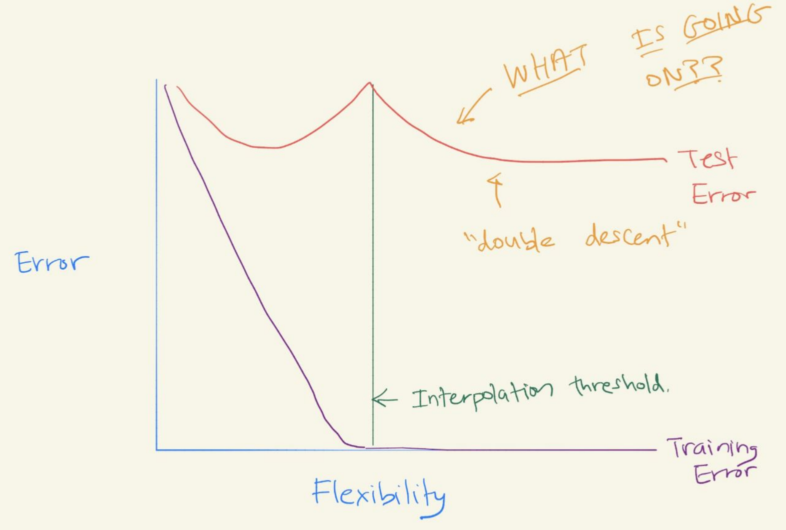
Ventajas y límites de la inferencia exacta

Semana 2

Aplicación estricta de las reglas de la probabilidad. Ejemplos Gaussian y Beta-Binomial El problema histórico de la probabilidad: el costo computacional. Efectos secundarios de la ruptura de las reglas de la probabilidad: el overfitting. Evaluación de modelos polinomiales de complejidad creciente. No identificabilidad de modelos



El costo computacional
Problema histórico de la probabilidad

La **aplicación estricta** de las reglas de la probabilidad obligan a **evaluar todo el espacio de hipótesis**.

El costo computacional

Problema histórico de la probabilidad

La **aplicación estricta** de las reglas de la probabilidad obligan a **evaluar todo el espacio de hipótesis**.

- Siglo 18: Nace la probabilidad (soluciones analíticas).

El costo computacional

Problema histórico de la probabilidad

La **aplicación estricta** de las reglas de la probabilidad obligan a **evaluar todo el espacio de hipótesis**.

- Siglo 18: Nace la probabilidad (soluciones analíticas).
- Siglo 19: Física estadística (contar estados).

El costo computacional
Problema histórico de la probabilidad

La **aplicación estricta** de las reglas de la probabilidad obligan a **evaluar todo el espacio de hipótesis**.

- Siglo 18: Nace la probabilidad (soluciones analíticas).
- Siglo 19: Física estadística (contar estados).
- Siglo 20: Estimadores puntuales (evitan evaluar todo el espacio de hipótesis).

El costo computacional

Problema histórico de la probabilidad

La **aplicación estricta** de las reglas de la probabilidad obligan a **evaluar todo el espacio de hipótesis**.

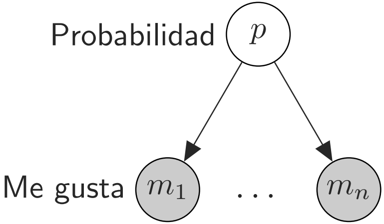
- Siglo 18: Nace la probabilidad (soluciones analíticas).
- Siglo 19: Física estadística (contar estados).
- Siglo 20: Estimadores puntuales (evitan evaluar todo el espacio de hipótesis).
- Siglo 21: Comienza a ser posible evaluar todo el espacio de hipótesis mediante la (aproximación a) la aplicación estricta de las reglas de la probabilidad de forma general

Siglo 18: la primera solución analítica.

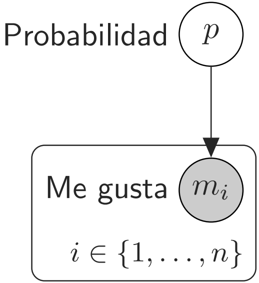
¿A quién le compro?

- 100% Me gusta de 10 Ratings
- 96% Me gusta de 50 Ratings
- 93% Me gusta de 200 Ratings

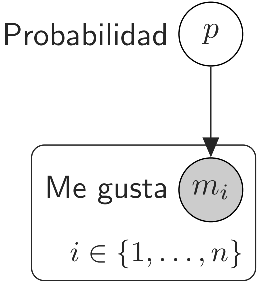
Siglo 18: la primera solución analítica.



Siglo 18: la primera solución analítica.

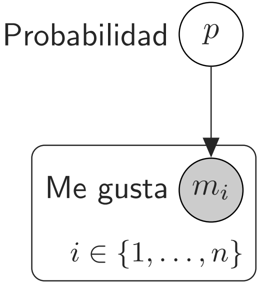


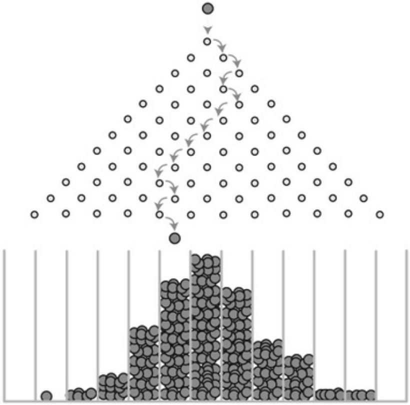
Siglo 18: la primera solución analítica.



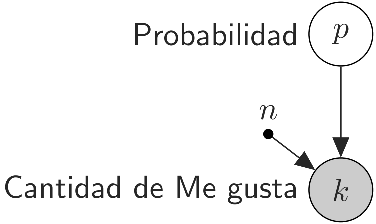
$$P(m_i \,|\, p) = p^{m_i} \, (1-p)^{1-m_i}$$

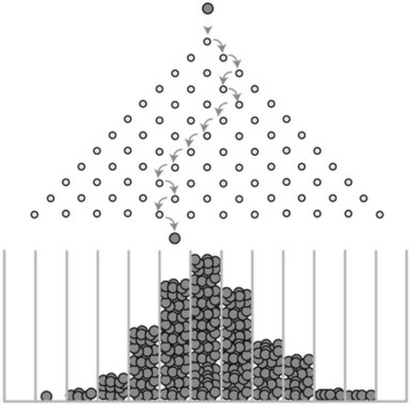
Siglo 18: la primera solución analítica.



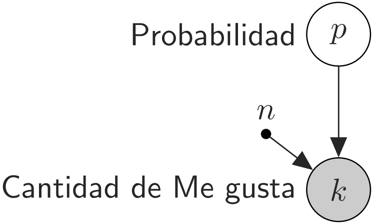


Siglo 18: la primera solución analítica.



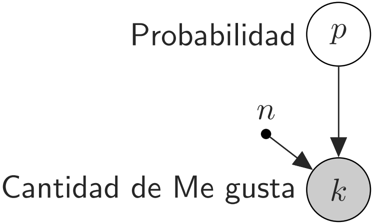


Siglo 18: la primera solución analítica.



$$P(k \,|\, p,\, n) = \frac{n!}{k!(n-k)!}\, p^k \,(1-p)^{n-k}$$

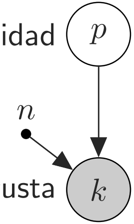
Siglo 18: la primera solución analítica.



$$P(p) = 1$$

$$P(k \mid p,\, n) = \tfrac{n!}{k!(n-k)!}\, p^k\, (1-p)^{n-k}$$

Siglo 18: la primera solución analítica.

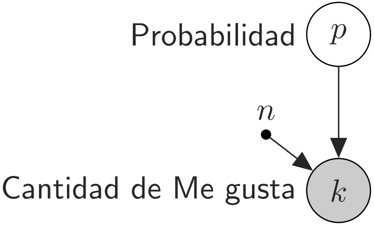


Probabilidad $\quad P(p) = 1$

Cantidad de Me gusta $\quad P(k \,|\, p,\, n) = \frac{n!}{k!(n-k)!}\, p^k\, (1-p)^{n-k}$

$$P(p \,|\, k, n) = \frac{P(k \,|\, p, n) P(p)}{P(k \,|\, n)}$$

Siglo 18: la primera solución analítica.

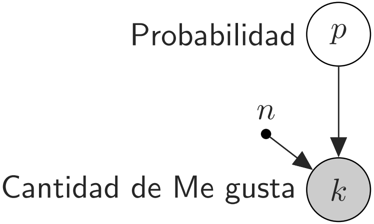


$P(p) = 1$

$P(k \,|\, p,\, n) = \frac{n!}{k!(n-k)!}\, p^k\, (1-p)^{n-k}$

$$P(p \,|\, k, n) = \frac{P(k \,|\, p, n) P(p)}{P(k \,|\, n)} = \underbrace{(n+1)}_{1/P(k \,|\, n)} \; \underbrace{\frac{n!}{k!(n-k)!}\, p^k\, (1-p)^{n-k}}_{P(k \,|\, p, n)} \; \underbrace{1}_{P(p)}$$

Siglo 18: la primera solución analítica.

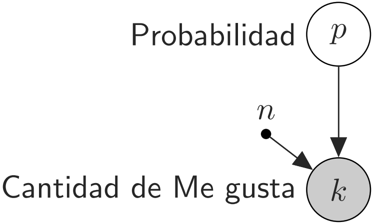


$P(p) = 1$

$P(k \,|\, p,\, n) = \frac{n!}{k!(n-k)!}\, p^k \,(1-p)^{n-k}$

$$P(p \,|\, k, n) = \frac{P(k \,|\, p, n) P(p)}{P(k \,|\, n)} = \frac{(n+1)!}{k!(n-k)!}\, p^k \,(1-p)^{n-k}$$

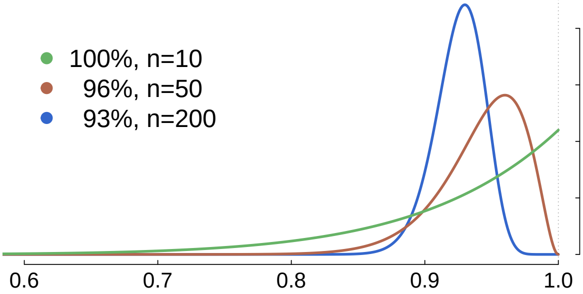
Siglo 18: la primera solución analítica.



$$P(p) = 1$$

$$P(k \,|\, p,\, n) = \frac{n!}{k!(n-k)!}\, p^k \,(1-p)^{n-k}$$

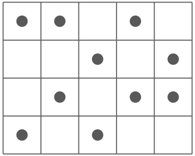
$$P(p \,|\, k, n) = \frac{P(k \,|\, p, n) P(p)}{P(k \,|\, n)}$$



Siglo 19: Física estadística

Gases

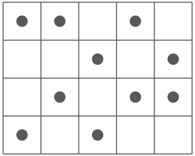
Micro Estados



$$\left.\begin{array}{l} N = 10 \\ V = 20 \end{array}\right\} \text{Macro estados}$$

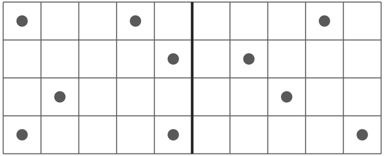
Siglo 19: Física estadística
Gases

Micro Estados



$$\left.\begin{array}{l} N = 10 \\ V = 20 \end{array}\right\} \text{Macro estados}$$

Expansión de volumen

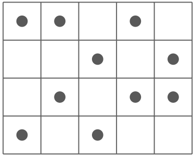


$$\begin{array}{l} N = 10 \\ V = 40 \end{array}$$

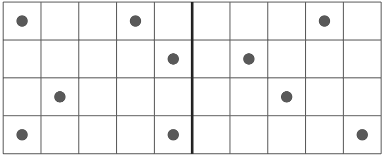
Siglo 19: Física estadística

Gases

Micro Estados



Expansión de volumen



$$\left.\begin{aligned} N &= 10 \\ V &= 20 \end{aligned}\right\} \text{Macro estados}$$

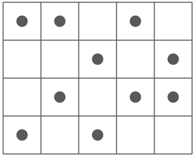
$$\begin{aligned} N &= 10 \\ V &= 40 \end{aligned}$$

¿Cuál es la probabilidad de que se contraiga nuevamente?

Siglo 19: Física estadística

Gases

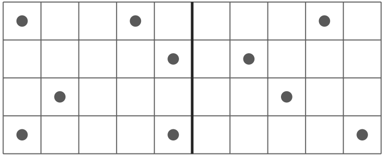
Micro Estados



$$\left.\begin{array}{l} N = 10 \\ V = 20 \end{array}\right\} \text{Macro estados}$$

$$W = \binom{20}{10} = 184\,756$$

Expansión de volumen

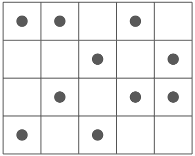


$$\begin{array}{l} N = 10 \\ V = 40 \end{array}$$

Siglo 19: Física estadística

Gases

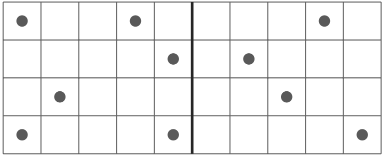
Micro Estados



$$\left.\begin{array}{l} N = 10 \\ V = 20 \end{array}\right\} \text{Macro estados}$$

$$W = \binom{20}{10} = 184\,756$$

Expansión de volumen

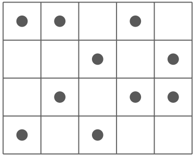


$$\begin{array}{l} N = 10 \\ V = 40 \end{array}$$

$$W = \binom{40}{10} = 847\,660\,528$$

Siglo 19: Física estadística
Gases

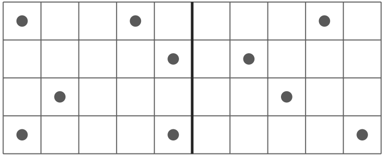
Micro Estados



$$\left.\begin{aligned} N &= 10 \\ V &= 20 \end{aligned}\right\} \text{Macro estados}$$

$$W = \binom{20}{10} = 184\,756$$

Expansión de volumen



$$\begin{aligned} N &= 10 \\ V &= 40 \end{aligned}$$

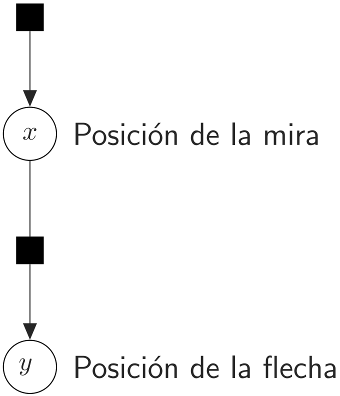
$$W = \binom{40}{10} = 847\,660\,528$$

$$P(\text{contraiga}|N = 10, V = 40) = \frac{\binom{20}{10}}{\binom{40}{10}} = 0.0002$$

Modelos gaussianos

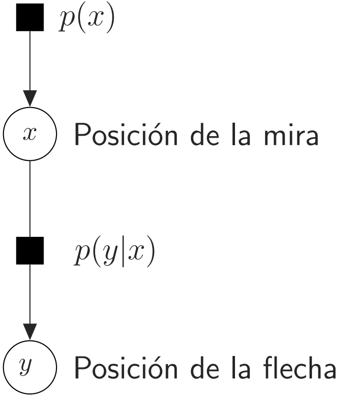
Modelos gaussianos

Arco y flecha | Habilidad y desempeño



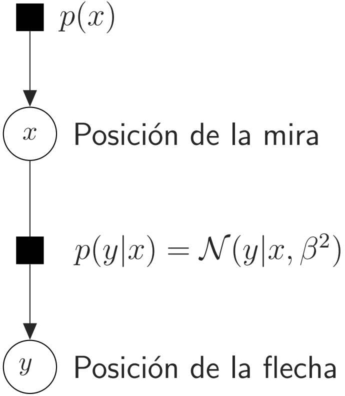
Modelos gaussianos

Arco y flecha | Habilidad y desempeño



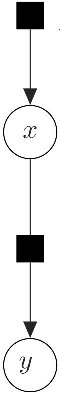
Modelos gaussianos

Arco y flecha | Habilidad y desempeño



Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu, \sigma^2)$

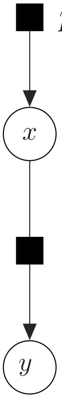
Posición de la mira

$p(y|x) = \mathcal{N}(y|x, \beta^2)$

Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

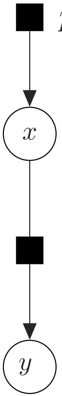
Posición de la mira

$$p(y|x) = \mathcal{N}(y|x, \beta^2)$$

Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño

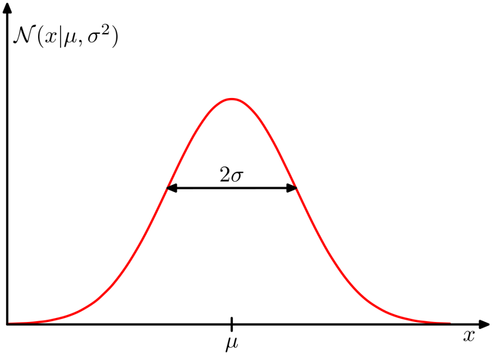


$$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Posición de la mira

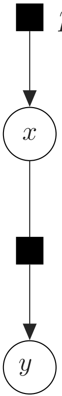
$$p(y|x) = \mathcal{N}(y|x, \beta^2)$$

Posición de la flecha



Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

x Posición de la mira

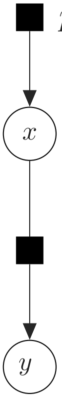
Predicción de $y :=$

$p(y|x) = \mathcal{N}(y|x, \beta^2)$

y Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

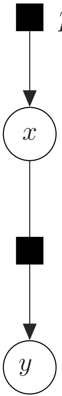
Predicción de $y := p(y)$

$p(y|x) = \mathcal{N}(y|x,\beta^2)$

Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Posición de la mira

Predicción de $y := p(y) = \int \mathcal{N}(y|x, \beta^2)\mathcal{N}(x|\mu, \sigma^2)dx$

$$p(y|x) = \mathcal{N}(y|x, \beta^2)$$

Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño

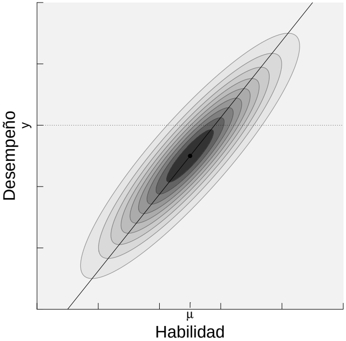
$$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

x Posición de la mira

$$p(y|x) = \mathcal{N}(y|x, \beta^2)$$

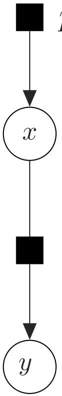
y Posición de la flecha

Predicción de $y := p(y) = \int \mathcal{N}(y|x, \beta^2)\mathcal{N}(x|\mu, \sigma^2)dx$



Modelos gaussianos

Arco y flecha | Habilidad y desempeño



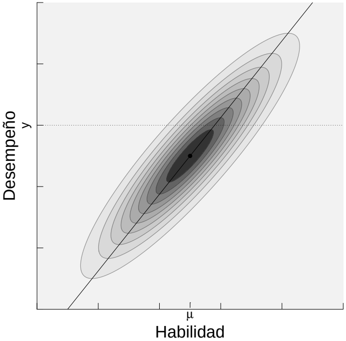
$$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

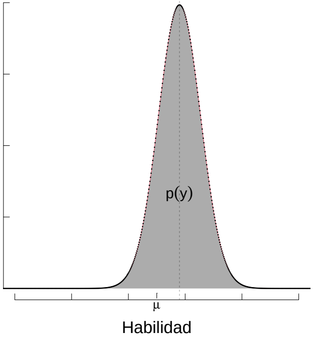
Posición de la mira

$$p(y|x) = \mathcal{N}(y|x, \beta^2)$$

Posición de la flecha

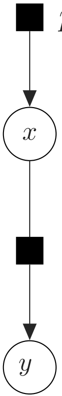
Predicción de $y := p(y) = \int \mathcal{N}(y|x, \beta^2)\mathcal{N}(x|\mu, \sigma^2)dx$





Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

$p(y|x) = \mathcal{N}(y|x, \beta^2)$

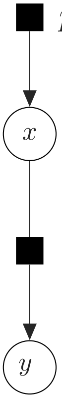
Posición de la flecha

$$\text{Predicción de } y := p(y) = \int \mathcal{N}(y|x, \beta^2)\mathcal{N}(x|\mu, \sigma^2)dx$$

$$\overset{*}{=} \int \mathcal{N}(y|\mu, \beta^2 + \sigma^2)\, \mathcal{N}(x|\mu_*, \sigma_*^2)\, dx$$

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\beta}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

$p(y|x) = \mathcal{N}(y|x,\beta^2)$

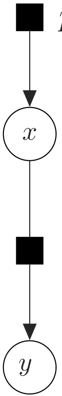
Posición de la flecha

Predicción de $y := p(y) = \int \mathcal{N}(y|x,\beta^2)\mathcal{N}(x|\mu,\sigma^2)dx$

$$\overset{*}{=} \int \underbrace{\mathcal{N}(y|\mu,\beta^2+\sigma^2)}_{\text{const}}\mathcal{N}(x|\mu_*,\sigma_*^2)\,dx$$

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

$p(y|x) = \mathcal{N}(y|x,\beta^2)$

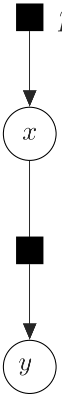
Posición de la flecha

$$\text{Predicción de } y := p(y) = \int \mathcal{N}(y|x,\beta^2)\mathcal{N}(x|\mu,\sigma^2)dx$$

$$\overset{*}{=} \int \underbrace{\mathcal{N}(y|\mu,\beta^2+\sigma^2)}_{\text{const}}\underbrace{\mathcal{N}(x|\mu_*,\sigma_*^2)}_{1}\,dx$$

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

$p(y|x) = \mathcal{N}(y|x, \beta^2)$

Posición de la flecha

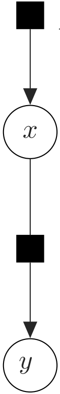
$$\text{Predicción de } y := p(y) = \int \mathcal{N}(y|x,\beta^2)\mathcal{N}(x|\mu,\sigma^2)dx$$

$$\overset{*}{=} \int \underbrace{\mathcal{N}(y|\mu,\beta^2+\sigma^2)}_{\text{const}} \underbrace{\mathcal{N}(x|\mu_*,\sigma_*^2)}_{1} dx$$

$$= \mathcal{N}(y|\mu,\beta^2+\sigma^2)$$

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

Predicción de $y := p(y) = \int \mathcal{N}(y|x, \beta^2)\mathcal{N}(x|\mu, \sigma^2)dx$

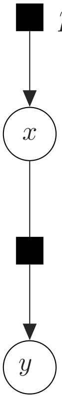
$p(y|x) = \mathcal{N}(y|x, \beta^2)$

Posterior de $x = p(x|y)$

Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

Predicción de $y := p(y) = \displaystyle\int \mathcal{N}(y|x, \beta^2)\mathcal{N}(x|\mu, \sigma^2)dx$

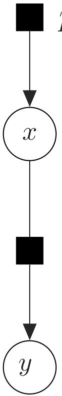
$p(y|x) = \mathcal{N}(y|x, \beta^2)$

Posterior de $x = p(x|y) \propto \mathcal{N}(x|\mu, \sigma^2)\mathcal{N}(y|x, \beta^2)$

Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

Predicción de $y := p(y) = \int \mathcal{N}(y|x,\beta^2)\mathcal{N}(x|\mu,\sigma^2)dx$

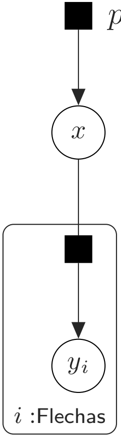
$p(y|x) = \mathcal{N}(y|x,\beta^2)$

Posterior de $x = p(x|y) \propto \mathcal{N}(x|\mu,\sigma^2)\mathcal{N}(y|x,\beta^2)$

Posición de la flecha

Modelos gaussianos

Arco y flecha | Habilidad y desempeño



$p(x) = \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$

Posición de la mira

Predicción de $y := p(y) = \int \mathcal{N}(y|x, \beta^2)\mathcal{N}(x|\mu, \sigma^2)dx$

$p(y|x) = \mathcal{N}(y|x, \beta^2)$

Posterior de $x = p(x|y) \propto \mathcal{N}(x|\mu, \sigma^2)\mathcal{N}(y|x, \beta^2)$

Posición de la flecha

Modelos lineales
(lineales en las hipótesis)

Modelos lineales
(lineales en las hipótesis)

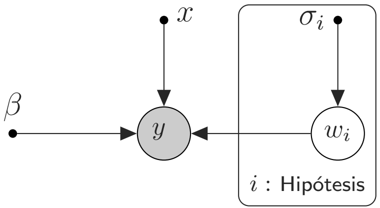
$$p(y|x, \boldsymbol{w}, \beta) = \mathcal{N}\left(y \,\middle|\, \sum_i w_i \cdot \phi_i(x)\,,\ \beta^2\right)$$

Modelos lineales
(lineales en las hipótesis)

$$p(y|x,\boldsymbol{w},\beta) = \mathcal{N}\left(y \,\middle|\, \sum_i w_i \cdot \phi_i(x)\,,\ \beta^2\right) \qquad p(w_i) = \mathcal{N}(w_i \,|\, 0, \sigma_i^2)$$

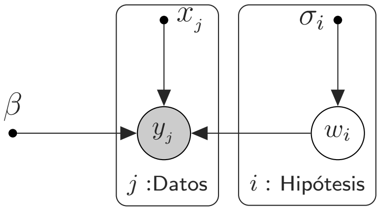
Modelos lineales
(lineales en las hipótesis)

$$p(y|x, \boldsymbol{w}, \beta) = \mathcal{N}\left(y \,\middle|\, \sum_i w_i \cdot \phi_i(x)\,,\ \beta^2\right) \qquad p(w_i) = \mathcal{N}(w_i \,|\, 0, \sigma_i^2)$$



Modelos lineales
(lineales en las hipótesis)

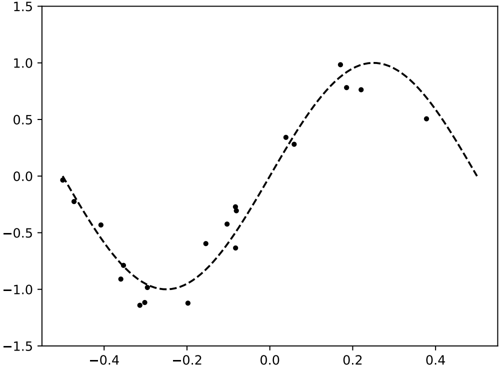
$$p(y|x,\boldsymbol{w},\beta) = \mathcal{N}\left(y \,\middle|\, \sum_i w_i \cdot \phi_i(x)\,,\ \beta^2\right) \qquad p(w_i) = \mathcal{N}(w_i\,|\,0,\sigma_i^2)$$



Modelos lineales

Función objetivo

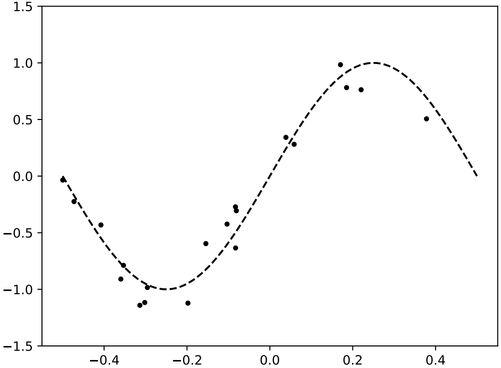
$$\mathcal{N}(y \,|\, \mathsf{sen}(2\pi x), \beta^2) \quad x \in [-\tfrac{1}{2}, \tfrac{1}{2}]$$



Modelos lineales

Función objetivo

$\mathcal{N}(y \,|\, \mathsf{sen}(2\pi x), \beta^2) \quad x \in [-\frac{1}{2}, \frac{1}{2}]$

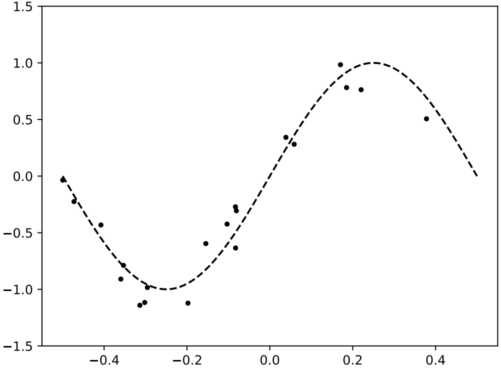


$$y = w_0 + w_1\, x + w_2\, x^2 + w_3\, x^3 + \ldots$$

Modelos lineales

Función objetivo

$\mathcal{N}(y \,|\, \mathrm{sen}(2\pi x), \beta^2) \quad x \in [-\frac{1}{2}, \frac{1}{2}]$



$$y = w_0 + w_1\, x + w_2\, x^2 + w_3\, x^3 + \ldots$$

$$y = \sum_{i=0}^{M} w_i \cdot x^i$$

¿Cuál es el mejor modelo polinomial?

Siglo 20: Estimadores puntuales

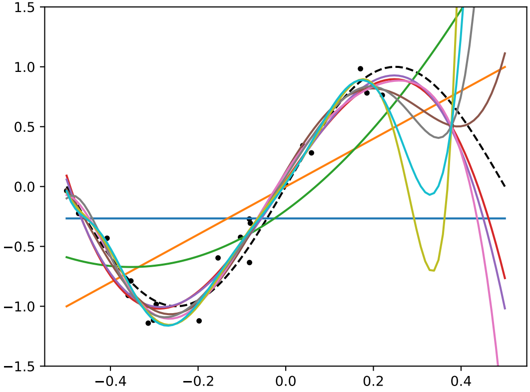
~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} P(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\beta) = \min_{\boldsymbol{w}} \sum_{i=1}^{n}(y_i - f(x_i,\boldsymbol{w}))^2$$

Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

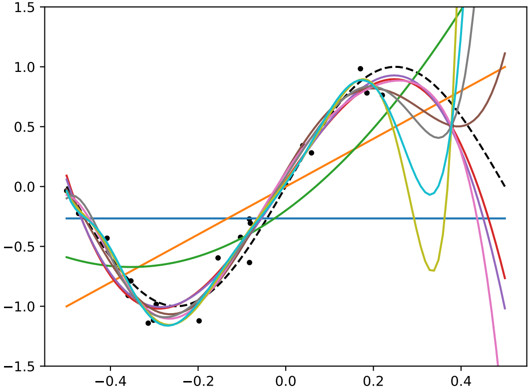
$$\max_{\boldsymbol{w}} \; P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \beta) = \; \min_{\boldsymbol{w}} \; \sum_{i=1}^{n} (y_i - f(x_i, \boldsymbol{w}))^2$$

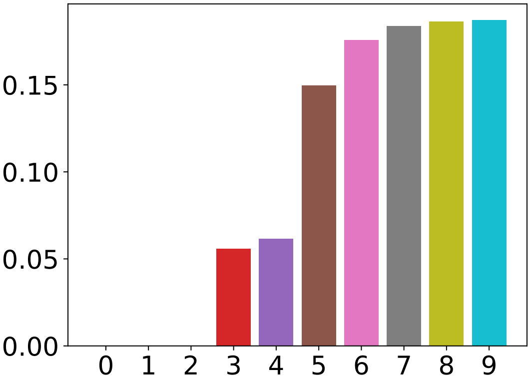


Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} P(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\beta) = \min_{\boldsymbol{w}} \sum_{i=1}^{n}(y_i - f(x_i,\boldsymbol{w}))^2$$

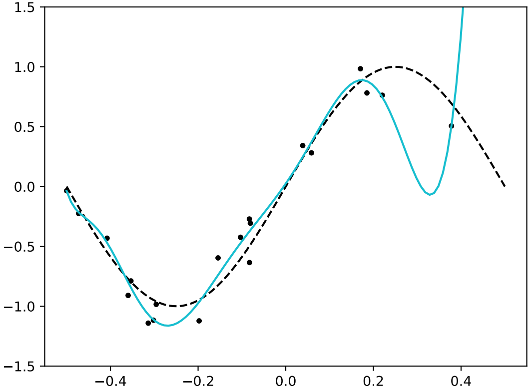


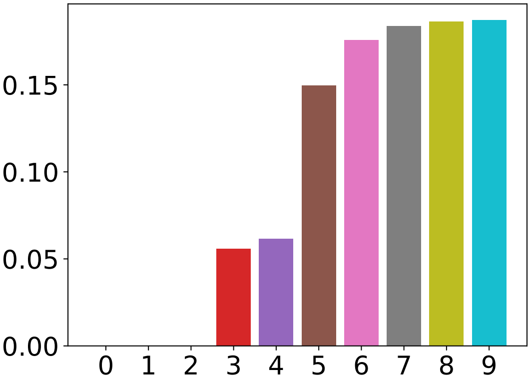


Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} \; P(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\beta) = \; \min_{\boldsymbol{w}} \; \sum_{i=1}^{n} (y_i - f(x_i,\boldsymbol{w}))^2$$





Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} P(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\beta) = \min_{\boldsymbol{w}} \sum_{i=1}^{n} (y_i - f(x_i,\boldsymbol{w}))^2$$

$$P(\text{dato}|\text{Modelo}) = \sum_{\text{hipótesis}} P(\text{dato}\,|\,\text{hipótesis},\text{Modelo})P(\text{hipótesis}|\text{Modelo})$$

Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \beta) = \min_{\boldsymbol{w}} \sum_{i=1}^{n} (y_i - f(x_i, \boldsymbol{w}))^2$$

$$P(\text{dato}|\text{Modelo}) = \qquad P(\text{dato}\,|\,\text{hipótesis}, \text{Modelo})$$

Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\underset{\boldsymbol{w}}{\text{max}}\ P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \beta) = \ \underset{\boldsymbol{w}}{\text{min}}\ \sum_{i=1}^{n}(y_i - f(x_i, \boldsymbol{w}))^2$$

$$P(\text{dato}|\text{Modelo}) = \qquad P(\text{dato}\,|\,\text{hipótesis}, \text{Modelo})$$

$$= P(\text{dato}\,|\,\overbrace{\underset{h}{\text{arg max}}\ P(\text{dato}|h, \text{Modelo})}^{\text{Hipótesis que mejor predice}}, \text{Modelo})$$

Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} P(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\beta) = \min_{\boldsymbol{w}} \sum_{i=1}^{n}(y_i - f(x_i,\boldsymbol{w}))^2$$

$$P(\text{dato}|\text{Modelo}) = \quad P(\text{dato}\,|\,\text{hipótesis},\text{Modelo})$$

$$= P(\text{dato}\,|\,\overbrace{\underset{h}{\arg\max}\ P(\text{dato}|h,\text{Modelo})}^{\text{Hipótesis que mejor predice}},\ \text{Modelo})$$

¿Predecimos o “post-decimos”?

Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} P(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\beta) = \min_{\boldsymbol{w}} \sum_{i=1}^{n}(y_i - f(x_i,\boldsymbol{w}))^2$$

$$P(\text{dato}|\text{Modelo}) = \qquad P(\text{dato}\,|\,\text{hipótesis}, \text{Modelo})$$

$$= P(\text{dato}\,|\,\overbrace{\underset{h}{\text{arg max}}\; P(\text{dato}|h,\text{Modelo})}^{\text{Hipótesis que mejor predice}},\ \text{Modelo})$$

$$P(\text{dato}_{\textbf{Testear}}|\text{Modelo}) = P(\text{dato}_{\textbf{Testear}}\,|\,\underset{h}{\text{arg max}}\; P(\text{dato}_{\textbf{Entrenar}}|h,\text{Modelo}),\ \text{Modelo})$$

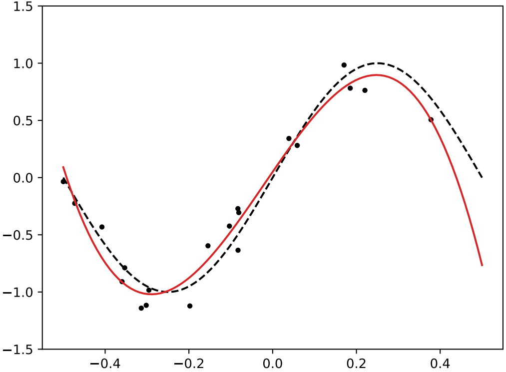
Testeo y Entrenamiento

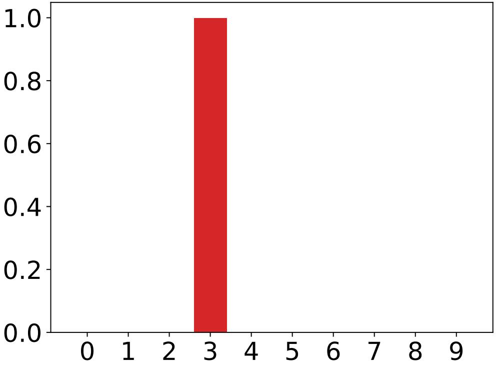
Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

$$\max_{\boldsymbol{w}} P(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\beta) = \min_{\boldsymbol{w}} \sum_{i=1}^{n} (y_i - f(x_i,\boldsymbol{w}))^2$$

Con testeo y entrenamiento



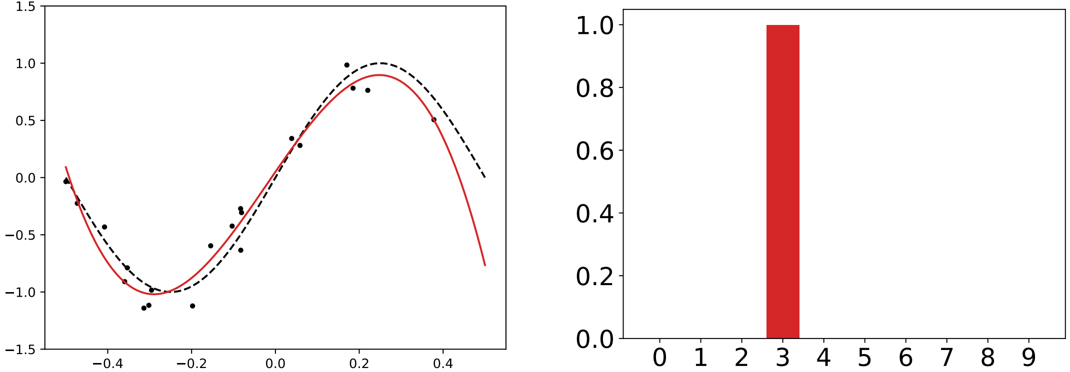


Siglo 20: Estimadores puntuales

~~Evaluación~~ Selección de hipótesis

Perio si empezamos a ver datos $x \notin [-\frac{1}{2}, \frac{1}{2}]$ este modelo no sirve!

Con testeo y entrenamiento



¿Será que la aplicación estricta de las reglas de la probabilidad produce sobreajuste (*overfitting*)?

¿Será que la aplicación estricta de las reglas de la probabilidad produce sobreajuste (*overfitting*)?

¿Y entonces el sistema de razonamiento en contextos de incertidumbre funciona mal?

Siglo 21: Inferencia exacta

Aplicación estricta de las reglas de la probabilidad

Siglo 21: Inferencia exacta

Aplicación estricta de las reglas de la probabilidad

El modelo lineal tiene solución analítica!

Siglo 21: Inferencia exacta

Aplicación estricta de las reglas de la probabilidad

El modelo lineal tiene solución analítica!

- Para evaluar las hipótesis al interior de los modelos (**posterior**),

$$P(\text{Hipótesis}|\text{Datos}, \text{Modelo})$$

Siglo 21: Inferencia exacta

Aplicación estricta de las reglas de la probabilidad

El modelo lineal tiene solución analítica!

- Para evaluar las hipótesis al interior de los modelos (**posterior**),

$$P(\text{Hipótesis}|\text{Datos}, \text{Modelo})$$

- Y para evaluar modelos alternativos (**evidencia**),

$$P(\text{Datos}|\text{Modelos})$$

Evidencia y Posterior

Dado el prior $p(\mathbf{w})$ y el likelihood $p(\mathbf{y}|\mathbf{w})$

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \qquad p(\mathbf{y}|\mathbf{w}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{w} + \mathbf{b}, \mathbf{L}^{-1})$$

Obtenemos la evidencia $p(\mathbf{y})$, y el posterior $p(\mathbf{w}|\mathbf{y})$.

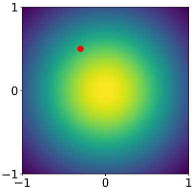
$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \;\; \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T)$$

$$p(\mathbf{w}|\mathbf{y}) = \mathcal{N}(\mathbf{w}|\boldsymbol{\Sigma}\left(\mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\right), \;\; \boldsymbol{\Sigma})$$

con $\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^T\mathbf{L}\mathbf{A})^{-1}$

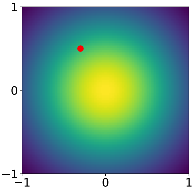
$$P(\text{Hipótesis}|\text{Datos, Modelo})$$

Priori

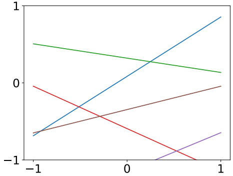


$$P(\text{Hipótesis}|\text{Datos}, \text{Modelo})$$

Priori

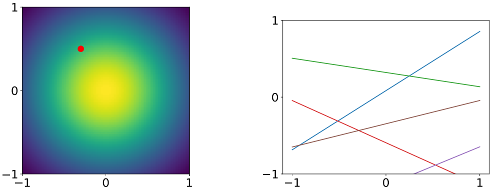


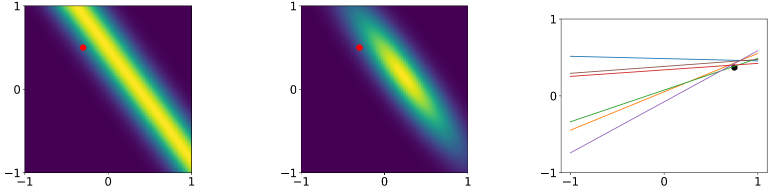
Espacio de datos



Verosimilitud | Priori/Posteriori | Espacio de datos

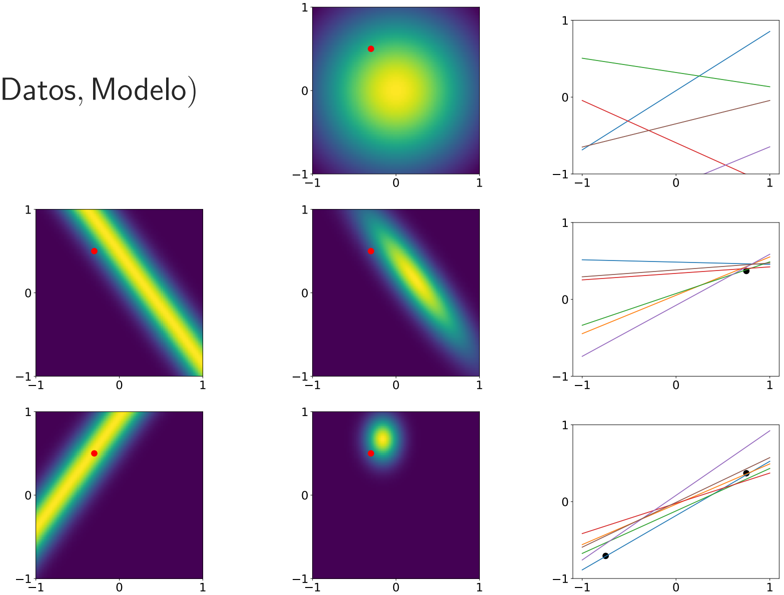
$$P(\text{Hipótesis}|\text{Datos}, \text{Modelo})$$





Verosimilitud Priori/Posteriori Espacio de datos

$$P(\text{Hipótesis}|\text{Datos, Modelo})$$



La función de costo epistémica

Todos los datos son de testeo y entrenamiento:

$$\underbrace{P(\text{Datos} = \{d_1, d_2, \dots\}|\text{Modelo})}_{\text{Evidencia: predicción del modelo}} = \underbrace{P(d_1|\text{Modelo})}_{\text{Predicción 1}} \underbrace{P(d_2|d_1, \text{Modelo})}_{\text{Predicción 2}} \dots$$

La función de costo epistémica

Todos los datos son de testeo y entrenamiento:

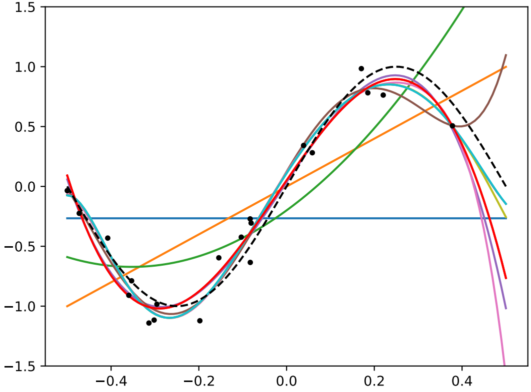
$$\underbrace{P(\text{Datos} = \{d_1, d_2, \dots\}|\text{Modelo})}_{\text{Evidencia: predicción del modelo}} = \underbrace{P(d_1|\text{Modelo})}_{\text{Predicción 1}} \underbrace{P(d_2|d_1, \text{Modelo})}_{\text{Predicción 2}} \dots$$

La predicción se hace con todas las hipótesis

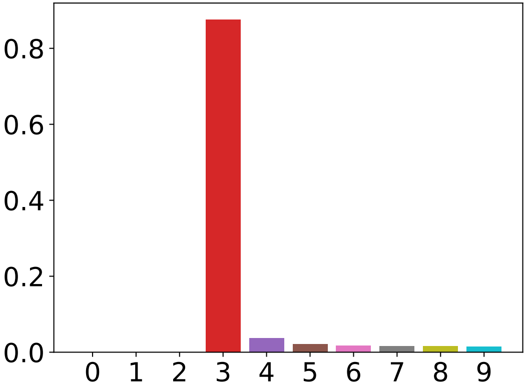
$$P(\text{dato}_1|\text{Modelo}) = \sum_{\text{hipótesis}} P(\text{dato}_1|\text{hipótesis}, \text{Modelo}) P(\text{hipótesis}|\text{Modelo})$$

Siglo 21: Inferencia exacta

Aplicación estricta de las reglas de la probabilidad

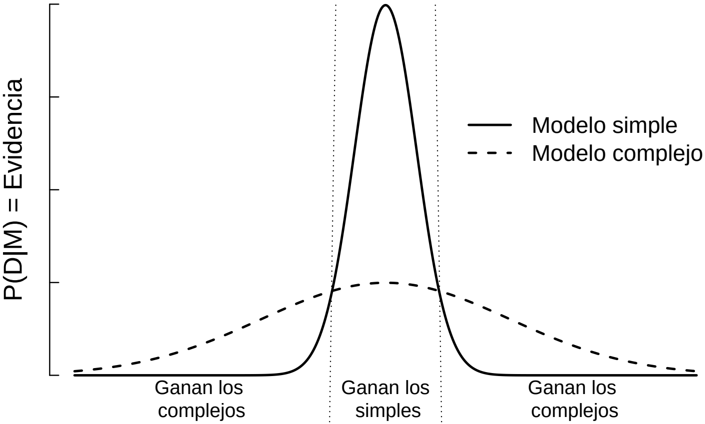


$P(\text{Modelo}|\text{Datos})$



Evidencia

Balance natural entre complejidad y predicción



La verdadera predicción

Predicción con la contribución de todos los modelos

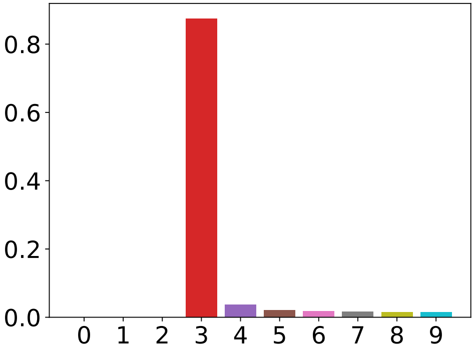
$$P(\text{Datos}) = \sum_{\text{Modelo}} P(\text{Datos}|\text{Modelo})P(\text{Modelo})$$

La verdadera predicción

Predicción con la contribución de todos los modelos

$$P(\text{Datos}) = \sum_{\text{Modelo}} P(\text{Datos}|\text{Modelo}) P(\text{Modelo})$$

Si aparecen datos $x \notin [-\frac{1}{2}, \frac{1}{2}]$ van a poder ser explicados con los modelos más complejos



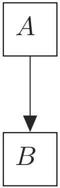
Ejercicio: Identificación de modelo causal

$P(A)$

$A=0$	$A=1$
0.5	0.5

$P(B|A)$

	$B=0$	$B=1$
$A=0$	0.95	0.05
$A=1$	0.05	0.95



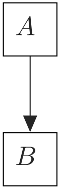
Ejercicio: Identificación de modelo causal

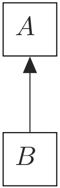
$P(A)$

$A=0$	$A=1$
0.5	0.5

$P(B|A)$

	$B=0$	$B=1$
$A=0$	0.95	0.05
$A=1$	0.05	0.95





$P(A|B)$

	$A=0$	$A=1$
$B=0$	0.95	0.05
$B=1$	0.05	0.95

$P(B)$

$B=0$	$B=1$
0.5	0.5

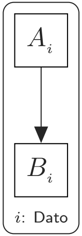
Ejercicio: Identificación de modelo causal

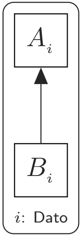
$P(A)$

$A=0$	$A=1$
0.5	0.5

$P(B|A)$

	$B=0$	$B=1$
$A=0$	0.95	0.05
$A=1$	0.05	0.95





$P(A|B)$

	$A=0$	$A=1$
$B=0$	0.95	0.05
$B=1$	0.05	0.95

$P(B)$

$B=0$	$B=1$
0.5	0.5

$$P(\text{Modelo}|\text{Datos} = \{(a_1, b_1), (a_2, b_2), \dots\}) =$$

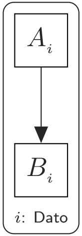
Ejercicio: Identificación de modelo causal

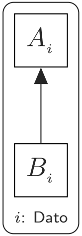
$P(A)$

$A=0$	$A=1$
0.5	0.5

$P(B|A)$

	$B=0$	$B=1$
$A=0$	0.95	0.05
$A=1$	0.05	0.95





$P(A|B)$

	$A=0$	$A=1$
$B=0$	0.95	0.05
$B=1$	0.05	0.95

$P(B)$

$B=0$	$B=1$
0.5	0.5

$$P(\text{Modelo}|\text{Datos} = \{(a_1, b_1), (a_2, b_2), \ldots\}) = \frac{\overbrace{P(\text{Datos}|\text{Modelo})}^{\text{Predicción}}\, P(\text{Modelo})}{P(\text{Datos})}$$

$$p = b$$

Laboratorios de

Métodos Bayesianos

Bibliografía Unidad 2

- Herbrich. **TrueSkill Through Time: A Bayesian Skill Rating System**. Advances in Neural Information Processing Systems. 2006. (Descargar). (lectura paper)

- MacKay. *Information theory, inference and learning algorithms.*. Cambridge university press; 2003 (Descargar). (lecturas 1.1, 2.4-6, 4.1)

Otras:

- Kschischang. *Factor graphs and the sum-product algorithm*; IEEE Transactions on information theory. 2001. (Descargar). (lectura partes del paper)

- Kelly. *A New Interpretation of Information Rate.*; Bell System Technical Journal. 1956. (Descargar). (lectura paper)