Introducción a la inferencia bayesiana

Clase 9: el Teorema de Bayes, *prior*, *posterior*, *MAP* y *LMS*

6.431x — Probability: The Science of Uncertainty and Data

(versión de clase: las soluciones se comparten al terminar)

B.Sc. Cristian Lazo Quispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Programa SDS (Statistics and Data Science, MITx) · **BREIT** (Brescia Institute of Technology), iniciativa de Aporta —Grupo Breca— en alianza con MIT IDSS.
Basado en MIT 6.431x / MIT OCW RES.6-012 (CC BY-NC-SA 4.0).

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

Plan de hoy: de lo fácil a lo difícil

Las cuatro partes (2 horas)

1. **Kahoot** (10 min): 5 preguntas de la Clase 8 y 5 de la Clase 9.
2. **Repaso de la Clase 8** (15 min): la **fórmula de los límites** de la convolución y **un** ejemplo (el problema 1 de la Clase 8, por el camino corto).
3. **Teoría de la Clase 9** (40 min): por qué Bayes (video, IA que genera imágenes), el **Teorema de Bayes** con la **tabla de Bayes**, MAP y LMS, y Bayes en Machine Learning.
4. **Cinco problemas de la Clase 9** (55 min), en escalera:

La escalera de los cinco problemas

P1	★	¿qué moneda tengo?
P2	★★	¿qué máquina falló?
P3	★★★	MAP frente a LMS
P4	★★★★	dos caras: actualizar y predecir
P5	★★★★★	adivinar con 3 tiros (PS 7a)

La idea de hoy

Cada problema se resuelve con **una receta corta** (una tabla, una fórmula o un dibujo). Si una cuenta se pone larga, **hay un atajo**: lo buscamos juntos.

Kahoot de calentamiento (Clases 8 y 9)

Jugamos ahora: 10 preguntas, unos 10 minutos

- Preguntas 1 a 5 (**Clase 8**): sumar dos v.a. contando pares, la forma de la suma de uniformes, $\mathrm{Var}(X+Y)$ con covarianza, incorrelado frente a independiente y esperanzas iteradas.
- Preguntas 6 a 10 (**Clase 9**): la tabla de Bayes, MAP y LMS, el filtro de spam, un segundo examen positivo y predecir la próxima cara.

Antes de responder

Las primeras son de 30 segundos: confía en la intuición. En las de 90 segundos hay una **trampa** (el error clásico) entre las opciones: escribe la cuenta antes de elegir.

¿Qué veremos hoy?

1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

Contenido

1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

La fórmula de los límites: la convolución de uniformes sin dolor

$X \sim \text{Unif}[a,b]$, $Y \sim \text{Unif}[c,d]$ independientes

Para que $X + Y = z$: si $X = x$, entonces $Y = z - x$. Un x **sirve** si $a \le x \le b$ **y** $c \le z - x \le d$. Juntando las dos condiciones:

$$\text{máx}(a,\ z-d) \ \le\ x\ \le\ \text{mín}(b,\ z-c).$$

La fórmula general

$$f_Z(z) = \text{altura} \times \text{largo}(z)$$

$$\text{altura} = \frac{1}{(b-a)(d-c)}$$

$$\text{largo}(z) = \text{mín}(b,\ z-c) - \text{máx}(a,\ z-d)$$

Si sale negativo, el largo es 0: ningún x sirve.

Probémosla: $X \sim \text{Unif}[0,2]$, $Y \sim \text{Unif}[3,4]$, $z = 3{,}5$

$$\text{Altura} = \frac{1}{(2-0)(4-3)} = \frac{1}{2}.$$

$$\text{largo} = \text{mín}(2;\ 0{,}5) - \text{máx}(0;\ -0{,}5)$$

$$= 0{,}5 - 0 = 0{,}5$$

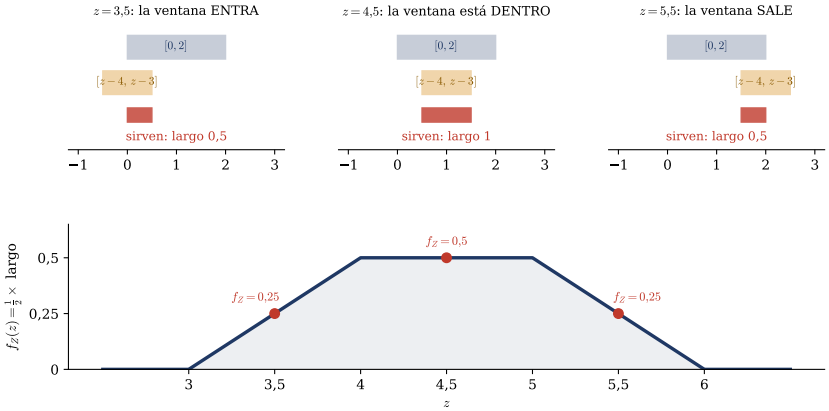
$$f_Z(3{,}5) = \tfrac{1}{2} \times 0{,}5 = 0{,}25.$$

Sin integrales: solo un mín y un máx.

Atajo: siempre sale un trapecio

Basta evaluar la fórmula en unos pocos z y **unir con rectas**. Empieza en $a + c$, termina en $b + d$; si los anchos son iguales, es un triángulo. (¿Por qué? Dos slides más adelante.)

Verlo: la ventana que se desliza



$X \sim \text{Unif}[0, 2]$: los x posibles (azul). Para un z dado, Y "pide" $x \in [z-4, z-3]$ (ámbar, una ventana de ancho 1). Lo que sirve es la parte común (rojo); su largo por la altura $\frac{1}{2}$ da $f_Z(z)$. Al subir z la ventana se corre a la derecha: **entra** (sube), está **dentro** (meseta) y **sale** (baja).

¿Cómo sé que hay tres tramos? ¿Siempre son tres?

Los tramos cambian donde cambia el máx o el mín

En largo$(z) = \text{mín}(b,\ z-c) - \text{máx}(a,\ z-d)$:

- $\text{máx}(a, z-d)$ cambia de "a" a "$z-d$" cuando $z - d = a$, o sea en $\boxed{z = a + d}$.
- $\text{mín}(b, z-c)$ cambia de "$z-c$" a "b" cuando $z - c = b$, o sea en $\boxed{z = b + c}$.

Más el inicio $a + c$ (lo mínimo que puede sumar) y el final $b + d$ (lo máximo): **4 puntos de quiebre ⇒ 3 tramos**.

Receta (sirve para *cualquier* par de uniformes)

1. Calcula los 4 números $a + c$, $a + d$, $b + c$, $b + d$ y **ordénalos**.
2. En los extremos $f_Z = 0$; en los dos del medio $f_Z = \dfrac{1}{\text{ancho mayor}}$.
3. Une con rectas: sube → meseta → baja.

Ejemplo: $[0, 2] + [3, 4]$ (la ventana entra, está dentro, sale)

$a + c = 3$, $a + d = 4$, $b + c = 5$, $b + d = 6$. Anchos 2 y 1: meseta en $\frac{1}{2}$.

tramo	$\text{máx}(a, z-d)$	$\text{mín}(b, z-c)$	largo
$3 \le z \le 4$	0	$z - 3$	$z - 3$
$4 \le z \le 5$	$z - 4$	$z - 3$	1
$5 \le z \le 6$	$z - 4$	2	$6 - z$

¿Siempre tres? Casi

- **Anchos distintos** ($a + d \neq b + c$): 3 tramos, **trapecio**.
- **Anchos iguales** ($a + d = b + c$): el tramo del medio mide 0 y quedan **2 tramos**, **triángulo**. Ej.: $[0, 1] + [0, 1]$ → puntos $0, 1, 1, 2$.
- Con densidades **no uniformes** (o por tramos) puede haber más quiebres: la idea es la misma, los tramos cambian donde cambia algún límite.

Repaso (1/2): el problema 1 de la Clase 8 con la fórmula

El problema y los datos de la fórmula

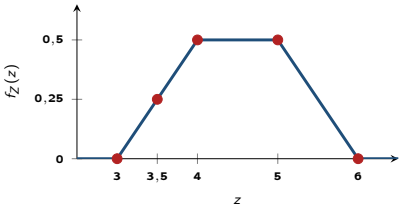
$X \sim \text{Unif}[0,2]$, $Y \sim \text{Unif}[3,4]$ independientes, $Z = X + Y$: halla f_Z y $\mathbb{P}(Z \le 3{,}5)$. Datos: $a = 0$, $b = 2$, $c = 3$, $d = 4$.

$$\text{altura} = \frac{1}{(2-0)(4-3)} = \frac{1}{2}$$

$$\text{largo}(z) = \text{mín}(2,\, z-3) - \text{máx}(0,\, z-4)$$

Paso 2: la tabla (solo restar; $f_Z = \frac{1}{2} \times \text{largo}$)

z	$\text{mín}(2, z-3)$	$\text{máx}(0, z-4)$	largo	f_Z
3	0	0	0	0
3,5	0,5	0	0,5	0,25
4	1	0	1	0,5
5	2	1	1	0,5
6	2	2	0	0



Se unen los puntos rojos con rectas: un **trapecio**.

Y por tramos, si lo piden

$f_Z = \frac{z-3}{2}$ en $[3,4]$; $\frac{1}{2}$ en $[4,5]$; $\frac{6-z}{2}$ en $[5,6]$; 0 fuera. (Son las rectas que unen los puntos.)

Repaso (2/2): $\mathbb{P}(Z \leq 3{,}5)$ con un dibujo, sin integrar

Paso 3: la conjunta es "plana"

(X, Y) es uniforme en el rectángulo $[0, 2] \times [3, 4]$, de área 2. Entonces

$$\mathbb{P}((X, Y) \in R) = \frac{\text{área de } R}{\text{área total}} = \frac{\text{área de } R}{2}.$$

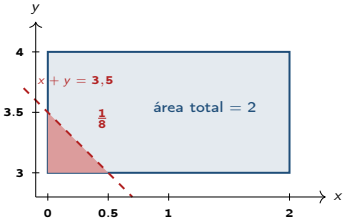
Paso 4: la región $x + y \leq 3{,}5$

Es el triangulito de la esquina, con catetos 0,5 y 0,5:

$$\text{área} = \tfrac{1}{2} \cdot 0{,}5 \cdot 0{,}5 = \tfrac{1}{8}, \qquad \mathbb{P}(Z \leq 3{,}5) = \frac{1/8}{2} = \frac{1}{16}.$$

Lo que hay que llevarse

Para **probabilidades** de una suma de uniformes, muchas veces no hace falta la densidad: se dibuja la recta $x + y = t$ en el rectángulo y se calcula un área.



Mismo resultado que integrar $\int_{\mathbf{3}}^{\mathbf{3,5}} \frac{z-\mathbf{3}}{\mathbf{2}}\,dz = \frac{\mathbf{1}}{\mathbf{16}}$, pero sin integrales.

Contenido

1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

Video: cómo escapar de la trampa bayesiana (Veritasium en español)



youtube.com/watch?v=D7KKlC0LOyw **(clic en la imagen)**

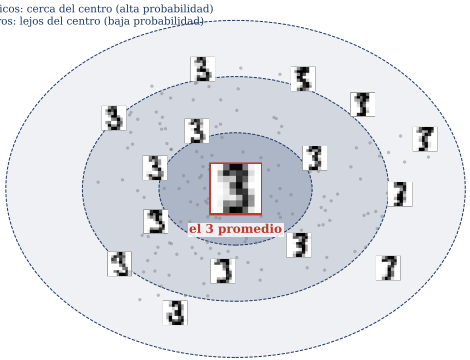
Qué ver (primeros 4 minutos)

- **0:00–2:50** El examen médico: sales positivo en una enfermedad rara. ¿Qué tan probable es que la tengas? (Lo resolvemos más adelante con la tabla de Bayes.)
- **3:19–4:21** La mesa de Bayes: adivinar dónde cayó una bola tirando más bolas y preguntando "¿a la izquierda o a la derecha?".

El pensamiento bayesiano

No estamos seguros de cómo es el mundo, pero **con cada nueva evidencia** corregimos lo que creíamos: la posterior de hoy es la **prior** de mañana.

Una imagen también tiene probabilidad



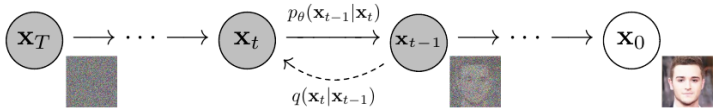


Imagina todas las fotos de **gatos** del mundo (aquí, para que sea real, 183 números **3** escritos a mano). Forman una **distribución**: en el centro el "gato promedio"; cerca, los gatos típicos (alta probabilidad); lejos, los raros (que ya casi parecen otra cosa: mira los 3 del borde, parecen 7).

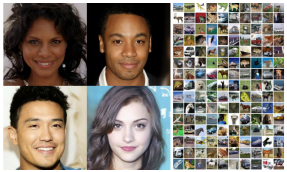
Sampling (muestrear)

Si conoces la distribución, puedes **sacar imágenes nuevas** de ella (derecha): no están en el dataset. Eso hace una IA que genera imágenes: aprende la distribución de los datos y muestrea.

Así genera imágenes una IA: modelos de difusión



Ho, Jain y Abbeel (2020), *Denoising Diffusion Probabilistic Models*, NeurIPS, Fig. 2: de x_T (ruido puro) a x_0 (una cara), paso a paso.



Imágenes generadas desde ruido (misma fuente, Fig. 1).

La idea en tres pasos

1. **Ensuciar**: a una foto real se le agrega ruido poco a poco hasta que no queda nada.
2. **Aprender a limpiar**: una red neuronal ve la foto con ruido y **adivina la foto limpia**. Eso es inferencia bayesiana: dato = foto con ruido, lo desconocido = foto limpia.
3. **Generar**: partir de ruido puro y limpiar un poquito, muchas veces.

Hoy: la versión de bolsillo

El mismo razonamiento (**qué creía** + **qué vi** → **qué creo ahora**) con monedas, radares y exámenes médicos: el **Teorema de Bayes**.

Contenido

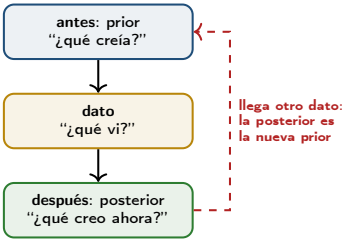
1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

El Teorema de Bayes: actualizar lo que creo con un dato

La fórmula (ya la vimos en la Clase 3)

$$\mathbb{P}(\text{hipótesis} \mid \text{dato}) = \frac{\mathbb{P}(\text{dato} \mid \text{hipótesis}) \times \mathbb{P}(\text{hipótesis})}{\mathbb{P}(\text{dato})}$$

- $\mathbb{P}(\text{hipótesis})$: la **prior**, lo que creía **antes**.
- $\mathbb{P}(\text{dato} \mid \text{hipótesis})$: la **likelihood** (en español, verosimilitud).
- $\mathbb{P}(\text{hipótesis} \mid \text{dato})$: la **posterior**, lo que creo **después**.
- $\mathbb{P}(\text{dato})$: la suma de prior $\times$ likelihood sobre todas las hipótesis (total probability). Solo sirve para que todo sume 1.



Lo nuevo de la Unit 7

Lo que no conocemos (qué moneda es, qué enfermedad, qué objeto) se trata como algo **aleatorio** con su prior. Inferir = calcular la posterior.

La receta más útil del día: la tabla de Bayes

Cinco columnas, siempre las mismas

1. **Hypothesis**: las posibles respuestas a "¿qué está pasando?" (los valores de Θ).
2. **Prior**: cuánto creía en cada una **antes** de ver el dato.
3. **Likelihood**: si esa hipótesis fuera cierta, ¿qué tan probable era ver **este** dato? $\mathbb{P}(\text{dato} \mid \text{hipótesis})$.
4. **Prior $\times$ likelihood**: multiplicar las columnas 2 y 3.
5. **Posterior**: (prior $\times$ likelihood) $\div$ (suma de la columna 4).

La suma de la columna 4 es $\mathbb{P}(\text{dato})$ (total probability).
No hay que memorizar Bayes: **la tabla es Bayes**.

Antes de llenarla: cuatro preguntas

1. ¿Qué **no sé** y quiero averiguar? $\rightarrow$ las hipótesis.
2. ¿Qué **sabía antes**? $\rightarrow$ la prior.
3. ¿Qué **observé**? $\rightarrow$ el dato.
4. ¿Cómo se comporta el dato bajo cada hipótesis? $\rightarrow$ la likelihood.

Ojo

La likelihood **no** es la respuesta:
$\mathbb{P}(\text{dato} \mid \text{hipótesis})$ va al revés de lo que queremos, $\mathbb{P}(\text{hipótesis} \mid \text{dato})$. La tabla da la vuelta.

Ejemplo de la Lecture 14: ¿avión o pájaro en el radar?

La historia

Un radar de aeropuerto detecta **un objeto** en el cielo. No sabemos qué es: puede ser un **avión** o un **pájaro**. Por los registros del aeropuerto, el 20 por ciento de los objetos detectados son aviones. El radar devuelve una señal **fuerte** o **débil**: un avión (grande, metálico) da señal fuerte el 90 por ciento de las veces; un pájaro, solo el 30 por ciento. **Hoy llegó una señal fuerte.** ¿Es un avión?

Se pide

$\mathbb{P}(\text{avión} \mid \text{señal fuerte})$, y decidir (MAP): ¿avisamos "avión" o "pájaro"?

Traducir a notación

frase	notación
"puede ser avión o pájaro"	hipótesis: avión o pájaro
"el 20 por ciento son aviones"	prior $\mathbb{P}(\text{avión}) = 0{,}2$, $\mathbb{P}(\text{pájaro}) = 0{,}8$
"hoy llegó señal fuerte"	dato $X = \text{fuerte}$
"un avión da señal fuerte el 90 por ciento"	$\mathbb{P}(\text{fuerte} \mid \text{avión}) = 0{,}9$
"un pájaro, el 30 por ciento"	$\mathbb{P}(\text{fuerte} \mid \text{pájaro}) = 0{,}3$

Intuición antes de calcular

La señal fuerte apunta a avión (0,9 contra 0,3)... pero hay **cuatro pájaros por cada avión**. ¿Quién gana?

Solución del radar con la tabla de Bayes

La tabla (dato: señal fuerte)

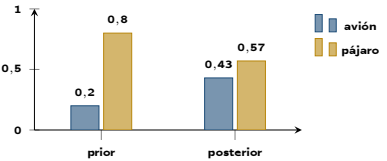
hypothesis	prior	$\mathbb{P}(\text{fuerte} \mid \cdot)$	prior $\times$likelihood	posterior
avión	0,2	0,9	0,18	$\frac{0,18}{0,42} = \frac{3}{7}$
pájaro	0,8	0,3	0,24	$\frac{0,24}{0,42} = \frac{4}{7}$
		suma $= \mathbb{P}(\text{fuerte})$	**0,42**	1

Respuesta

$\mathbb{P}(\text{avión} \mid \text{fuerte}) = \frac{3}{7} \approx 0,43$. MAP: la mayor posterior es **pájaro** ($\frac{4}{7}$).

Con 100 objetos (para verlo con personas, no fórmulas)

20 aviones $\to$ 18 dan señal fuerte. 80 pájaros $\to$ 24 dan señal fuerte. De las 42 señales fuertes, 18 son aviones: $\frac{18}{42} = \frac{3}{7}$.



Lectura

La señal fuerte **sube** avión de 0,2 a 0,43, pero no alcanza: hay tantos pájaros que también hay más pájaros con señal fuerte (24 contra 18).

El examen médico del video, con la tabla de Bayes

Datos: enfermedad en 0,1 por ciento; el examen acierta 99 por ciento

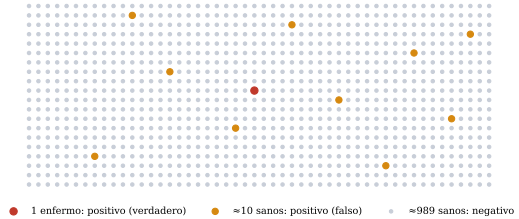
	prior	$\mathbb{P}(+\mid\cdot)$	prior $\times$likelihood	posterior
enfermo	0,001	0,99	0,00099	$\approx 0{,}09$
sano	0,999	0,01	0,00999	$\approx 0{,}91$
		suma $= \mathbb{P}(+)$	0,01098	1

$$\mathbb{P}(\text{enfermo}\mid+) = \frac{0{,}00099}{0{,}01098} = \frac{11}{122} \approx 9 \text{ por ciento.}$$

Un segundo examen positivo

La posterior 0,09 pasa a ser la **prior**. Misma tabla:

$$\frac{0{,}09 \cdot 0{,}99}{0{,}09 \cdot 0{,}99 + 0{,}91 \cdot 0{,}01} \approx 0{,}91.$$



En 1000 personas: los positivos son 1 enfermo y unos 10 sanos. Si saliste positivo eres uno de esos 11: probabilidad $\approx \frac{1}{11}$.

La trampa

Confundir $\mathbb{P}(+\mid\text{enfermo}) = 0{,}99$ con $\mathbb{P}(\text{enfermo}\mid+) \approx 0{,}09$. La **prior** (0,1 por ciento) pesa muchísimo: hay muchos más sanos que pueden dar falso positivo.

Contenido

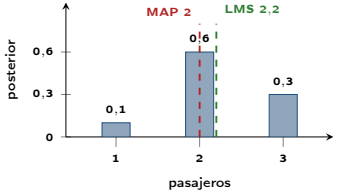
1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

MAP y LMS: dos formas de resumir la posterior

Ejemplo

Un sensor en la puerta de un auto intenta adivinar cuántos pasajeros hay (1, 2 o 3). Después de hacer la tabla de Bayes con su lectura, la posterior es:

pasajeros	1	2	3
posterior	0,1	0,6	0,3



MAP = la barra más alta. LMS = el punto de equilibrio de las barras.

MAP (*maximum a posteriori*): la más probable

La barra más alta: **2 pasajeros**. Se usa para **adivinar/decidir**: es la regla que menos se equivoca (aquí se equivoca con probabilidad 0,4).

LMS (*least mean squares*): el promedio

$1(0{,}1) + 2(0{,}6) + 3(0{,}3) = 2{,}2$ pasajeros. Se usa para **estimar un número**: es el que tiene menor error cuadrático promedio.

No siempre coinciden

2,2 pasajeros no existe: para **decidir** entre opciones se usa MAP. Si la posterior es simétrica (como una campana), MAP y LMS dan lo mismo.

Dos ideas más: actualizar en cadena y predecir

Actualizar en cadena (radar, segunda señal fuerte)

Tras la primera señal: avión $\frac{3}{7}$, pájaro $\frac{4}{7}$. Esa posterior es la **nueva prior**:

	prior	$\mathbb{P}(\text{fuerte} \mid \cdot)$	prior $\times$ likelihood	posterior
avión	$\frac{3}{7}$	0,9	$\frac{2,7}{7}$	$\frac{2,7}{3,9} = \frac{9}{13}$
pájaro	$\frac{4}{7}$	0,3	$\frac{1,2}{7}$	$\frac{1,2}{3,9} = \frac{4}{13}$

Con dos señales fuertes ahora gana **avión** ($\approx 0{,}69$).

Predecir el próximo dato

¿Qué tan probable es que la **próxima** señal sea fuerte, sabiendo que la primera lo fue? Promediar sobre las hipótesis con la posterior:

$$\frac{3}{7} \cdot 0{,}9 + \frac{4}{7} \cdot 0{,}3 = \frac{3{,}9}{7} \approx 0{,}56.$$

(Es la suma de la columna "prior $\times$ likelihood" de la tabla de la izquierda.)

Las dos reglas

1) La posterior de hoy es la prior de mañana. **2)** Para predecir: $\sum$ (posterior $\times$ probabilidad del dato bajo cada hipótesis).

Contenido

1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

Machine Learning es inferencia: datos → posterior

El diccionario: lo de hoy con otro nombre

Clase 9	Machine Learning
lo desconocido	parámetros del modelo
el dato	el *dataset*
prior	regularización
MAP	entrenar el modelo
LMS	predecir con poco error
ancho de la posterior	incertidumbre
tabla de Bayes	clasificador *Naive Bayes*

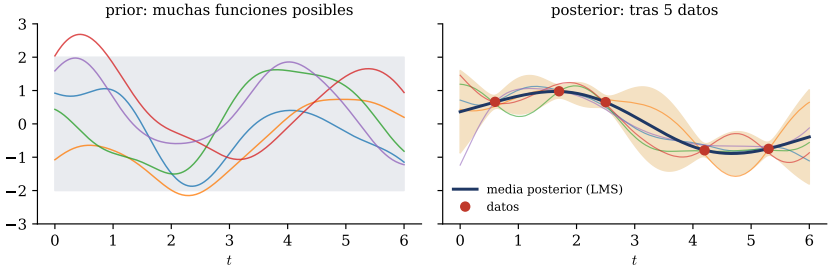
Naive Bayes: el filtro de spam es una tabla de Bayes

Hipótesis: **spam** o **normal**. Prior: qué fracción de correos es spam. Dato: las palabras del correo ("gratis", "ganaste"...). Likelihood: qué tan seguido aparece cada palabra en spam y en correos normales. Posterior → MAP: ¿a la carpeta de spam o no?

La prior en ML

Con pocos datos, la prior (lo que ya sabemos) evita conclusiones locas. Con muchos datos, los datos mandan.

Gaussian Processes: la posterior dibujada



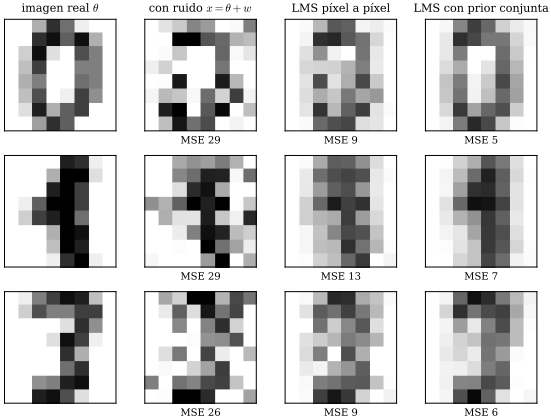
Qué se ve

Queremos adivinar una **curva** desconocida. Prior (izquierda): muchas curvas posibles. Tras ver 5 puntos (derecha), solo sobreviven las curvas que pasan cerca de ellos: esa es la posterior. La línea gruesa es su promedio (LMS).

El modelo sabe dónde no sabe

Cerca de los datos la banda es **angosta** (hay evidencia); lejos de los datos solo queda la prior y la banda es **ancha**.

Quitar ruido a una imagen = el estimador LMS



Dígitos escritos a mano (8×8 píxeles), datos reales. La prior se aprendió de 1794 imágenes.

El problema

Lo desconocido: la imagen limpia. El dato: la foto con ruido. Respuesta LMS: el **promedio** de las imágenes limpias que podrían haber producido esa foto.

Dos priors

- **Píxel a píxel** (cada píxel por su lado, prior = el "dígito promedio"): el error baja de ≈ 28 a ≈ 10.
- **Prior que sabe que los píxeles vecinos se parecen**: error ≈ 6.

La moraleja

Mejor prior $\Rightarrow$ **mejor estimación**. Un modelo de difusión es esto mismo, con una red neuronal que aprendió la prior de millones de imágenes.

Para verlo con dibujos (videos recomendados)

Bayes con áreas (3Blue1Brown, 15 min)

Bayes theorem, the geometry of changing beliefs.
`youtube.com/watch?v=HZGCoVF3YvM`

La posterior como la **fracción de un área**: la prior recorta una franja y la likelihood se queda con una parte de cada franja. Es exactamente la tabla de Bayes.

Convolución (3Blue1Brown, repaso de la Clase 8)

Convolutions: Why $X + Y$ in probability is a beautiful mess.
`3blue1brown.com/lessons/convolutions2`

La densidad de la suma como una ventana que se **desliza** sobre la otra densidad: la misma idea de la fórmula de los límites.

Sugerencia

Ver el de Bayes antes de los problemas 1 y 2; el de convolución, si el repaso de la Clase 8 aún no quedó claro.

Contenido

1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

Todo lo que necesitas para los 5 problemas

1. La tabla de Bayes

hypothesis	prior	likelihood	prior ×likelihood	posterior
H_1	$\mathbb{P}(H_1)$	$\mathbb{P}(\text{dato} \mid H_1)$	multiplicar	÷ suma
H_2	$\mathbb{P}(H_2)$	$\mathbb{P}(\text{dato} \mid H_2)$	multiplicar	÷ suma
		suma	$\mathbb{P}(\text{dato})$	1

2. Resumir la posterior

MAP = la hipótesis con mayor posterior (para **decidir**).
LMS = $\sum$ valor × posterior (para **estimar un número**).

3. Actualizar en cadena

La posterior de un dato es la **prior** del siguiente. (Da lo mismo que usar los dos datos a la vez.)

4. Predecir el próximo dato

$\mathbb{P}(\text{próximo dato}) = \sum \text{posterior} \times \mathbb{P}(\text{dato} \mid \text{hipótesis})$.

Truco: imaginar 100 (o 1000) casos

Convertir las probabilidades en **cantidades de personas, piezas o monedas** y contar. Es la misma tabla, sin fracciones.

Problema 1 (★ muy fácil): ¿qué moneda tengo?

En una bolsa hay dos monedas: una **justa** ($\mathbb{P}(\text{cara}) = \frac{1}{2}$) y una **cargada** ($\mathbb{P}(\text{cara}) = \frac{3}{4}$). Saco una al azar y la lanzo una vez: sale **cara**.

a. ¿Cuál es la probabilidad de que sea la moneda cargada?

b. Si tengo que adivinar qué moneda es (regla MAP), ¿qué digo? ¿Con qué probabilidad me equivoco?

Problema 2 (★★ fácil): ¿qué máquina hizo la pieza defectuosa?

Una fábrica tiene dos máquinas. La máquina A hace el 60 por ciento de las piezas y la B el 40 por ciento. El 2 por ciento de las piezas de A sale defectuoso; de B, el 5 por ciento. Un inspector encuentra **una pieza defectuosa**.

a. ¿Cuál es la probabilidad de que la haya hecho la máquina A? ¿Y la B?

b. Si hay que mandar a revisar una sola máquina (regla MAP), ¿cuál?

Problema 3 (★★★ intermedio): MAP frente a LMS

Una moneda puede ser de tres tipos: sale cara con probabilidad $\theta = 0{,}2$, $0{,}5$ u $0{,}8$. Antes de lanzarla creemos que es de cada tipo con probabilidad $\frac{1}{4}$, $\frac{1}{2}$ y $\frac{1}{4}$. Se lanza **dos** veces y salen **dos caras**.

a. Halla la posterior de θ (la probabilidad de cada tipo dado "dos caras").

b. Calcula el MAP y el LMS. ¿Coinciden? ¿Cuál usarías para *adivinar el sesgo exacto* y cuál para *estimarlo con poco error cuadrático*?

Problema 4 (★★★★ intermedio-alto): dos caras seguidas

Como en el problema 1: en una bolsa hay una moneda **justa** ($\mathbb{P}(\text{cara}) = \frac{1}{2}$) y una **cargada** ($\mathbb{P}(\text{cara}) = \frac{3}{4}$). Saco una al azar y la lanzo **dos** veces: sale **cara y cara**.

a. Actualiza en cadena: la posterior tras la primera cara es la prior de la segunda. ¿$\mathbb{P}(\text{cargada} \mid 2 \text{ caras})$?

b. Comprueba que da lo mismo usar las dos caras de una vez.

c. ¿Cuál es la probabilidad de que el **tercer** tiro salga cara?

Problema 5 (★★★★★ reto): adivinar la moneda con 3 tiros

Adaptado de "Hypothesis test between two coins" (Problem Set 7a, MIT 6.431x). Alicia tiene dos monedas que se ven iguales: la moneda 1 sale cara con probabilidad $\frac{1}{4}$ y la moneda 2 con $\frac{3}{4}$. Elige la moneda 1 con probabilidad $\frac{3}{4}$ y se la da a Bob, que la lanza **3 veces** y cuenta k caras ($k = 0, 1, 2, 3$).

a. Para cada k, ¿qué moneda debe decir Bob (regla MAP)?

b. ¿Con qué probabilidad acierta Bob? Compáralo con adivinar sin lanzar.

Contenido

1. Repaso de la Clase 8: la convolución sin complicarse
2. ¿Por qué Bayes? Inferencia bayesiana en el mundo real
3. El Teorema de Bayes y la tabla de Bayes
4. Responder con un solo número: MAP y LMS
5. Bayes en Machine Learning
6. Cinco problemas, de menos a más
7. Cierre

Resumen de la Clase 9

Lo que hay que saber hacer

1. **Traducir** el enunciado: hipótesis, prior, dato y likelihood.
2. Llenar la **tabla de Bayes**: prior $\times$ likelihood, sumar, dividir.
3. **MAP** = la hipótesis más probable (para decidir); **LMS** = el promedio (para estimar un número).
4. **Actualizar en cadena**: la posterior de hoy es la prior de mañana.
5. **Predecir**: promediar sobre las hipótesis con la posterior.

Errores clásicos

Confundir $\mathbb{P}(\text{dato} \mid \text{hipótesis})$ con $\mathbb{P}(\text{hipótesis} \mid \text{dato})$ (la trampa del examen médico) · olvidar la prior · predecir con una sola hipótesis en vez de promediar.

Siguiente clase (Clase 10)

El estimador **LMS** a fondo y el estimador lineal **LLMS**. Para quien quiera más: la posterior de una moneda con sesgo continuo y el modelo normal de la Lecture 15 están en el documento de teoría.

¡Gracias!

Clase 9 — Introducción a la inferencia bayesiana (6.431x, Unit 7)

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca · en alianza con MIT IDSS

Basado en MIT 6.431x / MIT OCW RES.6-012 (Bertsekas & Tsitsiklis), CC BY-NC-SA 4.0.