Bayesian decision making

Z. Straka, P. Švarný, J. Kostlivá

Today two examples:

1. Bayesian decision making basics
2. Prior probabilities in practice

Bayesian decision making basics

Bayesian decision making basics

What is correct?

A: $P(X = x_i) = \sum_j \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)}$

B: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$

C: $P(X = x_i) = \sum_i P(X = x_i, Y = y_j)$

D: $P(Y = y_i) = \sum_j P(X = x_i, Y = y_j)$

Bayesian decision making basics

What is correct?

A: $P(X = x_i) = \sum_j \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)}$

B: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$

C: $P(X = x_i) = \sum_i P(X = x_i, Y = y_j)$

D: $P(Y = y_i) = \sum_j P(X = x_i, Y = y_j)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$

What is correct?

A: $P(X = x_i | Y = y_j) = P(Y = y_j, X = x_i)P(X = x_i)$

B: $P(X = x_i, Y = y_j) = \frac{P(Y=y_j|X=x_i)}{P(X=x_i)}$

C: $P(X = x_i, Y = y_j) = P(Y = y_j|X = x_i)P(Y = y_i)$

D: $P(X = x_i, Y = y_j) = P(Y = y_j|X = x_i)P(X = x_i)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$

What is correct?

A: $P(X = x_i | Y = y_j) = P(Y = y_j, X = x_i) P(X = x_i)$

B: $P(X = x_i, Y = y_j) = \frac{P(Y = y_j | X = x_i)}{P(X = x_i)}$

C: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(Y = y_i)$

D: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$

What is correct?

A: $P(Y = y_j | X = x_i) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{P(X=x_i)}$

B: $P(Y = y_i, X = x_j) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{P(X=x_j)}$

C: $P(Y = y_i | X = x_j) = P(X = x_i | Y = y_j) P(Y = y_j)$

D: $P(Y = y_i | X = x_j) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{\sum_l P(X=x_l, Y=y_i)}$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$

What is correct?

A: $P(Y = y_j | X = x_i) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{P(X=x_i)}$

B: $P(Y = y_i, X = x_j) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{P(X=x_j)}$

C: $P(Y = y_i | X = x_j) = P(X = x_i | Y = y_j) P(Y = y_j)$

D: $P(Y = y_i | X = x_j) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{\sum_l P(X=x_l, Y=y_i)}$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$
- Bayes' theorem: $P(Y = y_j | X = x_i) = \frac{P(X=x_i|Y=y_j)P(Y=y_j)}{P(X=x_i)}$

What is correct?

A: $\delta^* = \arg\max_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

B: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x|s)$

C: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

D: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, x) P(x, s)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$
- Bayes' theorem: $P(Y = y_j | X = x_i) = \frac{P(X = x_i | Y = y_j) P(Y = y_j)}{P(X = x_i)}$

What is correct?

A: $\delta^* = \arg\max_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

B: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x|s)$

C: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

D: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, x) P(x, s)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$
- Bayes' theorem: $P(Y = y_j | X = x_i) = \frac{P(X=x_i|Y=y_j)P(Y=y_j)}{P(X=x_i)}$
- Bayes optimal strategy: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

What is correct?

A: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s|x)$

B: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s, x)$

C: $\delta^*(x) = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

D: $\delta^*(x) = \arg\min_s \sum_d l(s, d) P(s|x)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$
- Bayes' theorem: $P(Y = y_j | X = x_i) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{P(X=x_i)}$
- Bayes optimal strategy: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

What is correct?

A: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s|x)$

B: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s, x)$

C: $\delta^*(x) = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$

D: $\delta^*(x) = \arg\min_s \sum_d l(s, d) P(s|x)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$
- Bayes' theorem: $P(Y = y_j | X = x_i) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{P(X=x_i)}$
- Bayes optimal strategy: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$
- BOS solution: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s|x)$

Assume $l(s, d) = 1, \texttt{if } d \neq s, l(s, d) = 0 \texttt{ otherwise.}$ What is correct?

A: $\delta^*(x) = \arg\min_d P(d|x)$

B: $\delta^*(x) = \arg\max_d P(d|x)$

C: $\delta^*(x) = \arg\max_d P(d|x) P(x)$

D: $\delta^*(x) = \arg\max_d P(d|x) P(s)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$
- Bayes' theorem: $P(Y = y_j | X = x_i) = \frac{P(X=x_i|Y=y_j)P(Y=y_j)}{P(X=x_i)}$
- Bayes optimal strategy: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$
- BOS solution: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s|x)$

Assume $l(s, d) = 1, \texttt{if } d \neq s, l(s, d) = 0 \texttt{ otherwise}.$ What is correct?

A: $\delta^*(x) = \arg\min_d P(d|x)$

B: $\delta^*(x) = \arg\max_d P(d|x)$

C: $\delta^*(x) = \arg\max_d P(d|x) P(x)$

D: $\delta^*(x) = \arg\max_d P(d|x) P(s)$

Bayesian decision making basics

- Sum rule of probability: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- Product rule of probability: $P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$
- Bayes' theorem: $P(Y = y_j | X = x_i) = \frac{P(X=x_i | Y=y_j) P(Y=y_j)}{P(X=x_i)}$
- Bayes optimal strategy: $\delta^* = \arg\min_\delta \sum_x \sum_s l(s, \delta(x)) P(x, s)$
- BOS solution: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s|x)$
- $L_{0,1}$ classification: $\delta^*(x) = \arg\max_d P(d|x)$

Prior probabilities in practice

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female.

A: Male

B: Female

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female.

A: Male

B: Female

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

$\frac{x}{\text{cm}}$	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female.

A: Male

B: Female (if we assume that there are same the number of men and women.)

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

$\frac{x}{\text{cm}}$	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female. **Female**
Task 2: What would be the minimum ratio of men to change the previous answer into the opposite one.

Right step?

A: $P(X = male, Y = L) = P(X = female, Y = L)$

B: $P(X = male|Y = L) = P(X = female|Y = L)$

C: $P(X = male|Y > L) = P(X = female|Y < L)$

D: $P(X = male|Y > L) > P(X = female|Y < L)$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female. **Female**
Task 2: What would be the minimum ratio of men to change the previous answer into the opposite one.

Right step?

A: $P(X = male, Y = L) = P(X = female, Y = L)$

B: $P(X = male|Y = L) = P(X = female|Y = L)$

C: $P(X = male|Y > L) = P(X = female|Y < L)$

D: $P(X = male|Y > L) > P(X = female|Y < L)$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x \mid \text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x \mid \text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female. **Female**
Task 2: What would be the minimum ratio of men to change the previous answer into the opposite one.

Right step?

A: $P(X = male, Y = L) = P(X = female, Y = L)$

B: $P(X = male | Y = L) = P(X = female | Y = L)$

C: $P(X = male | Y > L) = P(X = female | Y < L)$

D: $P(X = male | Y > L) > P(X = female | Y < L)$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

$\frac{x}{\text{cm}}$	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female. **Female**
Task 2: What would be the minimum ratio of men to change the previous answer into the opposite one. $P(X = male|Y = L) = P(X = female|Y = L)$

From the equation get value of?

A: $P(X = male)$

B: $P(X = male|Y = L)$

C: $P(X = female|Y < L)$

D: $P(X = male|Y > L)$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female. **Female**
Task 2: What would be the minimum ratio of men to change the previous answer into the opposite one. $P(X = male|Y = L) = P(X = female|Y = L)$

From the equation get value of?

A: $P(X = male)$

B: $P(X = male|Y = L)$

C: $P(X = female|Y < L)$

D: $P(X = male|Y > L)$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female. **Female**
Task 2: What would be the minimum ratio of men to change the previous answer into the opposite one. $P(X = male|Y = L) = P(X = female|Y = L)$

From the equation get value of? $P(X = male)$
Calculate $P(X = male)$:

A: $\frac{5}{11}$

B: $\frac{6}{11}$

C: $\frac{6}{10}$

D: $\frac{7}{12}$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\vert male)$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\vert female)$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 1: Estimate whether a 168cm tall (i.e. L) person is male or female. **Female**
Task 2: What would be the minimum ratio of men to change the previous answer into the opposite one. $P(X = male|Y = L) = P(X = female|Y = L)$

From the equation get value of? $P(X = male)$
Calculate $P(X = male)$:

B: $\frac{6}{11}$

$P(X = male|Y = L) = P(X = female|Y = L)$
$\frac{P(L|male)\cdot P(male)}{P(L)} = \frac{P(L|female)\cdot P(female)}{P(L)}, P(female) = 1 - P(male)$
$P(L|male) \cdot P(male) = P(L|female) \cdot (1 - P(male))$
$0.25 \cdot P(male) = 0.3 - 0.3 \cdot P(male) \Rightarrow P(male) = \frac{6}{11}$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

$\frac{x}{\text{cm}}$	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x \mid \text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x \mid \text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 3: Assuming there are 70% men and 30% women, consider the loss function l (s - state, d - decision): $l(s = \text{female}, d = \text{male}) = 2, l(s = \text{male}, d = \text{female}) = 1,$
$l(s = \text{male}, d = \text{male}) = l(s = \text{female}, d = \text{female}) = 0.$
How do you classify a person under consideration of L?
How?

A: $\delta^*(X = L) = \mathrm{argmin}_s \sum_s l(s, d) \cdot P(s|X = L)$

B: $\delta^*(X = L) = \mathrm{argmin}_d l(s, d) \cdot P(s|X = L)$

C: $\delta^*(X = L) = \mathrm{argmin}_s \sum_d l(s, d) \cdot P(s|X = L)$

D: $\delta^*(X = L) = \mathrm{argmin}_d \sum_s l(s, d) \cdot P(s|X = L)$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 3: Assuming there are 70% men and 30% women, consider the loss function l (s = state, d = decision): $l(s = \text{female}, d = \text{male}) = 2, l(s = \text{male}, d = \text{female}) = 1,$
$l(s = \text{male}, d = \text{male}) = l(s = \text{female}, d = \text{female}) = 0.$
How do you classify a person under consideration of L?
How?

D: $\delta^*(X = L) = \text{argmin}_d \sum_s l(s, d) \cdot P(s|X = L)$

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 3: Assuming there are 70% men and 30% women, consider the loss function l (s = state, d = decision): $l(s = \text{female}, d = \text{male}) = 2, l(s = \text{male}, d = \text{female}) = 1$,
$l(s = \text{male}, d = \text{male}) = l(s = \text{female}, d = \text{female}) = 0$.
How do you classify a person under consideration of L?
How? $\delta^*(X = L) = \text{argmin}_d \sum_s l(s, d) \cdot P(s|X = L)$
Result?

A: female

B: male

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

x cm	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 3: Assuming there are 70% men and 30% women, consider the loss function l (s = state, d = decision): $l(s = \text{female}, d = \text{male}) = 2, l(s = \text{male}, d = \text{female}) = 1$,
$l(s = \text{male}, d = \text{male}) = l(s = \text{female}, d = \text{female}) = 0$.
How do you classify a person under consideration of L?
How? $\delta^*(X = L) = \text{argmin}_d \sum_s l(s, d) \cdot P(s|X = L)$
Result?

A: female

B: male

Prior probabilities in practice

The probability distribution of height of men and women is known (see table).

$\frac{x}{\text{cm}}$	XS (0–100)	S (100–125)	M (125–150)	L (150–175)	XL (175–200)	XXL (200-∞)	$\sum$
$P(x\|\text{male})$	0.05	0.15	0.2	0.25	0.3	0.05	**1**
$P(x\|\text{female})$	0.05	0.1	0.3	0.3	0.25	0.0	**1**

Task 3: Assuming there are 70% men and 30% women, consider the loss function l (s = state, d = decision): $l(s = \text{female}, d = \text{male}) = 2, l(s = \text{male}, d = \text{female}) = 1,$
$l(s = \text{male}, d = \text{male}) = l(s = \text{female}, d = \text{female}) = 0.$

How do you classify a person under consideration of L?

$P(male|L) = \frac{P(L|male) \cdot P(male)}{P(L)} = \frac{P(L|male) \cdot P(male)}{P(L|male) \cdot P(male) + P(L|female) \cdot P(female)} = \frac{0.25 \cdot 0.7}{0.25 \cdot 0.7 + 0.3 \cdot 0.3} = 0.66$

$P(female|L) = 1 - 0.66 = 0.34$

$\delta^*(X) = \operatorname{argmin}_d \left(l(female, d) \cdot P(female|L) + l(male, d) \cdot P(male|L) \right)$

$$\delta^*(X) = \operatorname{argmin}_d \left\{ \begin{array}{c} d = female : 0 \cdot 0.34 + 1 \cdot 0.66 = 0.66 \\ d = male : 2 \cdot 0.34 + 0 \cdot 0.66 = 0.68 \end{array} \right\} \Rightarrow d = female$$