

Zestaw pytań z krótkimi odpowiedziami z przedmiotu Metody Inżynierii Wiedzy:

1. Czym jest reguła Bayesa? Do czego stosuje się regułę Bayesa?

Metoda obliczania prawdopodobieństwa zdarzenia na podstawie znajomości warunkowego prawdopodobieństwa innych zdarzeń. Stosuje się ją do aktualizacji prawdopodobieństw na podstawie nowych danych.

2. Co to jest sieć Bayesowska? Jakie są elementy sieci Bayesowskiej?

Graficzny model probabilistyczny reprezentujący zmienne losowe i ich zależności. Węzły (zmienne) i krawędzie (zależności).

3. Czym jest ukryty łańcuch Markowa (HMM)? Jakie są podstawowe elementy HMM? Do czego stosuje się ukryte łańcuchy Markowa?

HMM to Model statystyczny, w którym system jest opisywany przez sekwencję stanów, które są ukryte, a obserwacje są widoczne. Podstawowe elementy to stany, obserwacje, prawdopodobieństwa przejść i prawdopodobieństwa emisji. HMM stosuje się do modelowania sekwencyjnych danych, takich jak rozpoznawanie mowy czy analiza sekwencji biologicznych.

4. Czym jest wektor stacjonarny? Jakie jest zastosowanie wektora stacjonarnego?

Wektor, który nie zmienia się w wyniku działania macierzy przejścia w łańcuchu Markowa. Stosuje się go do określania długoterminowych proporcji stanów w łańcuchu Markowa.

5. Czym jest zadanie klasyfikacji? Czym jest klasyfikacja binarna? Czym jest klasyfikacja wieloklasowa?

Proces przypisywania danych wejściowych do jednej z predefiniowanych klas. Klasyfikacja binarna, w której dane wejściowe są przypisywane do jednej z dwóch możliwych klas. Klasyfikacja wieloklasowa, w której dane wejściowe mogą być przypisane do jednej z wielu klas.

6. Jakie są różnice między klasyfikacją, regresją i klastrowaniem?

Klasyfikacja przypisuje do klas, regresja przewiduje wartości ciągłe, klastrowanie grupuje dane bez etykiet.

7. Czym jest algorytm KNN? Jak działa KNN?

Algorytm k najbliższych sąsiadów, używany do klasyfikacji i regresji. Przypisuje klasę na podstawie większości klas wśród k najbliższych sąsiadów.

8. Czym jest algorytm k-means clustering? Jak działa k-means clustering?

Algorytm klasteryzacji dzielący dane na k grup w taki sposób, że każdy punkt należy do najbliższego centroidu. Iteracyjnie przypisuje punkty do najbliższego centroidu, a następnie aktualizuje położenie centroidów.

9. Czym jest perceptron? Jak działa perceptron?

Najprostszy model sieci neuronowej, używany do klasyfikacji liniowej. Sumuje wejścia, mnoży przez wagi, dodaje bias i przepuszcza przez funkcję aktywacji.

10. Czym jest regresja logistyczna? Jak działa regresja logistyczna?

Model statystyczny do binarnej klasyfikacji, który wykorzystuje funkcję logistyczną do modelowania prawdopodobieństwa przynależności do klasy. Oblicza prawdopodobieństwo klasy za pomocą funkcji sigmoidalnej na liniowej kombinacji wejść.

11. Co to jest bias (stronniczość)? Co to jest wariacja? Co to jest błąd modelu?

Bias to tendencja modelu do systematycznego błędu w przewidywaniu. Wariacja to wrażliwość modelu na zmiany w danych treningowych. Błąd modelu to rozbieżność między przewidywaniami modelu a rzeczywistymi wartościami.

12. Czym jest drzewo decyzyjne? Jakie są elementy drzewa decyzyjnego? Czym jest wskaźnik rozgałęzienia?

Model klasyfikacji lub regresji, który używa struktury drzewa do podejmowania decyzji. Składowe to węzły decyzyjne, gałęzie i liście (klasy). Wskaźnik rozgałęzienia to miara jakości podziału danych w węźle drzewa decyzyjnego.

13. Czym jest algorytm lasów losowych? Jak działa las losowy?

Zespół drzew decyzyjnych używany do poprawy dokładności klasyfikacji i redukcji przeuczenia. Trenuje wiele drzew decyzyjnych na losowych podzbiorach danych i średnią wyniki.

14. Czym jest uczenie zespołów?

Metoda łączenia wielu modeli w celu uzyskania lepszych wyników predykcji.

15. Czym jest regresja liniowa? Jak działa regresja liniowa?

Model statystyczny do przewidywania wartości ciągłych na podstawie liniowej kombinacji zmiennych wejściowych. Minimalizuje sumę kwadratów różnic między przewidywanymi a rzeczywistymi wartościami.

16. Czym są modele parametryczne? Czym są modele nieparametryczne?

Modele parametryczne mają ustaloną liczbę parametrów niezależnie od ilości danych. Modele nieparametryczne mogą mieć liczbę parametrów zależną od ilości danych.

17. Co to jest interpolacja? Co to jest aproksymacja?

Interpolacja to metoda przewidywania wartości w obrębie zakresu danych. Aproksymacja to metoda przewidywania wartości na podstawie funkcji przybliżającej dane.

18. Co to jest metoda najmniejszych kwadratów?

Metoda estymacji parametrów modelu, która minimalizuje sumę kwadratów błędów.

19. Czym jest współczynnik determinacji (r^2)?

Miara jakości dopasowania modelu regresji, wskazująca procent wariancji wyjaśnionej przez model.

20. Czym są sieci neuronowe?

Modele parametryczne inspirowane strukturą i funkcjonowaniem mózgu, używane do uczenia maszynowego.

21. Co to jest wsteczna propagacja?

Algorytm używany do trenowania sieci neuronowych przez minimalizację błędu za pomocą gradientu.

22. Czym są sieci konwolucyjne (CNN)? Jakie są główne elementy CNN?

Sieci neuronowe stosowane głównie do analizy obrazów, które wykorzystują operacje konwolucji. Główne elementy CNN to warstwy konwolucyjne, warstwy poolingowe i warstwy w pełni połączone.

23. Czym są sieci rekurencyjne (RNN)?

Sieci neuronowe przeznaczone do przetwarzania sekwencyjnych danych, takich jak tekst czy dane czasowe.

24. Co to jest autodekoder, enkoder, dekodek?

Autodekoder to rodzaj sieci neuronowej używanej do uczenia efektywnych reprezentacji danych poprzez odtwarzanie wejściowych danych na wyjściu. Enkoder to komponent sieci neuronowej, który przekształca dane wejściowe w zredukowaną reprezentację wektorową. Dekoder to komponent sieci neuronowej, który przekształca zredukowaną reprezentację wektorową z powrotem w dane wyjściowe.