

Bayesiaans leren

Les 1: Bayesiaanse statistiek

Joris Bierkens

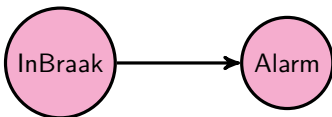
Vakantiecursus

augustus 2019

Next Section

1 Bayesiaanse statistiek

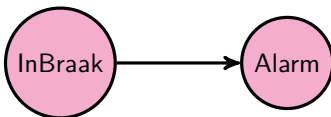
2 Neurale netwerken



Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

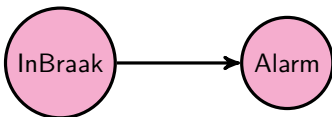


Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

$$P(B | A)$$

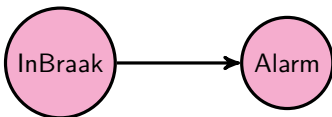


Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

$$P(B | A) = \frac{P(B \text{ and } A)}{P(A)}$$

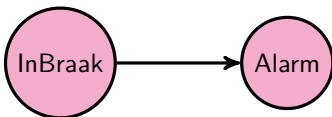


Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

$$P(B | A) = \frac{P(B \text{ and } A)}{P(A)} = \frac{P(A | B)P(B)}{P(A)}$$

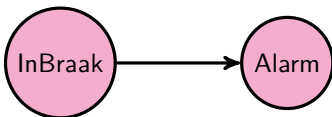


Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

$$P(B | A) = \frac{P(B \text{ and } A)}{P(A)} = \frac{P(A | B)P(B)}{P(A)} = \frac{P(A | B)P(B)}{P(A \text{ and } B) + P(A \text{ and not } B)}$$

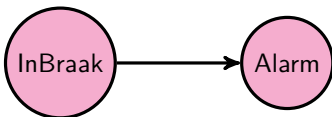


Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

$$\begin{aligned} P(B | A) &= \frac{P(B \text{ and } A)}{P(A)} = \frac{P(A | B)P(B)}{P(A)} = \frac{P(A | B)P(B)}{P(A \text{ and } B) + P(A \text{ and not } B)} \\ &= \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | \text{not } B)P(\text{not } B)} \end{aligned}$$

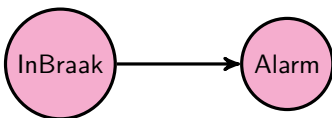


Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

$$\begin{aligned} P(B | A) &= \frac{P(B \text{ and } A)}{P(A)} = \frac{P(A | B)P(B)}{P(A)} = \frac{P(A | B)P(B)}{P(A \text{ and } B) + P(A \text{ and not } B)} \\ &= \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | \text{not } B)P(\text{not } B)} = \frac{0.99 \times 10^{-4}}{0.99 \times 10^{-4} + 0.001 \times (1 - 10^{-4})} \approx 0.090 \end{aligned}$$



Wordt er ingebroken?

- Als er een inBraak is (B), gaat het Alarm af (A) met kans 0.99.
- Een vals Alarm treedt op met kans 0.001.

In mijn wijk is de kans op een inbraak ongeveer 1×10^{-4} per nacht per woning.

$$\begin{aligned} P(B | A) &= \frac{P(B \text{ and } A)}{P(A)} = \frac{P(A | B)P(B)}{P(A)} = \frac{P(A | B)P(B)}{P(A \text{ and } B) + P(A \text{ and not } B)} \\ &= \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | \text{not } B)P(\text{not } B)} = \frac{0.99 \times 10^{-4}}{0.99 \times 10^{-4} + 0.001 \times (1 - 10^{-4})} \approx 0.090 \end{aligned}$$

Wet van Bayes

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)}$$



Thomas Bayes, c.
1701 – 1761

Bayesiaanse statistiek

Combineer

- “a priori” informatie (kans op inbraak)
- observaties (gaat het alarm af?)

om “a posteriori” informatie te verkrijgen (kans op inbraak, gegeven de observaties).

Bayesiaanse statistiek

Combineer

- “a priori” informatie (kans op inbraak)
- observaties (gaat het alarm af?)

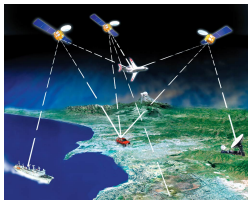
om “a posteriori” informatie te verkrijgen (kans op inbraak, gegeven de observaties).

Bayesiaanse statistiek

- **Prior** kansverdeling π_0 over de parameter $\theta \in \mathcal{S}$.
- Kansverdeling (**model**) $p(y | \theta)$ voor observaties y conditioneel op de parameter θ .
- Geeft: **Posterior** kansverdeling

$$\pi(\theta) = \frac{p(y | \theta)\pi_0(\theta)}{\sum_{\theta' \in \mathcal{S}} p(y | \theta')\pi_0(\theta')}, \quad \theta \in \mathcal{S}.$$

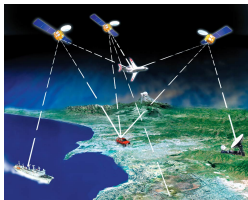
Bayesiaanse statistiek met continue parameter



Voorbeeld: GPS signaal

- observaties: signaal ontvangen door GPS ontvanger
- parameter: voertuigpositie in \mathbb{R}^d

Bayesiaanse statistiek met continue parameter



Voorbeeld: GPS signaal

- observaties: signaal ontvangen door GPS ontvanger
- parameter: voertuigpositie in \mathbb{R}^d

Bayesiaanse statistiek in de Euclidische ruimte

- **Prior dichtheidsfunctie** $\pi_0(\theta)$, $\theta \in \mathcal{S} \subset \mathbb{R}^d$.
- Voorwaardelijke **dichtheidsfunctie/kansfunctie** $p(y | \theta)$, $\theta \in \mathcal{S}$, $y \in \mathcal{O}$.
- Geeft **posterior dichtheidsfunctie**

$$\pi(\theta) = \frac{p(y | \theta)\pi_0(\theta)}{\int_{\mathcal{S}} p(y | \theta')\pi_0(\theta') d\theta'} \propto p(y | \theta)\pi_0(\theta), \quad \theta \in \mathcal{S}.$$

Onafhankelijke waarnemingen

Stel we hebben n observaties $\mathbf{y} = (y_1, \dots, y_n)$ volgens dichtheidsfunctie $f(y_i | \theta)$. Dan is het kansmodel

$$p(\mathbf{y} | \theta) = \prod_{i=1}^n f(y_i | \theta),$$

Onafhankelijke waarnemingen

Stel we hebben n observaties $\mathbf{y} = (y_1, \dots, y_n)$ volgens dichtheidsfunctie $f(y_i | \theta)$. Dan is het kansmodel

$$p(\mathbf{y} | \theta) = \prod_{i=1}^n f(y_i | \theta),$$

en dus wordt de posterior dichtheidsfunctie

$$\pi(\theta) = \frac{\prod_{i=1}^n f(y_i | \theta) \pi_0(\theta)}{\int_{\mathcal{S}} \prod_{i=1}^n f(y_i | \theta') \pi_0(\theta') d\theta'} \propto p(\mathbf{y} | \theta) \pi_0(\theta), \quad \theta \in \mathbb{R}^d.$$

Onafhankelijke waarnemingen

Stel we hebben n observaties $\mathbf{y} = (y_1, \dots, y_n)$ volgens dichtheidsfunctie $f(y_i | \theta)$. Dan is het kansmodel

$$p(\mathbf{y} | \theta) = \prod_{i=1}^n f(y_i | \theta),$$

en dus wordt de posterior dichtheidsfunctie

$$\pi(\theta) = \frac{\prod_{i=1}^n f(y_i | \theta) \pi_0(\theta)}{\int_{\mathcal{S}} \prod_{i=1}^n f(y_i | \theta') \pi_0(\theta') d\theta'} \propto p(\mathbf{y} | \theta) \pi_0(\theta), \quad \theta \in \mathbb{R}^d.$$

Voorbeeld

- $y_i | \mu \sim \mathcal{N}(\mu, \sigma^2)$
- $\pi_0: \mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$

Onafhankelijke waarnemingen

Stel we hebben n observaties $\mathbf{y} = (y_1, \dots, y_n)$ volgens dichtheidsfunctie $f(y_i | \theta)$. Dan is het kansmodel

$$p(\mathbf{y} | \theta) = \prod_{i=1}^n f(y_i | \theta),$$

en dus wordt de posterior dichtheidsfunctie

$$\pi(\theta) = \frac{\prod_{i=1}^n f(y_i | \theta) \pi_0(\theta)}{\int_{\mathcal{S}} \prod_{i=1}^n f(y_i | \theta') \pi_0(\theta') d\theta'} \propto p(\mathbf{y} | \theta) \pi_0(\theta), \quad \theta \in \mathbb{R}^d.$$

Voorbeeld

- $y_i | \mu \sim \mathcal{N}(\mu, \sigma^2)$
- $\pi_0: \mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$

Opgave

- $y_i | p \sim \text{Ber}(p)$: $P(y_i = 1) = p$, en $P(y_i = 0) = 1 - p$.
- $\pi_0(p) \propto p^{a-1}(1-p)^{b-1}$, $a, b > 0$.

Voorbeeld: overstromingskans

(Thomas, 1948)

- Een 'kistdam' die een bouwplaats beschermt is ontworpen om stromingen van 5.3×10^4 liter per seconde (l/s) te weerstaan.

Voorbeeld: overstromingskans

(Thomas, 1948)

- Een 'kistdam' die een bouwplaats beschermt is ontworpen om stromingen van 5.3×10^4 liter per seconde (l/s) te weerstaan.
- Over de afgelopen $n = 25$ jaar, zijn overstromingsniveau's gehaald tussen 1.8×10^4 l/s en 1.3×10^5 l/s. Het kritieke niveau van 5.3×10^4 l/s is $y = 5$ keer overschreden in deze periode.



Voorbeeld: overstromingskans

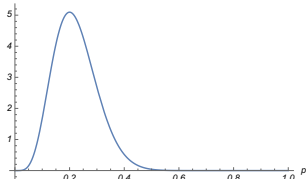
(Thomas, 1948)

- Een 'kistdam' die een bouwplaats beschermt is ontworpen om stromingen van 5.3×10^4 liter per seconde (l/s) te weerstaan.
- Over de afgelopen $n = 25$ jaar, zijn overstromingsniveau's gehaald tussen 1.8×10^4 l/s en 1.3×10^5 l/s. Het kritieke niveau van 5.3×10^4 l/s is $y = 5$ keer overschreden in deze periode.
- Met een uniforme prior op p vinden we de jaarlijkse kans op een overstroming sterker dan het kritieke niveau als

$$\pi(p | y) \propto p^y (1 - p)^{n-y} = p^5 (1 - p)^{20}.$$



Posterior density

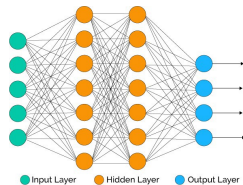


Next Section

1 Bayesiaanse statistiek

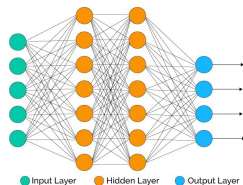
2 Neurale netwerken

Neuraal netwerk



- Geparametriseerde verzameling functies $(f(\cdot; \theta))_{\theta \in \mathcal{S}}$.

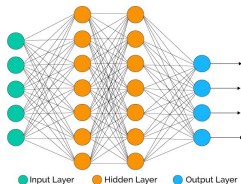
Neuraal netwerk



- Geparametriseerde verzameling functies $(f(\cdot; \theta))_{\theta \in \mathcal{S}}$.

$$f(\cdot; \theta) = f_{\theta^{(L)}}^{(L)} \circ f_{\theta^{(L-1)}}^{(L-1)} \circ \dots \circ f_{\theta^{(2)}}^{(2)} \circ f_{\theta^{(1)}}^{(1)},$$

Neuraal netwerk

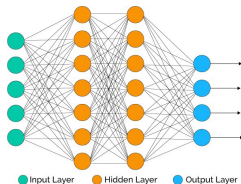


- Geparametriseerde verzameling functies $(f(\cdot; \theta))_{\theta \in \mathcal{S}}$.

$$f(\cdot; \theta) = f_{\theta^{(L)}}^{(L)} \circ f_{\theta^{(L-1)}}^{(L-1)} \circ \dots \circ f_{\theta^{(2)}}^{(2)} \circ f_{\theta^{(1)}}^{(1)}, \quad \text{waarbij}$$

$$f_{\theta^{(l)}}^{(l)}(\xi) = g^{(l)} \left(\underbrace{\mathbf{W}^{(l)\top}}_{\text{matrix}} \xi + \underbrace{\mathbf{b}^{(l)}}_{\text{vector}} \right), \quad \theta^{(l)} = (\underbrace{\mathbf{W}^{(l)}}_{\text{matrix}}, \underbrace{\mathbf{b}^{(l)}}_{\text{vector}}).$$

Neuraal netwerk



- Geparametriseerde verzameling functies $(f(\cdot; \theta))_{\theta \in \mathcal{S}}$.

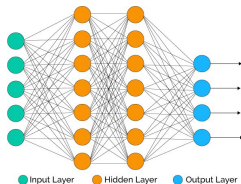
$$f(\cdot; \theta) = f_{\theta^{(L)}}^{(L)} \circ f_{\theta^{(L-1)}}^{(L-1)} \circ \dots \circ f_{\theta^{(2)}}^{(2)} \circ f_{\theta^{(1)}}^{(1)}, \quad \text{waarbij}$$

$$f_{\theta^{(l)}}^{(l)}(\xi) = g^{(l)} \left(\mathbf{W}^{(l)\top} \xi + \mathbf{b}^{(l)} \right), \quad \theta^{(l)} = \underbrace{(\mathbf{W}^{(l)})}_{\text{matrix}}, \underbrace{(\mathbf{b}^{(l)})}_{\text{vector}}.$$

- $g^{(l)} : \mathbb{R} \rightarrow \mathbb{R}$, $l = 1, \dots, L - 1$ zijn niet-linear en worden componentsgewijs toegepast.

sigmoid: $g^{(l)}(z) = (1 + e^{-z})^{-1}$,

Neuraal netwerk



- Geparametriseerde verzameling functies $(f(\cdot; \theta))_{\theta \in \mathcal{S}}$.

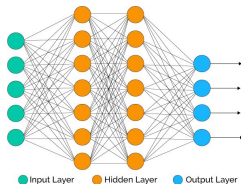
$$f(\cdot; \theta) = f_{\theta^{(L)}}^{(L)} \circ f_{\theta^{(L-1)}}^{(L-1)} \circ \dots \circ f_{\theta^{(2)}}^{(2)} \circ f_{\theta^{(1)}}^{(1)}, \quad \text{waarbij}$$

$$f_{\theta^{(l)}}^{(l)}(\xi) = g^{(l)} \left(\underbrace{\mathbf{W}^{(l)\top}}_{\text{matrix}} \xi + \underbrace{\mathbf{b}^{(l)}}_{\text{vector}} \right), \quad \theta^{(l)} = (\underbrace{\mathbf{W}^{(l)}}_{\text{matrix}}, \underbrace{\mathbf{b}^{(l)}}_{\text{vector}}).$$

- $g^{(l)} : \mathbb{R} \rightarrow \mathbb{R}$, $l = 1, \dots, L - 1$ zijn niet-linear en worden componentsgewijs toegepast.

sigmoid: $g^{(l)}(z) = (1 + e^{-z})^{-1}$, **ReLU:** $g^{(l)}(z) = \max(0, z)$.

Neuraal netwerk



- Geparametriseerde verzameling functies $(f(\cdot; \theta))_{\theta \in \mathcal{S}}$.

$$f(\cdot; \theta) = f_{\theta^{(L)}}^{(L)} \circ f_{\theta^{(L-1)}}^{(L-1)} \circ \dots \circ f_{\theta^{(2)}}^{(2)} \circ f_{\theta^{(1)}}^{(1)}, \quad \text{waarbij}$$

$$f_{\theta^{(l)}}^{(l)}(\xi) = g^{(l)} \left(\underbrace{\mathbf{W}^{(l)\top}}_{\text{matrix}} \xi + \underbrace{\mathbf{b}^{(l)}}_{\text{vector}} \right), \quad \theta^{(l)} = (\underbrace{\mathbf{W}^{(l)}}_{\text{matrix}}, \underbrace{\mathbf{b}^{(l)}}_{\text{vector}}).$$

- $g^{(l)} : \mathbb{R} \rightarrow \mathbb{R}$, $l = 1, \dots, L-1$ zijn niet-linear en worden componentsgewijs toegepast.

sigmoid: $g^{(l)}(z) = (1 + e^{-z})^{-1}$, **ReLU:** $g^{(l)}(z) = \max(0, z)$.

- Laatste laag bijvoorbeeld

linear : $g^{(L)}(\mathbf{z}) = \mathbf{z}$, **softmax:** $g^{(L)}(\mathbf{z})_i = \exp(z_i) / (\sum_j \exp(z_j))$.

Neuraal netwerk voor classificatie

- Bij classificatie zoeken we een kansverdeling $p(y | \mathbf{x})$ waarbij $y \in \{1, \dots, K\}$.

Neuraal netwerk voor classificatie

- Bij classificatie zoeken we een kansverdeling $p(y | \mathbf{x})$ waarbij $y \in \{1, \dots, K\}$.
- Benaderende kansverdeling uit verzameling $(f(\cdot; \boldsymbol{\theta}))_{\boldsymbol{\theta} \in \mathcal{S}}$.

Neuraal netwerk voor classificatie

- Bij classificatie zoeken we een kansverdeling $p(y | \mathbf{x})$ waarbij $y \in \{1, \dots, K\}$.
- Benaderende kansverdeling uit verzameling $(f(\cdot; \boldsymbol{\theta}))_{\boldsymbol{\theta} \in \mathcal{S}}$.
- Softmax-functie in laatste laag geeft kansvector als output, en we kiezen

$$p_{\text{model}}(y | \mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}; \boldsymbol{\theta})(y).$$

Neuraal netwerk voor classificatie

- Bij classificatie zoeken we een kansverdeling $p(y | \mathbf{x})$ waarbij $y \in \{1, \dots, K\}$.
- Benaderende kansverdeling uit verzameling $(f(\cdot; \boldsymbol{\theta}))_{\boldsymbol{\theta} \in \mathcal{S}}$.
- Softmax-functie in laatste laag geeft kansvector als output, en we kiezen

$$p_{\text{model}}(y | \mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}; \boldsymbol{\theta})(y).$$

- Optimaliteitscriterium

$$J(\boldsymbol{\theta}) = - \sum_{i=1}^n \log p_{\text{model}}(y_i | \mathbf{x}_i, \boldsymbol{\theta}).$$

Neuraal netwerk voor classificatie

- Bij classificatie zoeken we een kansverdeling $p(y | \mathbf{x})$ waarbij $y \in \{1, \dots, K\}$.
- Benaderende kansverdeling uit verzameling $(f(\cdot; \boldsymbol{\theta}))_{\boldsymbol{\theta} \in \mathcal{S}}$.
- Softmax-functie in laatste laag geeft kansvector als output, en we kiezen

$$p_{\text{model}}(y | \mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}; \boldsymbol{\theta})(y).$$

- Optimaliteitscriterium

$$J(\boldsymbol{\theta}) = - \sum_{i=1}^n \log p_{\text{model}}(y_i | \mathbf{x}_i, \boldsymbol{\theta}).$$

- Probleem: **overfitting**

Neuraal netwerk voor classificatie

- Bij classificatie zoeken we een kansverdeling $p(y | \mathbf{x})$ waarbij $y \in \{1, \dots, K\}$.
- Benaderende kansverdeling uit verzameling $(f(\cdot; \boldsymbol{\theta}))_{\boldsymbol{\theta} \in \mathcal{S}}$.
- Softmax-functie in laatste laag geeft kansvector als output, en we kiezen

$$p_{\text{model}}(y | \mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}; \boldsymbol{\theta})(y).$$

- Optimaliteitscriterium

$$J(\boldsymbol{\theta}) = - \sum_{i=1}^n \log p_{\text{model}}(y_i | \mathbf{x}_i, \boldsymbol{\theta}).$$

- Probleem: **overfitting**
- Aanpak: **regularisatie** door nieuw optimaliteitscriterium

$$J(\boldsymbol{\theta}) = - \sum_{i=1}^n \log p_{\text{model}}(y_i | \mathbf{x}_i, \boldsymbol{\theta}) + \text{pen}(\boldsymbol{\theta}).$$

Bayesiaanse interpretatie neuraal netwerk

- Optimaliteitscriterium

$$J(\boldsymbol{\theta}) = - \sum_{i=1}^n \log p_{\text{model}}(y_i | \mathbf{x}_i, \boldsymbol{\theta}) + \text{pen}(\boldsymbol{\theta}).$$

- Gegeven een minimaliserende $\boldsymbol{\theta}_0$ van $J(\boldsymbol{\theta})$, kunnen we een nieuwe voorspelling doen d.m.v. $p_{\text{model}}(y_{n+1} = i | \mathbf{x}_{n+1}, \boldsymbol{\theta}_0)$.
- **Bayesiaanse interpretatie:** We herkennen

$$J(\boldsymbol{\theta}) = - \log \pi(\boldsymbol{\theta} | \mathbf{X}, \mathbf{y}) + \text{const},$$

waarbij

$$\pi(\boldsymbol{\theta} | \mathbf{X}, \mathbf{y}) \propto p_{\text{model}}(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta}) \pi_0(\boldsymbol{\theta}) \quad \text{en} \quad \pi_0(\boldsymbol{\theta}) = \exp(-\text{pen}(\boldsymbol{\theta})).$$

- **Nuttig?** Een **posterior kansverdeling** is een rijker object dan enkel een optimaliteitscriterium. We kunnen nu een Bayesiaanse voorspelling doen:

$$P(y_{n+1} = i | \mathbf{x}_{n+1}) = \int_{\mathcal{S}} \pi(\boldsymbol{\theta}) p_{\text{model}}(y_{n+1} = i | \mathbf{x}_{n+1}, \boldsymbol{\theta}) d\boldsymbol{\theta}$$

- Quantificeert de **onzekerheid** in $\boldsymbol{\theta}$.
- Helaas is er geen eenvoudige vorm voor de posterior (zoals er wel was bij eerdere voorbeelden).

Voorbeeld: binaire classificatie (1)

# drankjes (x)	4	3	1	5	1	2	3	3	2	1
hoofdpijn (y)	1	0	0	1	0	1	0	1	1	0

Voorbeeld: binaire classificatie (1)

# drankjes (x)	4	3	1	5	1	2	3	3	2	1
hoofdpijn (y)	1	0	0	1	0	1	0	1	1	0

■ Model: $p(y = 1 | x, \mathbf{w}) = 1/(1 + \exp(-w_0 - w_1x))$

Voorbeeld: binaire classificatie (1)

# drankjes (x)	4	3	1	5	1	2	3	3	2	1
hoofdpijn (y)	1	0	0	1	0	1	0	1	1	0

- Model: $p(y = 1 | x, \mathbf{w}) = 1/(1 + \exp(-w_0 - w_1x))$
- Dit mini-neurale netwerk heet ook wel **logistische regressie**

Voorbeeld: binaire classificatie (1)

# drankjes (x)	4	3	1	5	1	2	3	3	2	1
hoofdpijn (y)	1	0	0	1	0	1	0	1	1	0

- Model: $p(y = 1 | x, \mathbf{w}) = 1/(1 + \exp(-w_0 - w_1x))$
- Dit mini-neurale netwerk heet ook wel **logistische regressie**
- Prior: uniforme verdeling over $\mathbf{w} \in \mathbb{R}^2$ (!)

Voorbeeld: binaire classificatie (1)

# drankjes (x)	4	3	1	5	1	2	3	3	2	1
hoofdpijn (y)	1	0	0	1	0	1	0	1	1	0

- Model: $p(y = 1 | x, \mathbf{w}) = 1/(1 + \exp(-w_0 - w_1x))$
- Dit mini-neurale netwerk heet ook wel **logistische regressie**
- Prior: uniforme verdeling over $\mathbf{w} \in \mathbb{R}^2$ (!)
- Posterior:

$$\pi(\mathbf{w} | \mathbf{x}, \mathbf{y}) \propto \prod_{i=1}^n p(y_i | x_i, \mathbf{w}).$$

Voorbeeld: binaire classificatie (1)

# drankjes (x)	4	3	1	5	1	2	3	3	2	1
hoofdpijn (y)	1	0	0	1	0	1	0	1	1	0

- Model: $p(y = 1 | x, \mathbf{w}) = 1/(1 + \exp(-w_0 - w_1x))$
- Dit mini-neurale netwerk heet ook wel **logistische regressie**
- Prior: uniforme verdeling over $\mathbf{w} \in \mathbb{R}^2$ (!)
- Posterior:

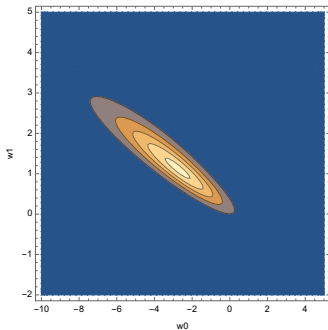
$$\pi(\mathbf{w} | \mathbf{x}, \mathbf{y}) \propto \prod_{i=1}^n p(y_i | x_i, \mathbf{w}).$$

Voorbeeld: binaire classificatie (1)

# drankjes (x)	4	3	1	5	1	2	3	3	2	1
hoofdpijn (y)	1	0	0	1	0	1	0	1	1	0

- Model: $p(y = 1 \mid x, \mathbf{w}) = 1/(1 + \exp(-w_0 - w_1x))$
- Dit mini-neurale netwerk heet ook wel **logistische regressie**
- Prior: uniforme verdeling over $\mathbf{w} \in \mathbb{R}^2$ (!)
- Posterior:

$$\pi(\mathbf{w} \mid \mathbf{x}, \mathbf{y}) \propto \prod_{i=1}^n p(y_i \mid x_i, \mathbf{w}).$$

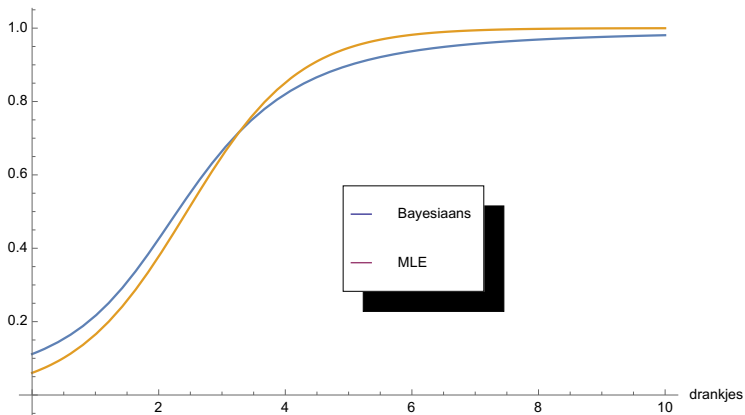


posterior verdeling

Voorbeeld: binaire classificatie (2)

Kans op hoofdpijn

voorspelde kans op hoofdpijn



Voorbeeld: binaire classificatie (3)

Variantie in hoofdpijn

voorspelde hoofdpijn variantie

