Powtórzenie materiału

Probabilistyczne Uczenie Maszynowe

Tomasz Kajdanowicz, Piotr Bielak, Maciej Falkiewicz, Kacper Kania, Piotr Zieliński

Procedura Egzaminów

Termin:

- "0" - 9.06.2020 godz. 9:00 - 10:30
- "1" - 18.06.2020 godz. 11:15 - 13:00
- "2" - 25. 06. 2020 godz. 11:15 - 13:00

Sposób przeprowadzenia:

1. Na maile studenckie zostaną wysłane pliki PDF z pytaniami.
2. Egzamin będzie się składał zarówno z pytań zamkniętych jak i otwartych (proporcja 50/50).
3. W zależności od dostępnych narzędzi, należy rozwiązać test albo nanosząc odpowiedzi bezpośrednio na test (tablet graficzny / wydruk arkusza), albo zapisując **tylko** odpowiedzi na osobnej kartce, podpisanej imieniem, nazwiskiem oraz numerem indeksu.
4. Na koniec należy zrobić skan / zdjęcie (np. Camscanner) odpowiedzi i wysłać na adres tomasz.kajdanowicz@pwr.edu.pl z tytułem **[PUMA][Egzamin] Termin 0 - <nazwisko>**
5. Proszę przećwiczyć procedurę skanowania wcześniej!
6. Na czas egzaminu łączymy się na kanale głosowym Discord.

Q1

(a) [3 points] For data D and hypothesis H, say whether or not the following equations must always be true.

$\sum_h P(H = h|D = d) = 1$ ... is this always true?

$\sum_h P(D = d|H = h) = 1$ ... is this always true?

$\sum_h P(D = d|H = h)P(H = h) = 1$ ... is this always true?

(b) [2 points] For the following equations, describe the relationship between them. Write one of four answers:

(1) "$=$" (2) "$\leq$" (3) "$\geq$" (4) "(depends)"

Choose the most specific relation that always holds; "(depends)" is the least specific. Assume all probabilities are non-zero.

$P(H = h\|D = d)$	$P(H = h)$
$P(H = h\|D = d)$	$P(D = d\|H = h)P(H = h)$

Q1

(a) [3 points] For data D and hypothesis H, say whether or not the following equations must always be true.

$\sum_h P(H=h|D=d) = 1$ ... is this always true? **YES**

$\sum_h P(D=d|H=h) = 1$ ... is this always true? **NO**

$\sum_h P(D=d|H=h)P(H=h) = 1$ ... is this always true? **NO**

(b) [2 points] For the following equations, describe the relationship between them. Write one of four answers:

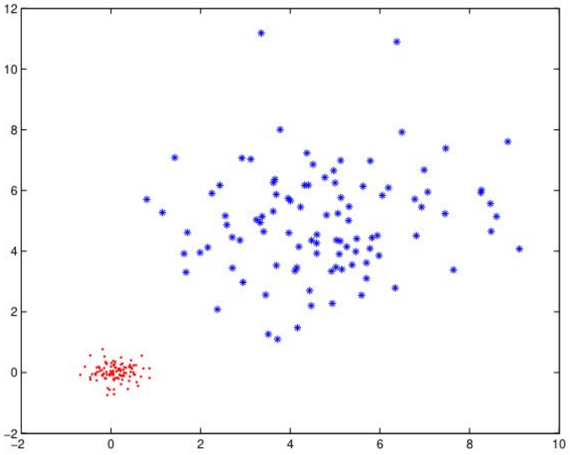
(1) "$=$" (2) "$\leq$" (3) "$\geq$" (4) "(depends)"

Choose the most specific relation that always holds; "(depends)" is the least specific. Assume all probabilities are non-zero.

$P(H=h\|D=d)$	**DEPENDS**	$P(H=h)$
$P(H=h\|D=d)$	**≥**	$P(D=d\|H=h)P(H=h)$

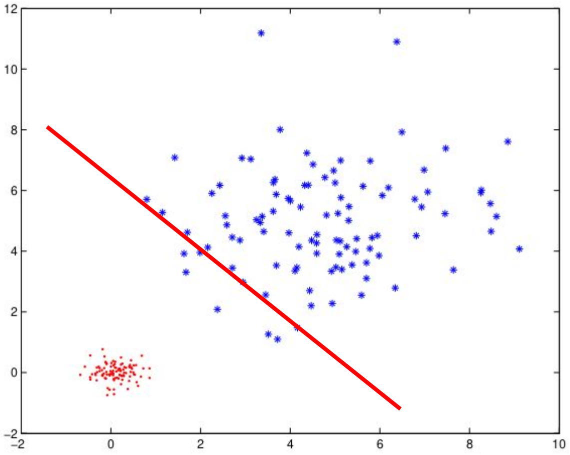
Q1

(c) [2 points] Suppose you are training Gaussian Naive Bayes (GNB) on the training set shown below. The dataset satisfies Gaussian Naive Bayes assumptions. Assume that the variance is independent of instances but dependent on classes, i.e. $\sigma_{ik} = \sigma_k$ where i indexes instances $X^{(i)}$ and $k \in 1, 2$ indexes classes. Draw the decision boundaries when you train GNB using the **same** variance for both classes, $\sigma_1 = \sigma_2$



Q1

(c) [2 points] Suppose you are training Gaussian Naive Bayes (GNB) on the training set shown below. The dataset satisfies Gaussian Naive Bayes assumptions. Assume that the variance is independent of instances but dependent on classes, i.e. $\sigma_{ik} = \sigma_k$ where i indexes instances $X^{(i)}$ and $k \in 1, 2$ indexes classes. Draw the decision boundaries when you train GNB using the **same** variance for both classes, $\sigma_1 = \sigma_2$



Q2

For the following parts, choose true or false with an explanation in **one sentence**

i. [1 point] Gamma mixture model can capture overlapping clusters, like Gaussian mixture model.

ii. [1 point] As you increase K, you will **always** get better likelihood of the data.

Q2

For the following parts, choose true or false with an explanation in **one sentence**

i. [1 point] Gamma mixture model can capture overlapping clusters, like Gaussian mixture model.

True. E-step performs soft assignment.

ii. [1 point] As you increase K, you will **always** get better likelihood of the data.

False. Won't improve after K > N

Q3

Given model parameters `w` and data `D` apply the Bayes formula to obtain `p(w|D)`. Name each component and tell which can be (in most cases) computed easily or not.

Q3

Given model parameters `w` and data `D` apply the Bayes formula to obtain `p(w|D)`. Name each component and tell which can be (in most cases) computed easily or not.

Likelihood (easy)

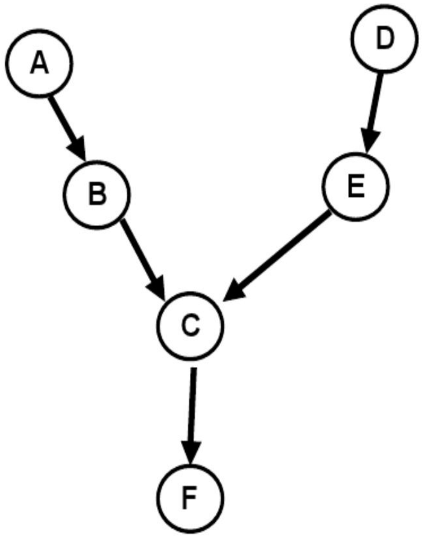
Prior (easy)

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)}$$

Normalizing factor / probability of data (hard)

Q4

Consider the following Bayesian network of 6 variables.



(a) [3 points] Set $X = \{B\}$ and $Y = \{E\}$. Specify two distinct (not-overlapping) sets Z such that: $X \perp\!\!\!\perp Y | Z$ (in other words, X is independent of Y given Z).

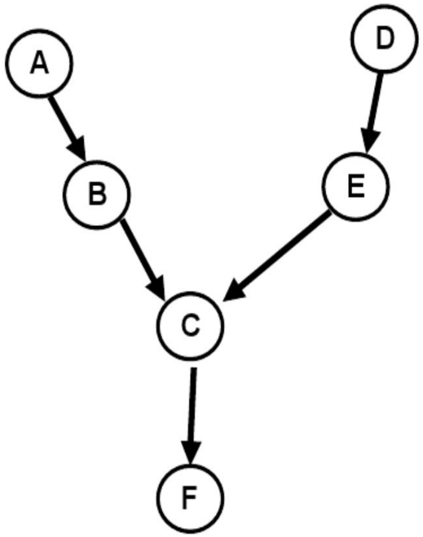
(b) [2 points] Can you find another distinct set for Z (i.e. a set that does not intersect with any of the sets listed in 1)?

(c) [2 points] How many distinct Z sets can you find if we replace B with A while Y stays the same (in other words, now $X = \{A\}$ and $Y = \{E\}$)? What are they?

(d) [2 points] If $W \perp\!\!\!\perp X | Z$ and $X \perp\!\!\!\perp Y | Z$ for some distinct variables W, X, Y, Z, can you say $W \perp\!\!\!\perp Y | Z$? If so, show why. If not, find a counterexample from the graph above.

Q4

Consider the following Bayesian network of 6 variables.



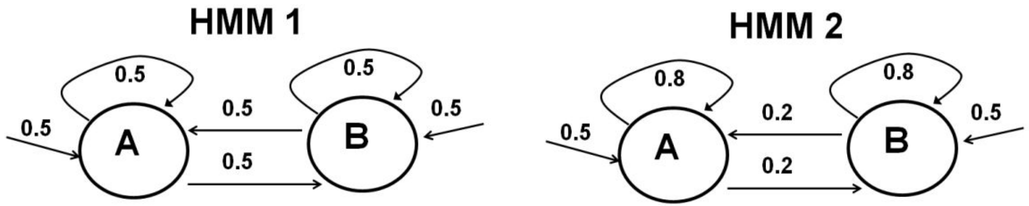
(a) [3 points] Set $X = \{B\}$ and $Y = \{E\}$. Specify two distinct (not-overlapping) sets Z such that: $X \perp\!\!\!\perp Y | Z$ (in other words, X is independent of Y given Z). **Z={A} and Z={D}**

(b) [2 points] Can you find another distinct set for Z (i.e. a set that does not intersect with any of the sets listed in 1)? **Z={} (empty set)**

(c) [2 points] How many distinct Z sets can you find if we replace B with A while Y stays the same (in other words, now $X = \{A\}$ and $Y = \{E\}$)? What are they? **Z = {}, Z = {B} and Z = {D}**

(d) [2 points] If $W \perp\!\!\!\perp X | Z$ and $X \perp\!\!\!\perp Y | Z$ for some distinct variables W, X, Y, Z, can you say $W \perp\!\!\!\perp Y | Z$? If so, show why. If not, find a counterexample from the graph above. **No. A⊥F |B and D⊥A|B but D and F are not independent given B.**

Q5



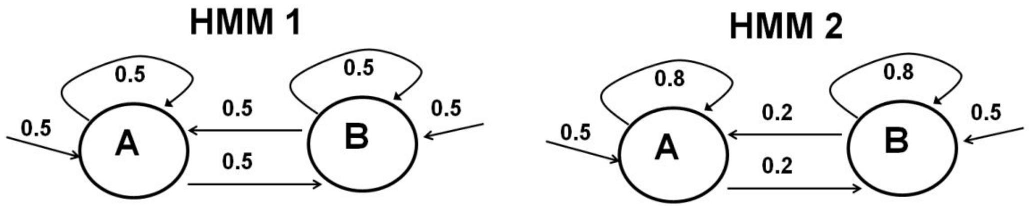
The figure above presents two HMMs. States are represented by circles and transitions by edges. In both, emissions are deterministic and listed inside the states.

Transition probabilities and starting probabilities are listed next to the relevant edges. For example, in HMM 1 we have a probability of 0.5 to start with the state that emits A and a probability of 0.5 to transition to the state that emits B if we are now in the state that emits A.

In the questions below, O_{100}=A means that the 100th symbol emitted by the HMM is A.

(a) [3 points] What is $P(O_{100} = A, O_{101} = A, O_{102} = A)$ for HMM1?

Q5



The figure above presents two HMMs. States are represented by circles and transitions by edges. In both, emissions are deterministic and listed inside the states.

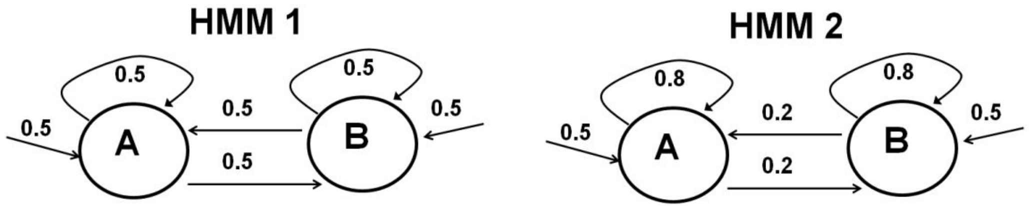
Transition probabilities and starting probabilities are listed next to the relevant edges. For example, in HMM 1 we have a probability of 0.5 to start with the state that emits A and a probability of 0.5 to transition to the state that emits B if we are now in the state that emits A.

In the questions below, O_{100}=A means that the 100th symbol emitted by the HMM is A.

(a) [3 points] What is $P(O_{100} = A, O_{101} = A, O_{102} = A)$ for HMM1?

The emission probabilities in the above equation are all 1. The transitions are all 0.5. So the only question is: What is $P(S_{100}$=A)? Since the model is fully symmetric, the answer to this is 0.5 and so the total equation evaluates to: 0.5^3

Q5



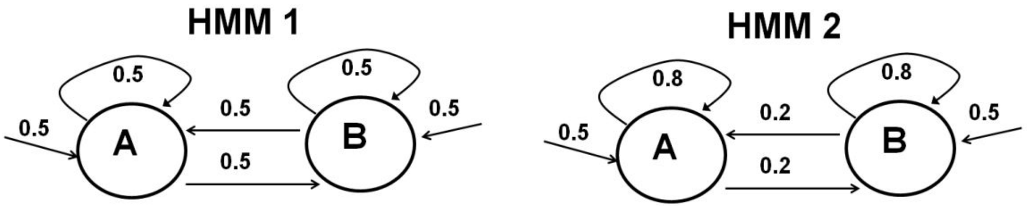
The figure above presents two HMMs. States are represented by circles and transitions by edges. In both, emissions are deterministic and listed inside the states.

Transition probabilities and starting probabilities are listed next to the relevant edges. For example, in HMM 1 we have a probability of 0.5 to start with the state that emits A and a probability of 0.5 to transition to the state that emits B if we are now in the state that emits A.

In the questions below, O_{100}=A means that the 100th symbol emitted by the HMM is A.

(b) [3 points] What is $P(O_{100} = A, O_{101} = A, O_{102} = A)$ for HMM2?

Q5



The figure above presents two HMMs. States are represented by circles and transitions by edges. In both, emissions are deterministic and listed inside the states.

Transition probabilities and starting probabilities are listed next to the relevant edges. For example, in HMM 1 we have a probability of 0.5 to start with the state that emits A and a probability of 0.5 to transition to the state that emits B if we are now in the state that emits A.

In the questions below, O_{100}=A means that the 100th symbol emitted by the HMM is A.

(b) [3 points] What is $P(O_{100} = A, O_{101} = A, O_{102} = A)$ for HMM2?

0.5 x 0.8^2

Q6

Załóżmy następujący rozkład łączny:

A	B	C	$P(A,B,C)$
a^0	b^0	c^0	0.2
a^0	b^0	c^1	0.2
a^0	b^1	c^0	0.1
a^0	b^1	c^1	0.1
a^1	b^0	c^0	0.1
a^1	b^0	c^1	0.1
a^1	b^1	c^0	0.1
a^1	b^1	c^1	0.1

(a) (3 punkty) Uzupełnij poniższą tabelę wykonując odpowiednie operacje.

A	B	C	$P(A,B,c^1)$
a^0	b^0	c^1	
a^0	b^1	c^1	
a^1	b^0	c^1	
a^1	b^1	c^1	

(b) (3 punkty) Uzupełnij poniższą tabelę wykonując odpowiednie operacje.

A	B	$P(A,B\|c^1)$
a^0	b^0	
a^0	b^1	
a^1	b^0	
a^1	b^1	

Q6

Załóżmy następujący rozkład łączny:

A	B	C	$P(A,B,C)$
a^0	b^0	c^0	0.2
a^0	b^0	c^1	0.2
a^0	b^1	c^0	0.1
a^0	b^1	c^1	0.1
a^1	b^0	c^0	0.1
a^1	b^0	c^1	0.1
a^1	b^1	c^0	0.1
a^1	b^1	c^1	0.1

(a) (3 punkty) Uzupełnij poniższą tabelę wykonując odpowiednie operacje.

A	B	C	$P(A,B,c^1)$
a^0	b^0	c^1	
a^0	b^1	c^1	
a^1	b^0	c^1	
a^1	b^1	c^1	

(b) (3 punkty) Uzupełnij poniższą tabelę wykonując odpowiednie operacje.

A	B	$P(A,B\|c^1)$
a^0	b^0	
a^0	b^1	
a^1	b^0	
a^1	b^1	

Q7

A. Rozwiń skrót ELBO ______________
B. Podaj wzór na ELBO ______________
C. Rozwiń skrót SVI _____________
D. Zadaniem SVI jest: ____________

Q8

W jakim celu używamy kryteriów AIC (Akaike information criterion) i BIC (Bayesian information criterion)?

- do wyboru konkretnej realizacji modelu
- jako funkcja straty w uczeniu nadzorowanym
- do zwiększenia wartości likelihood dla struktur bardziej złożonych
- by nie dopuścić do regularyzacji

Q9

What does it mean a prior is conjugate? Give examples of at least two conjugate priors and their likelihood distributions

Q9

What does it mean a prior is conjugate? Give examples of at least two conjugate priors and their likelihood distributions

The posterior has the same form as the prior, e.g.:

- lk: Binomial, prior: Beta
- lk: Multinomial, prior: Dirichlet

Q10

Czy ktoś już się uczył?