

Student: **Michał Filek**

Promotor: **dr Marek Śmieja**

Recenzent: **prof. dr hab. Igor Podolak**

Specjalność: **nauczanie maszynowe**

Temat pracy: **Text classification using the semi-supervised methods**

1. Ekstrema funkcji dwóch zmiennych: definicja i sposoby znajdowania.
2. Dyskretne zmienne losowe oraz ich najważniejsze rozkłady.
3. Graf eulerowski, graf hamiltonowski, liczba chromatyczna grafu; definicje i związane z tymi pojęciami twierdzenia.
4. Wektory i wartości własne macierzy; numeryczne algorytmy ich wyznaczania.
5. Rozwiązanie układów równań liniowych: metoda eliminacji Gaussa i metoda Gaussa-Seidla.
6. Pojęcia relacji równoważności i zbioru ilorazowego.
7. Omówić pojęcia przypadku użycia systemu i scenariusza przypadku użycia.
8. Omówić znane metody synchronizacji procesów.
9. Przedstawić i omówić hierarchię Chomsky'ego ze szczególnym uwzględnieniem występujących w niej definicji gramatyk.
10. Model ISO OSI. Przykłady protokołów w poszczególnych warstwach.
11. Podstawowe metody indeksowania w systemach baz danych.
12. Omówić metody reprezentacji grafów oraz metody przeglądania (DFS,BFS).
13. Omówić znane szybkie algorytmy sortowania przez porównanie.
14. Omówić znane algorytmy znajdowania najkrótszych ścieżek w grafie.
15. Zdefiniować pojęcie semantycznej poprawności algorytmu (częściowa poprawność, własność stopu, określoność obliczeń).
16. Zdefiniować pojęcia asercji i niezmiennika; omówić sposób wykorzystania tych pojęć w dowodzeniu semantycznej poprawności algorytmów.
17. Na czym polega programowanie dynamiczne? Własności i przykłady zastosowań.

## Analiza Danych

- \* Proszę porównać podejścia do k-means Lloyd'a vs Hartigana.
- \* Proszę opisać klastrowanie DBScan.
- \* Proszę omówić PCA, wraz z przykładem zastosowania.

## Bioinformatyka

- \* Problem zarządzania informacją biologiczną. Bazy danych pierwotne i wtórne.
- \* Dopasowanie par sekwencji - istota problemu, algorytmy Needlemana-Wunscha i Smitha-Watermana, metody heurystyczne.
- \* Dopasowania wielosekwencyjne.
- \* Cele i metody analizy filogenetycznej.

## Nauczanie maszynowe

- \* Proszę przedstawić regułę uczenia Bayesowskiego. Proszę uwzględnić w odpowiedzi definicję pojęcia likelihood, reguły Bayesa, pojęć prior i posterior. Czym różni się podejście Bayesowskie od podejścia Maximum Likelihood?
- \* Proszę przedstawić podstawowe podejścia do problemu regresji z wykorzystaniem metod uczenia maszynowego. Co oznacza pojęcie modelu liniowego? Co to są funkcje bazowe? Proszę przedstawić rozwiązanie problemu za pomocą metod spadku gradientu i trzy przykładowe algorytmy.
- \* Proszę zdefiniować pojęcie drzewa decyzyjnego w uczeniu maszynowym. W jaki sposób takie drzewo jest budowane? Czy drzewa decyzyjne można wykorzystać zarówno w problemach regresji jak i klasyfikacji? W jaki sposób modele drzew decyzyjnych można wykorzystać w modelach typu ensemble?

## Obliczalność i złożoność

- \* Wymienić kilka modeli obliczeń równoważnych maszynie Turinga. Uzasadnić, że niedeterministyczna maszyna Turinga jest obliczeniowo równoważna deterministycznej.
- \* Problemy rozstrzygalne, częściowo rozstrzygalne i nierozstrzygalne; podać przykład problemów z każdej z tych klas. Podać dowód nierozstrzygalności przykładowego problemu.
- \* Definicja złożoności obliczeniowej. Klasy złożoności: P, NP, PSPACE, NSPACE; zupełność w klasie. Podać dowód NP-zupełności przykładowego problemu.

## Programowanie niskopoziomowe

- \* Potokowe przetwarzanie instrukcji we współczesnych procesorach - zyski i zagrożenia.
- \* Konwencje wywołania funkcji, różnice między cdecl a System V AMD64 ABI.
- \* Operacje wektorowe SSE.

## Programowanie w logice

- \* Przedstaw zalety i wady oraz przykłady zastosowań programowania w logice.
- \* Co to są klauzule Horna? Przedstaw szkic algorytmu SLD rezolucji.

## Przetwarzanie języka naturalnego

- \* Na czym polega naiwny algorytm bayesowski klasyfikacji tekstów? Co w nim zakładamy i jaką funkcję minimalizujemy?
- \* Omów jeden z algorytmów word2vec lub GloVe.

## Statystyka bayesowska

- \* Podaj twierdzenie Bayesa. Czym różni się wnioskowanie bayesowskie od wnioskowania częstotliwościowego stosowanego w podejściu klasycznym?
- \* Co to jest funkcja wiarygodności? Czym różnią się funkcje wiarygodności w statystyce klasycznej, a bayesowskiej?
- \* Jaka jest różnica między weryfikacją hipotez w podejściu klasycznym, a w podejściu bayesowskim?
- \* Do czego służy metoda Monte Carlo oparta na łańcuchach Markowa?
- \* Co to jest próbnik Gibbsa? Kiedy nie można użyć próbnika Gibbsa?

## **Teoria informacji w nauczaniu maszynowym**

- \* Związek pomiędzy entropią Shannona i kodowaniem bezstratnym - przykład algorytmu kodowania bezstratnego.
- \* Wzajemna informacja - definicja, własności oraz zastosowania.
- \* Zasada maksymalnej entropii - sformułowanie problemu oraz przykład zastosowania.

## **Zarządzanie projektami IT**

- \* Inżynieria wymagań w projekcie: wymień przykładowe metody pozyskiwania, dokumentowania, walidacji i zarządzania wymaganiami.
- \* Omów metodę PERT/CPM.
- \* Omów model kosztu jakości (Cost of Quality).

## **Pytania własne Recenzenta**

.....  
.....