Predikovanie prežitia pasažierov Titanicu

Fakulta matematiky, fyziky a informatiky

Univerzity Komenského

Dátová veda:

Tomáš Varga, Erik Kolesár, Tadeáš Kaminský

Podiel Práce: ⅓, ⅓, ⅓

*Obsah*

[**Zadanie 2**](#_eqc0vdpwyuzn)

[**Časť A 3**](#_grniy0gfpq0p)

[Dôkaz: 3](#_9qxawhgmmtaf)

[Úloha LP: 5](#_ycl3d3o5h97d)

[**Časť B 6**](#_dhheaxod4bii)

# 

# 

# Zadanie

Jednou z typických úloh v strojovom učení je úloha klasifikácie dát. Dané

sú dve množiny dát



V lineárnej klasifikácii ide o nájdenie nadroviny danej vektorom *a* a skalárom *b*, ktorá dané dáta separuje, t.j. platí:



Keď že tieto nerovnosti sú homogénne v a, b (t.j. možno ich násobiť ľubovoľným kladným číslom a nič sa nezmení), možno ich ekvivalentne vyjadriť ako:



Takáto úloha sa dá riešiť ako úloha LP s premennými a, b a konštantnou (nulovou) účelovou funkciou.

# Časť A

Pre riešenie zadaného problému, bolo trebalo 2 základné homogenne nerovnosti s ostrou nerovnosťou premeniť na nerovnosti v ktorých sa nerovnosť zmení z ostrej na neostrú a pravá strana sa zmení z 0 na 1 a -1. Tento problém sme sa rozhodli riešiť spôsobom výberu najmenšieho možného ε kde



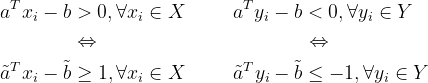
teda ε je najmenšie možné číslo z množiny možných výsledkov ktoré bude vždy kladné kvôli absolutnej hodnote. Tým, že dané číslo je najmenšie zo všetkých možných nám zaruči to že, po vynásobení rovníc ε sa nestane to, že by pre ani jednu rovnicu neplatila nerovnosť ktorú sa snažíme ukázať čiže.



z pôvodných homogénnych nerovností sa teda stanú nami potrebné rovnice.

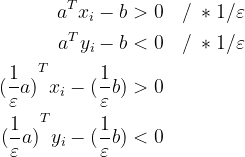
## Dôkaz:

Chceme teda dokázať ekvivalenciu 2 výrazov:



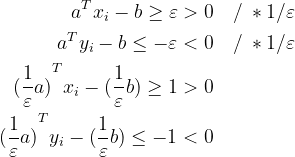
kde 

postupnou úpravou dostávame teda



Následne si môžeme všimnúť že sme sa stále nezbavili ostrých nerovností. Tu nastáva najdôležitejšia časť našej úvahy a dokazu keďže sme zadefinovali ε ako najmenší možný výsledok vieme našu pôvodnú úpravu spraviť znova a lepšie.

Čiže



Ako si môžeme všimnúť využili sme spomenutú uvahu toho, že ε je najmenší možný výsledok. To sme aplikovali tým, že sme povedali že všetky výsledky sú buď väčšie rovné ako ε ale menšie rovné ako ε. Následným vynásobením 1/ε dostávame nami potrebné rovnice a to teda.



Čo je ekvivalentné s



Čo je to čo sme chceli dokázať.

Dalo by sa povedať že sme nerovnice istým spôsobom naškálovali aby z nich bolo možno spraviť Ulohu Lineárneho Programovania a nasledne odstránili nerovnosti voči nule kedže nerovnosti voči jednotke nestrácajú žiadnu informáciu.

## 

## 

## 

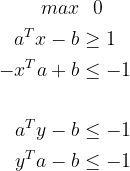
## 

## 

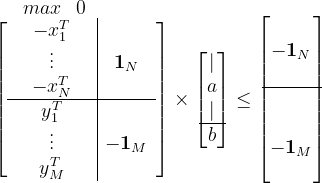
## Úloha LP:

Z nasledujúcich rovníc sme schopný vyjadriť tieto rovnice v Kompaktnom maticovom tvare a to nasledovne

Rovnice:



Kompaktný tvar:



Dôležité je tiež si uvedomiť, že nulová účelová funkcia znamená, že každé prípustné riešenie úlohy je zároveň aj optimálne.

# 

# 

# Časť B

V tejto časti sme sa zamerali na praktickú implementáciu poznatkov získaných v predchádzajúcej časti ([časť A](#_grniy0gfpq0p)) na dáta, pre ktoré sme schopní identifikovať separujúcu nadrovinu. Proces začal transformáciou dát do vhodnej formy, kde sme vytvorili matice X a Y, reprezentujúce body z jednotlivých tried. Tieto matice sme potom upravili tak, aby bolo možné aplikovať lineárne programovanie.

Ďalej sme vytvorili nové matice a následne zostavili maticu A, ktorá zahŕňa všetky potrebné informácie o úlohe lineárneho programovania. Pravá strana nerovnice (b) a účelová funkcia (c) boli ďalej definované podľa potrebných parametrov.

Využili sme rôzne metódy lineárneho programovania, konkrétne 'simplex', 'interior-point', 'highs-ds' a 'revised simplex', pre riešenie úlohy. Pre každú z týchto metód sme získali optimálne riešenie, ktoré obsahuje koeficienty separujúcej priamky. Tieto koeficienty sme využili na vizualizáciu separujúcej priamky v rámci jednotlivých grafov pre každú metódu.

Grafy sme následne zobrazili v mriežke 2x2, kde každý graf predstavuje separujúcu priamku pre jednu z použitých metód. Body z matic X a Y sme farebne odlišili a separujúce priamky sme zobrazili rôznymi farbami a typmi čiar. Na záver sme vytvorili jediný graf, ktorý zahrňuje všetky separujúce priamky pre lepšie porovnanie.

Výsledky ukázali, že tri zo štyroch použitých metód (simplex, highs-ds, revised simplex) poskytujú rovnaké riešenie, čo poukazuje na podobný algoritmický prístup. Naopak, metóda interior-point poskytla odlišnú separujúcu priamku, naznačujúc odlišný spôsob jej výpočtu v porovnaní so simplexovými metódami. Tým sme demonštrovali aplikáciu teoretických konceptov v praxi a porovnali výsledky rôznych metód lineárneho programovania v kontexte našej úlohy klasifikácie.