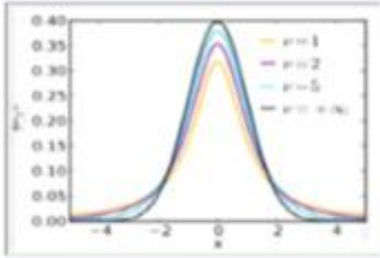
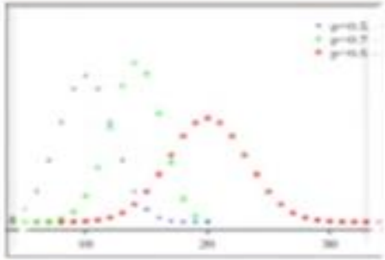
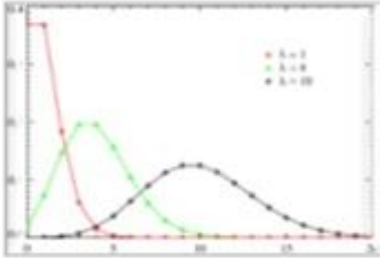
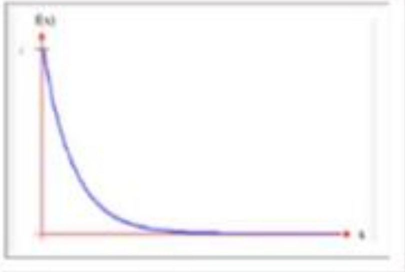
Distribuciones de datos

Definición

Las **distribuciones de datos** describen cómo se reparten los valores de un conjunto de datos. Nos muestran **qué valores son más frecuentes** y **cómo varían**.

Ejemplos comunes de distribuciones son la **normal** (en forma de campana), la **uniforme** (todos los valores igual de probables) y la **sesgada** (más acumulación en un extremo).

Ejemplos de distribuciones

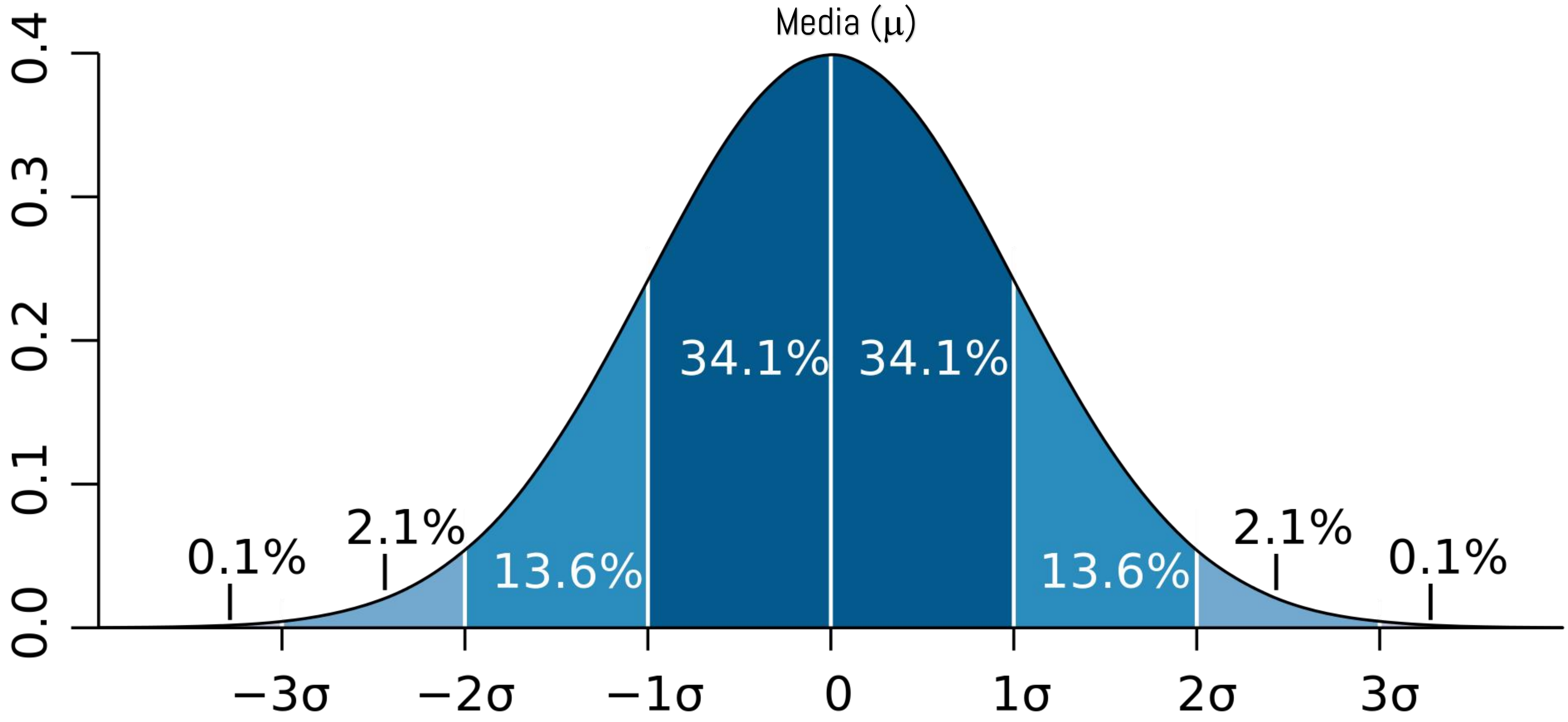
	Distribución				
	Normal	**T - Student**	**Binomial**	**Poisson**	**Exponencial**
Visualización					
Características	Forma de Campana	Más corta que una distribución Normal	Dos resultados: Éxito / Fallo	Diferentes formas, pero válida solo para enteros en el eje x	Modeliza el tiempo entre eventos
Ejemplos y Aplicaciones	Modelización de fenómenos naturales	Cuando tienes pequeñas muestras y desconoces la varianza		Probabilidad de número de eventos en un intervalo fijo	cuánto tiempo pasará antes de que un gran huracán golpee el Atlántico
	Ajuste de mínimos cuadrados o propagación de incertidumbre		Lanzamiento de una moneda. Debes utilizarla en cualquier conjunto de datos binarios (Sí/no; Vivo/Muerto)		construyendo cadenas de markov de tiempo continuo

Distribución normal

La **distribución normal** es una distribución de probabilidad continua con forma de campana simétrica. Se caracteriza por dos parámetros: la media μ, que determina el centro de la distribución, y la desviación estándar σ, que indica el grado de dispersión de los datos.

En esta distribución, los valores se concentran alrededor de μ, y la frecuencia disminuye progresivamente a medida que los valores se alejan. Es fundamental en estadística porque muchos fenómenos naturales y procesos aleatorios tienden a seguir este patrón.

Distribución Normal



Distribución Normal

- A la distribución normal también se le conoce como **campana de Gauss**, en honor al matemático alemán **Carl Friedrich Gauss**, quien la empleó para modelar errores de medición en astronomía.
- Aunque no todos los fenómenos naturales siguen esta distribución, su importancia se basa en el **teorema del límite central**, que establece que la suma de un gran número de variables aleatorias tiende a aproximarse a una distribución normal.

Distribución Normal

- El término "normal" fue popularizado por **Karl Pearson**, no porque los datos del mundo real sean necesariamente normales, sino porque esta distribución sirve como modelo teórico central en muchos métodos estadísticos.
- La primera formulación matemática de esta curva se atribuye a **Abraham de Moivre** en el siglo XVIII.

Formulación Matemática

La distribución normal obedece al siguiente modelo matemático:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde

- x es la variable aleatoria continua
- μ es la **media** de la distribución
- σ es la **desviación estándar**
- e es la base del logaritmo natural
- π es el número pi

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

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

Distribución normal

El modelo matemático de la distribución normal muestra que para describir la forma de su curva sólo necesitamos conocer dos parámetros: la **media μ**, que indica el valor alrededor del cual se agrupan la mayoría de los datos, y la **desviación estándar σ**, que nos dice cuánto se dispersan esos datos respecto a la media.

Con sólo estos dos valores, podemos representar completamente la distribución.

Naïve Bayes Classifier

Clasificador Naïve Bayes

El clasificador Naïve Bayes es un modelo de clasificación basado en el Teorema de Bayes con el supuesto de que las variables predictoras son independientes entre sí (de ahí el término naïve, ingenuo).

Calcula la probabilidad de que una instancia pertenezca a una clase dada, combinando la probabilidad previa de la clase y las probabilidades condicionales de las características observadas.

Modelo matemático

Dado un conjunto de características $X = (x_1, x_2, \ldots, x_n)$, queremos predecir la clase C_k más probable.

El clasificador Naive Bayes calcula:

$$P(C_k \mid X) = \frac{P(X \mid C_k) \cdot P(C_k)}{P(X)}$$

Como $P(X)$ es constante para todas las clases, podemos enfocarnos en maximizar el numerador:

$$P(C_k \mid X) \propto P(X \mid C_k) \cdot P(C_k)$$

Modelo matemático

Por lo tanto, el modelo se reduce a:

$$P(C_k \mid X) \propto P(C_k) \cdot \prod_{i=1}^{n} P(x_i \mid C_k)$$

Finalmente, se elige la clase que **maximiza esta expresión**:

$$\hat{C} = \arg\max_{C_k} \left(P(C_k) \cdot \prod_{i=1}^{n} P(x_i \mid C_k) \right)$$

Variantes de Naïve Bayes

- **Gaussian Naive Bayes:** Se usa cuando los features son **contínuos** y se asume que siguen una **distribución normal**.
- **Multinomial Naive Bayes:** Se utiliza para características que representan **conteos discretos**, como el número de veces que una palabra aparece en un documento.
- **Bernoulli Naive Bayes:** Se emplea cuando las características son **binarias** (por ejemplo: presencia o ausencia de una palabra en un documento).

Gaussian **Naïve Bayes**

Gaussian Naive Bayes se usa cuando las variables predictoras x_i son **continuas** y se asume que, para cada clase C_k, cada característica sigue una **distribución normal** (gaussiana).

Es decir:

$$P(x_i \mid C_k) \sim \mathcal{N}(\mu_{ik}, \sigma_{ik}^2)$$

donde:

- μ_{ik} es la media de la característica x_i para la clase C_k
- σ_{ik}^2 es la varianza de x_i para la clase C_k

Gaussian **Naïve Bayes**

En GNB, el cálculo de la **likelihood** $P(x_i \mid C_k)$ para cada característica x_i se hace usando la **función de densidad de una distribución normal** (la fórmula de Gauss).

Formalmente:

$$P(x_i \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} e^{-\frac{(x_i - \mu_{ik})^2}{2\sigma_{ik}^2}}$$

Por lo tanto, el **cálculo de la likelihood en Gaussian Naive Bayes se basa en la fórmula de Gauss**, usando los parámetros μ_{ik} (media) y σ_{ik}^2 (varianza) estimados a partir de los datos de entrenamiento para cada clase C_k.

Luego, se combinan las likelihoods de todas las características (producto de las likelihoods) y se multiplican por la probabilidad previa $P(C_k)$ para obtener $P(C_k \mid X)$.

Gaussian Naïve Bayes

En Gaussian Naive Bayes, el Teorema de Bayes permite calcular la probabilidad de que una observación pertenezca a una clase combinando la probabilidad previa de la clase (prior), la verosimilitud de los datos bajo esa clase (likelihood, modelada con la función de Gauss), y la evidencia total observada (marginal), para obtener una probabilidad actualizada (posterior).

El modelo es:

$$P(C_k \mid X) = \frac{P(X \mid C_k) \cdot P(C_k)}{P(X)}$$

donde

$$P(X \mid C_k) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} e^{-\frac{(x_i - \mu_{ik})^2}{2\sigma_{ik}^2}}$$

Gaussian Naïve Bayes

En Gaussian Naive Bayes, las probabilidades de cada característica se combinan bajo el supuesto de independencia condicional. Es decir, se asume que cada característica x_i es independiente de las demás, dado que conocemos la clase C_k.

Gracias a este supuesto, la probabilidad conjunta $P(X \mid C_k)$ se puede descomponer como el producto de las probabilidades individuales:

$$P(X \mid C_k) = \prod_{i=1}^{n} P(x_i \mid C_k)$$

Gaussian Naïve Bayes

Dado que el producto de múltiples probabilidades pequeñas puede dar lugar a números muy cercanos a cero (problemas de subdesbordamiento numérico), en la práctica este cálculo se realiza en el dominio logarítmico:

$$\log P(X \mid C_k) = \sum_{i=1}^{n} \log P(x_i \mid C_k)$$

Finalmente, la decisión del clasificador se toma eligiendo la clase C_k que maximiza la probabilidad posterior. Esto se expresa como:

$$\hat{C} = \arg\max_{C_k} \left[\log P(C_k) + \sum_{i=1}^{n} \log P(x_i \mid C_k) \right]$$

Ejemplo: Pima Indian Diabetes Dataset

Cálculo de priors para cada clase

Se parte de la versión limpia del conjunto de datos:

- Total de muestras: 392
- Muestras positivas (clase 1): 130
- Muestras negativas (clase 0): 392 - 130 = 262

Los priors se calculan como la proporción de muestras de cada clase respecto al total:

$$P(\text{clase } 1) = \frac{130}{392} \approx 0.3316 \quad (33.16\%)$$

$$P(\text{clase } 0) = \frac{262}{392} \approx 0.6684 \quad (66.84\%)$$

Estos valores representan la probabilidad a priori de cada clase antes de observar los datos.

Cálculos estadísticos por clase: media y desviación estándar

Clase 0 (sin diabetes)

Variable	Media	Desviación estándar
Pregnancies	2.72	2.65
Glucose	111.19	24.42
Blood Pressure	68.96	12.37
Skin Thickness	27.31	10.42
Insulin	131.69	103.16
BMI	31.92	6.90
Diabetes Pedigree Function	0.47	0.31
Age	28.01	8.50

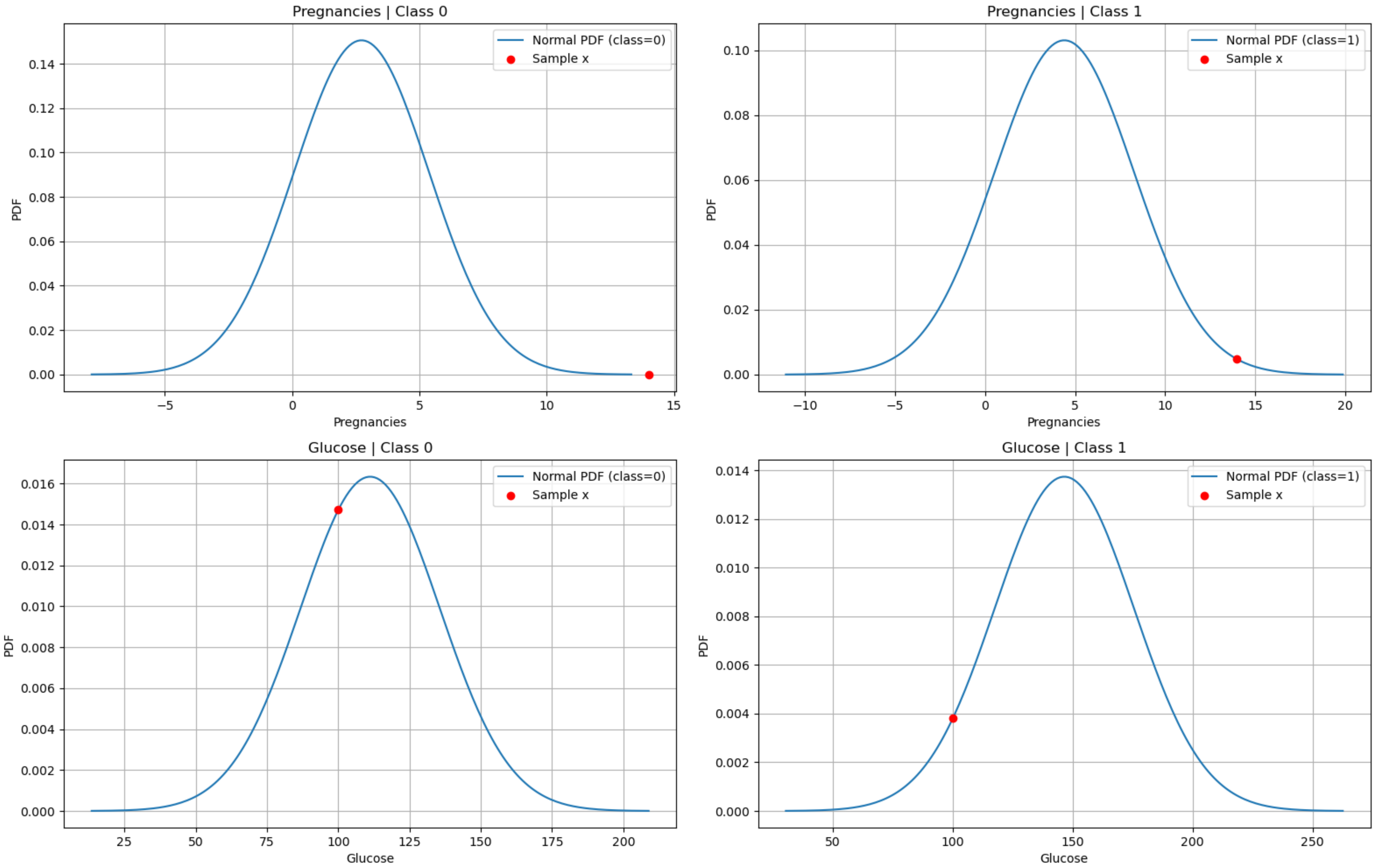
Selección de una muestra

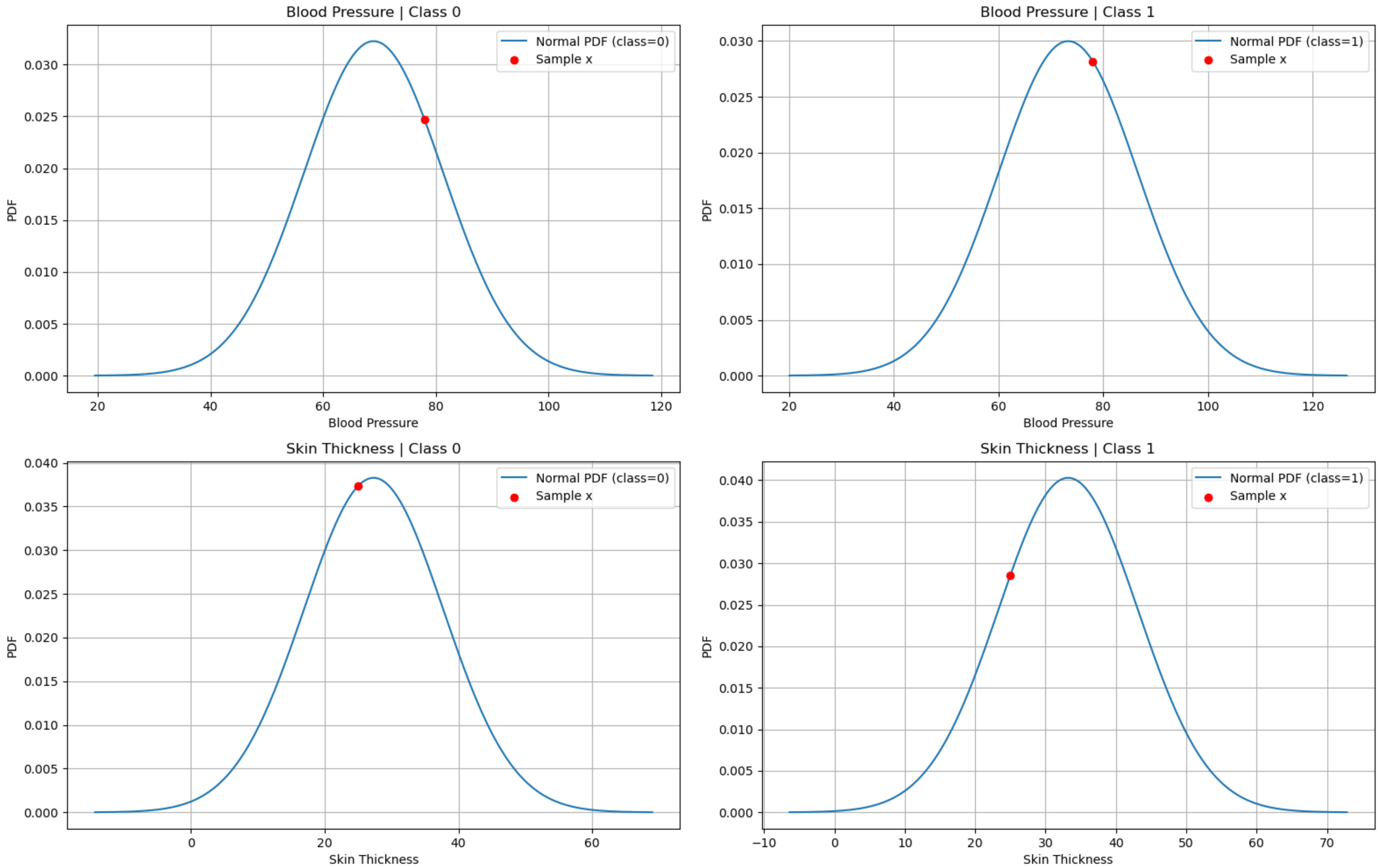
Variable	Valor observado
Pregnancies	14
Glucose	100
Blood Pressure	78
Skin Thickness	25
Insulin	184
BMI	36.6
Diabetes Pedigree Function	0.412
Age	46
Outcome (real)	1

Cálculo del likelihood

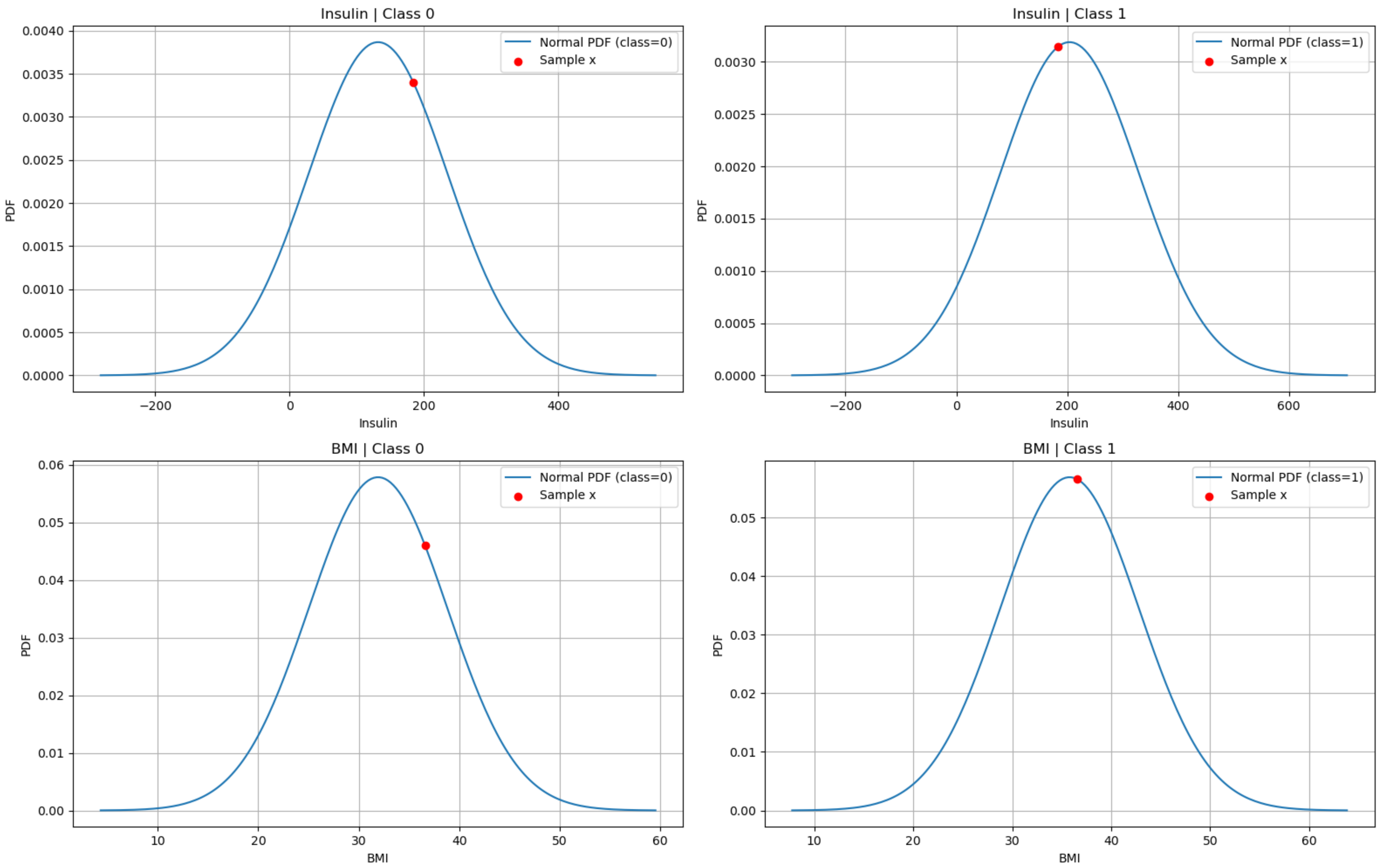
Para cada clase C_k (Outcome = 0 o 1), se calcula la verosimilitud de observar la muestra $\mathbf{x} = (x_1, x_2, ..., x_n)$, bajo la suposición de independencia condicional entre las variables.

Clase C_k	$P(x_{\text{Pregnancies}} \mid C_k)$	$P(x_{\text{Glucose}} \mid C_k)$	$P(x_{\text{Blood Pressure}} \mid C_k)$	$P(x_{\text{Skin Thickness}} \mid C_k)$	$P(x_{\text{Insulin}} \mid C_k)$	$P(x_{\text{BMI}} \mid C_k)$
0	0.000017	0.014708	0.024692	0.037351	0.003401	0.045950
1	0.004768	0.003819	0.028177	0.028566	0.003148	0.056542


Pregnancies | Class 0
Normal PDF (class=0)
Sample x
PDF
Pregnancies
Pregnancies | Class 1
Normal PDF (class=1)
Sample x
PDF
Pregnancies
Glucose | Class 0
Normal PDF (class=0)
Sample x
PDF
Glucose
Glucose | Class 1
Normal PDF (class=1)
Sample x
PDF
Glucose



Blood Pressure | Class 0
Normal PDF (class=0)
Sample x
PDF
Blood Pressure
Blood Pressure | Class 1
Normal PDF (class=1)
Sample x
PDF
Blood Pressure
Skin Thickness | Class 0
Normal PDF (class=0)
Sample x
PDF
Skin Thickness
Skin Thickness | Class 1
Normal PDF (class=1)
Sample x
PDF
Skin Thickness


Insulin | Class 0
Normal PDF (class=0)
Sample x
PDF
Insulin
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
−200
0
200
400
Insulin | Class 1
Normal PDF (class=1)
Sample x
PDF
Insulin
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
−200
0
200
400
600
BMI | Class 0
Normal PDF (class=0)
Sample x
PDF
BMI
0.06
0.05
0.04
0.03
0.02
0.01
0.00
10
20
30
40
50
60
BMI | Class 1
Normal PDF (class=1)
Sample x
PDF
BMI
0.05
0.04
0.03
0.02
0.01
0.00
10
20
30
40
50
60

Probabilidades posteriores

Feature	$P(x_i \mid C_0) \cdot P(C_0)$	$P(x_i \mid C_1) \cdot P(C_1)$
Pregnancies	0.000017×0.6684	0.004768×0.3316
Glucose	0.014708×0.6684	0.003819×0.3316
Blood Pressure	0.024692×0.6684	0.028177×0.3316
Skin Thickness	0.037351×0.6684	0.028566×0.3316
Insulin	0.003401×0.6684	0.003148×0.3316
BMI	0.045950×0.6684	0.056542×0.3316
Diabetes Pedigree Function	1.261183×0.6684	0.866924×0.3316
Age	0.005001×0.6684	0.024575×0.3316

Teorema de Bayes Simplificado

En Naïve Bayes no hace falta calcular explícitamente la probabilidad marginal $P(\mathbf{x})$, ya que esta es la misma para todas las clases.

Al comparar $P(C_k \mid \mathbf{x})$ entre clases, $P(\mathbf{x})$ aparece en el denominador de todas ellas y se cancela.

Por lo tanto, basta con comparar los valores no normalizados $P(\mathbf{x} \mid C_k) \cdot P(C_k)$.

Cálculo de probabilidad de cada clase

Asumiendo total independencia entre los features, la probabilidad de cada clase se puede calcular como el producto de las verosimilitudes de cada feature, multiplicado por la probabilidad a priori de cada clase, lo cual convierte el cálculo en un cálculo de probabilidad posterior:

$$P(C_k \mid \mathbf{x}) \propto P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)$$

Cálculo de probabilidad de cada clase

$$P(C_0 \mid \mathbf{x}) \propto 0.6684 \times 0.000017 \times 0.014708 \times 0.024692 \times 0.037351 \times 0.003401 \times 0.045950 \times 1.261183 \times 0.005001$$
$$= 2.940 \times 10^{-16}$$

Para C_1:

$$P(C_1 \mid \mathbf{x}) \propto 0.3316 \times 0.004768 \times 0.003819 \times 0.028177 \times 0.028566 \times 0.003148 \times 0.056542 \times 0.866924 \times 0.024575$$
$$= 1.153 \times 10^{-13}$$

Cálculo de probabilidad con logaritmos

En la práctica, se trabaja con log-verosimilitudes para evitar problemas de precisión al multiplicar muchos números pequeños.

$$\log P(C_k \mid \mathbf{x}) \propto \log P(C_k) + \sum_{i=1}^{n} \log P(x_i \mid C_k)$$

Para C_0:

$$\log P(C_0 \mid \mathbf{x}) \propto \log(0.6684) + \log(0.000017) + \log(0.014708) + \log(0.024692) + \log(0.037351) + \log(0.003401) + \log(0.045950) + \log(1.261183) + \log(0.005001)$$

$$= -0.4023 + (-10.9823) + (-4.2239) + (-3.7049) + (-3.2852) + (-5.6801) + (-3.0818) + 0.2321 + (-5.2987)$$

$$= -36.4268$$

Para C_1:

$$\log P(C_1 \mid \mathbf{x}) \propto \log(0.3316) + \log(0.004768) + \log(0.003819) + \log(0.028177) + \log(0.028566) + \log(0.003148) + \log(0.056542) + \log(0.866924) + \log(0.024575)$$

$$= -1.1044 + (-5.3466) + (-5.5707) + (-3.5669) + (-3.5546) + (-5.7590) + (-2.8713) + (-0.1430) + (-3.7042)$$

$$= -31.6207$$

Conversión a probabilidad

Se parte del cálculo en logaritmos:

$$\log P(C_0 \mid \mathbf{x}) = -36.4268$$

$$\log P(C_1 \mid \mathbf{x}) = -31.6207$$

Se pasa de logaritmos a escala normal mediante la exponencial:

$$\text{score}_0 = e^{-36.4268} = 2.94 \times 10^{-16}$$

$$\text{score}_1 = e^{-31.6207} = 1.153 \times 10^{-13}$$

Luego se normalizan los scores para obtener probabilidades que sumen 1:

$$P(C_0 \mid \mathbf{x}) = \frac{2.94 \times 10^{-16}}{2.94 \times 10^{-16} + 1.153 \times 10^{-13}} = 0.0025$$

$$P(C_1 \mid \mathbf{x}) = \frac{1.153 \times 10^{-13}}{2.94 \times 10^{-16} + 1.153 \times 10^{-13}} = 0.9975$$

Resultado

Se obtiene:

$$P(C_0 \mid \mathbf{x}) = 0.0025$$

$$P(C_1 \mid \mathbf{x}) = 0.9975$$

Decisión:

$$P(C_1 \mid \mathbf{x}) > P(C_0 \mid \mathbf{x})$$

Clasificación final:

Se asigna la muestra a la clase C_1.

Interpretación:

Existe una alta probabilidad de que la muestra pertenezca a la clase C_1 (99.75%), por lo que el modelo predice que la paciente tiene diabetes.