



VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS FAKULTETAS  
KOMPIUTERIJOS KATEDRA

Baigiamasis bakalauro darbas

**Duomenų dimensiškumo mažinimas ir klasifikavimas**

Atliko:

Donatas Kučinskas

parašas

Vadovas:

Vytautas Valaitis

Vilnius  
2015

# **Turinys**

<b>Sutartinis terminų žodynas</b>	<b>3</b>
<b>Santrauka</b>	<b>4</b>
<b>Summary</b>	<b>5</b>
<b>Ivydas</b>	<b>6</b>
<b>1. Dirbtinių neuronų tinklas</b>	<b>7</b>
<b>2. Dirbtinių neuronų tinklas</b>	<b>7</b>
2.1. Dirbtinis neuronas . . . . .	7
2.2. Dirbtiniai neuronai/tinklas?TODO . . . . .	8
<b>3. Dimensiškumo mažinimas</b>	<b>8</b>
<b>4. Iriso gėlių duomenys</b>	<b>9</b>
4.1. Pirmojo skyriaus poskyris . . . . .	9
4.1.1. Pirmojo skyriaus pirmo poskyrio poskyris . . . . .	9
<b>Išvados ir rekomendacijos</b>	<b>10</b>
<b>Ateities tyrimų planas</b>	<b>11</b>
<b>Literatūros šaltiniai</b>	<b>12</b>
<b>Priedai</b>	<b>12</b>
<b>A. Pirmojo priedo pavadinimas</b>	<b>13</b>
<b>B. Antrojo priedo pavadinimas</b>	<b>14</b>

## **Sutartinis terminų žodynas**

Pateikiamas terminų sąrašas (jei reikia)

## **Santrauka**

Santraukos tekstas rašto darbo kalba...

## **Summary**

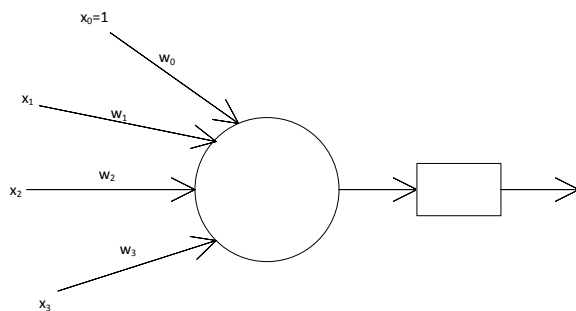
**Darbo pavadinimas kita kalba**

This is a summary in English...

## Ivadas

Klasifikavimas - tai dažnai sutinkama užduotis, turinti įvairių sprendimo būdų. Šios uždavinio tikslas - identifikuoti, kuriai grupei priklauso tiriamas objektas. Tiriamieji objektai dažniausiai būna vienos rūšies, aprašomi tam tikrais parametrais, o grupės, kuriems jie yra priskiriami - iš anksto žinomos. Pavyzdžiui, galima klasifikuoti gyvūnus pagal tam tikras jų fizines savybes - kojų ilgį, storį, kitas kūno apimtis, kailio ilgį ir pan. Natūralu, kad kiekvienas net ir tos pačios rūšies gyvūnas turės šiek tiek kitokius parametrus, tačiau šie parametrai dažniausiai turi įvairius proporcingumus, pagal kuriuos galima bandyti atspėti, kuriai rūšiai tam tikras gyvūnas priklauso.

Norint išspręsti konkretų klasifikavimo uždavinį, paprasčiausias sprendimas atrodo galėtų būti šių grupių parametrų ištyrimas - pavyzdžiui, norint mokėti atskirti triušius nuo liūtų turint jų ilgius nėra sunki užduotis. Tačiau problema kyla, kai atskiriamos klasės yra labai panašios viena į kitą - tokiu atveju pastebėti tam tikrus dėsningumus ir juos sumodeliuoti bei realizuoti ir kur kas sunkiau. Be to, sprendžiant konkretų klasifikavimo uždavinį, tektų gilintis į klasifikuojamus objektus - pavyzdžiui, norint sukurti tam tikrų kiškių rūšių klasifikavimą, gilios žinios apie šias kiškių rūšių savybes būtų privalomos.



1 pav. Dirbtinis neuronas

## 1. Dirbtinių neuronų tinklas

## 2. Dirbtinių neuronų tinklas

Dirbtinis neuronų tinklas - tai tarpusavyje susijungusių dirbtinių neuronų tinklas, kurio užduotis yra spręsti tam tikrą užduotį arba užduotis. Dirbtinis neuronų tinklas gavęs pradinį užduoties duomenį, juos apdoroja ir taip gaunamas tam tikras atsakymas. Šis atsakymas nebūtinai yra teisingas - neuronų tinklai suprojektuoti taip, kad galėtų būti mokomi kai gauna neteisingą atsakymą.

### 2.1. Dirbtinis neuronas

Dirbtinių neuronų tinklas sudarytas iš daugybės dirbtinių neuronų, todėl norint suprasti tinklą, reikia pradėti nuo vieno dirbtinio neurono. Žmogaus smegenys sudarytos iš daugybės neuronų. Dirbtinis neuronas - tai supaprastintas šių biologinių neuronų modelis. Jo modelis pavaizduotas 1 paveiksliuke. Dirbtinio neurono veikimo principas gan paprastas - per kairėje esančias jungtis dirbtinis neuronas gauna signalus iš kitų dirbtinių neuronų - iš  $k$ -tosios jungties gaunamas  $x_k$  dydžio signalas. Šiuos signalus neuronas apjungia ir pertvarko, ir taip sugeneruojamas dirbtinio neurono išeinamasis signalas. Šis išeinamasis signalas gali būti siunčiamas daugybei kitų neuronų - dešinėje esančios jungtys yra neurono išeinamojo signalo jungtys, kuriomis ir yra siunčiamas išeinamasis signalas.

Dirbtinis neuronas generuoja išeinamąjį signalą pagal tam tikrą modelį. Pirmiausia, kiekviena įeinančioji jungtis turi savo svorį  $w_i$  - šis svoris yra padauginamas iš įeinančio signalo dydžio  $x_k$ . Tada visos šios signalų dydžių ir svorių sandaugos yra susumuojamos - taip gaunamas skaičius  $a$  (2.1 formulė). Tada šis skaičius  $a$  yra paduodamas kaip argumentas tam tikrai funkcijai  $f$  ir gaunamas neurono išeities signalas  $y = f(a)$ . Ši funkcija  $f$  yra vadinama aktyvacijos funkcija - ją galima keisti pagal tai, kokio tikslo siekiama iš šio dirbtinio neurono. Populiariausios aktyvacijos funkcijos - slenkstinė, tiesinė, hiperbolinis tangentas bei sigmoidinė (2.2 formulė). Iš esmės aktyvacijos funkcija gali būti bet kokia funkcija, tačiau vėliau norint apmokyti dirbtinį neuronų tinklą, reikia rasti šios funkcijos išvestinę. Dėl šios priežasties dažniausiai pasirenkamos tokios aktyvacijos funkcijos, kurios ne tik tinkamai pertvarko signalą išvedimui, tačiau ir kurios išvestinės yra paprastos.

$$a = \sum_{k=1}^N w_k x_k \quad (2.1)$$

$$f(a) = \frac{1}{1 + e^{-a}} \quad (2.2)$$

TODO: bias

## 2.2. Dirbtiniai neuronai/tinklas?TODO

Visi

[TODO: dirbtinio neurono paveikslukas]

[TODO: citata?]

[TODO: 110 iš knygos]

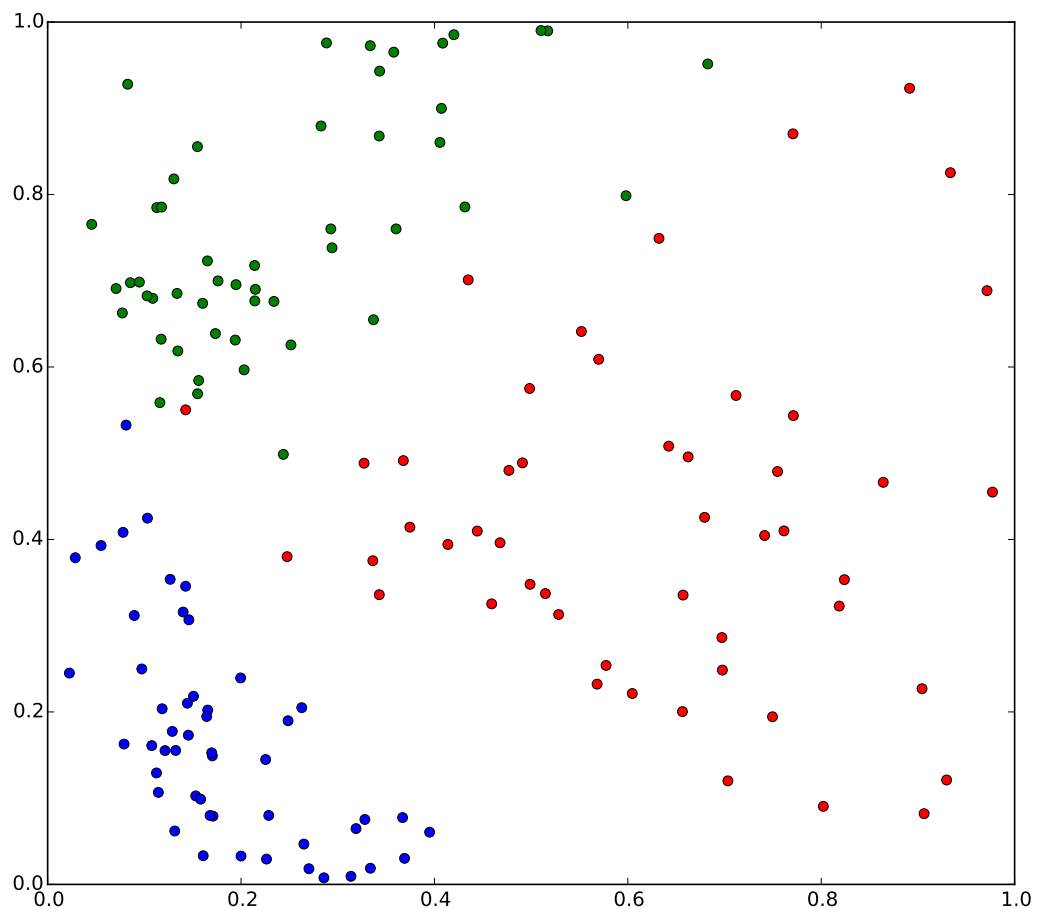
## 3. Dimensiškumo mažinimas

Klasifikavimo problema

Galimi sprendimai:

\* statistinis sprendimas \* neuroniniai tinklai \* veikimas \* apmokymas \* validavimas? \* Klasifikavimas požymių išskyrimui \* Dimensiškumo mažinimas -> Klasifikavimas





## 4. Iriso gėlių duomenys

Programuojant neuroninius tinklus, testavimui buvo panaudoti vilkdagių (angl. Iris flower) duomenys. Tai plačiai taikomi ir viešai pasiekiami duomenys, aprašantys 3 rūšių vilkdagius. Aprašyta po 50 kiekvienos rūšies vilkdagių. Kiekvienas vilkdagis aprašomas pateikiant 4 dydžius: taurėlapio ilgis, taurėlapio plotis, vainiklapio ilgis bei vainiklapio plotis. Šiuos vilkdagių duomenis sudaro 150 gėlių, kurių kiekviena aprašyta 4 parametrais bei priskirta vienai iš 3 vilkdagių grupių.

Šie duomenys puikiai tinka klasifikavimo tinklo apmokymui - tinklo tikslas yra kuo mažiau klystant pasakyti, kuriai iš 3 vilkdagių rūšių tam tikra gėlė su tam tikrais parametrais priklauso. Be to, yra pakankamai duomenų, kad būtų galima dalį jų panaudoti tinklo apmokymui, o kitą dalį - testavimui. Tokiu būdu bus užtikrinama, kad tinklas teisingai išmoko atskirti vilkdagių rūšis pagal parametrus, o ne tiesiog prisitaikė prie mokymui panaudotų duomenų.

### 4.1. Pirmojo skyriaus poskyris

Pateikiamas 4.1 poskyrio tekstas. Vienas iš šaltinių [?]. Visas turinys priklauso ?? skyriui.

#### 4.1.1. Pirmojo skyriaus pirmo poskyrio poskyris

Pateikiamas trečio lygio poskyrio tekstas.

1 lentelė. Lentelė ...

test	test
test	test

## **Išvados ir rekomendacijos**

Išvados bei rekomendacijos.

## **Ateities tyrimų planas**

Pristatomi ateities darbai ir/ar jų planas, gairės tolimesniems darbams....

# Priedai

Dokumentą sudaro du priedai: A priede ....

## **A. Pirmojo priedo pavadinimas**

Pirmojo priedo tekstas ...

## **B. Antrojo priedo pavadinimas**

Antrojo priedo tekstas ...