

1. Podaj wzór na prawdopodobieństwo warunkowe, a następnie wyprowadź regułę łańcuchową prawdopodobieństwa.

2. Zastosuj regułę łańcuchową do prawdopodobieństwa $P(A, B, C, D)$.

3. Wyraż $P(w_2, w_3, w_4 | w_1, w_0)$ tylko za pomocą rozkładu łącznego.

4. Mając następujący zbiór uczący uzyskany z rzutów kostką sześcienną

1, 1, 5, 2, 2, 6, 2, 3, 6, 3, 4

. Wyestymuj prawdopodobieństwo uzyskania poniższych wartości:

$$P(X = 1) = \quad P(X = 2) = \quad P(X = 3) = \quad P(X = 4) = \quad P(X = 5) = \quad P(X = 6) =$$

Następnie zakładając, że podany zbiór jest sekwencją policz prawdopodobieństwa występowania po sobie różnych wartości $P(X_t | X_{t-1})$

$$P(X_t = 2 | X_{t-1} = 6) = \quad P(X_t = 1 | X_{t-1} = 1) = \quad P(X_t = 3 | X_{t-1} = 2) = \quad P(X_t = 1 | X_{t-1} = 5) =$$

5. Dla podanego zbioru trzech sekwencji: „Ala ma kota”, „Jurek ma kota”, „Kamil ma psa” wyznacz prawdopodobieństwo sekwencji „Ala ma kota” i „Kamil ma kota”.

$$P(\text{Ala ma kota}) =$$

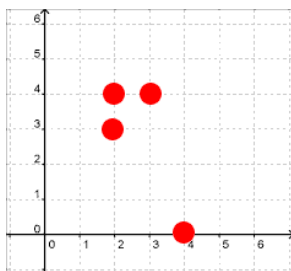
$$P(\text{Kamil ma kota}) =$$

6. Za rozmiar słownika przyjmij $|V|$ i spróbuj obliczyć liczbę parametrów: „pełnego” rozkładu $P(w_1, w_2, w_3, \dots, w_n)$ (tj. rozkładu łącznego przed zastosowaniem reguły łańcuchowej) oraz tego samego rozkładu po zastosowaniu reguły łańcuchowej.

7. Wyjaśnij na czym polega założenie Markowa.



8. Zapisz rozkład sekwencji $P(w_1, w_2, w_3, \dots, w_n)$ przy użyciu reguły łańcuchowej i założenia Markowa trzeciego rzędu
9. Korzystając z postaci rozkładu uzyskanej w poprzednim zadaniu - w jaki sposób wygenerować losową sekwencję?
10. Zakładając modelowanie rozkładu prawdopodobieństwa z założeniem Markowa drugiego rzędu na zbiorze sekwencji „Ala ma kota”, „Jurek ma kota”, „Kamil ma psa”, oblicz prawdopodobieństwo sekwencji „Kamil ma kota”.
11. Zakładając, że w n -elementowych sekwencjach występuje $|V|$ unikalnych słów – ile parametrów będzie miał model z założeniem Markowa rzędu: pierwszego, drugiego, trzeciego i czwartego?
12. Rozważ poniżej zwizualizowany 4-elementowy zbiór danych i wykonaj na nim operację grupowania algorytmem AHC. Zastosuj metrykę Manhatańską oraz single linkage. Wyniki grupowania przedstaw na dendogramie.



Fundusze Europejskie
Polska Cyfrowa



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

