

II užduotis (Vieno neurono (perceptrono) mokymas sprendžiant klasifikavimo uždavinį)

Užduoties tikslas – apmokyti vieną neuroną spręsti dviejų klasių uždavinį, atlikti tyrimą su dviem duomenų aibėmis.

Šiai užduočiai reiks naudoti dvi duomenų aibes (Irisų ir Krūties vėžio):

- **Irisų duomenų aibę** galima parsisiųsti iš <https://archive.ics.uci.edu/ml/datasets/iris>. Šiuose duomenyse yra trys klasės Setosa, Versicolor ir Virginica. Analizei reikia imti tik dvi: Versicolor ir Virginica.
- **Krūties vėžio duomenų aibę** galima parsisiųsti iš [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)). Šiuose duomenyse yra dvi klasės: 2 – nepiktybinis navikas, 4 – piktybinis navikas.

Užduoties punktai:

1. Parsisiųsti ir paruošti duomenis. Klasių žymes (*label*) pakeiskite 0 arba 1. Ištrinkite nereikalingus atributus (stulpelius), pvz., ID. Krūties vėžio duomenyse yra trūkstamų reikšmių, tose vietose yra klaustukai (?), todėl tokius įrašus (eilutes) reikia ištrinti.
2. Sukurti programą (parašyti kodą), kuri įgyvendintų vieno neurono (perceptrono) mokymo ir testavimo procesą, sprendžiant klasifikavimo uždavinį. Galima naudoti bet kurią programavimo kalbą. Turi būti įgyvendintas neurono mokymas ir testavimas, todėl turimi duomenys turi būti padalinami į mokymo ir testavimo aibes. Įprastai santykis tarp mokymo ir testavimo aibių yra 80:20 arba 70:30.

Programoje turi būti įgyvendinta:

- Duomenų nuskaitymas iš failo.
- Galimybė keisti tokius hiperparametrus, kaip mokymo greitis (*learning rate*), epochų skaičius.
- Dvi aktyvacijos funkcijos: ir slenkstinė, ir sigmoidinė.

Programos rezultatas turi būti:

- Gauti svoriai (analizuojant irisų duomenis, būtų penki svoriai w_0, w_1, w_2, w_3, w_4 , čia w_0 – poslinkis (bias); analizuojant krūties vėžio duomenis, būtų 10 svorių).
- Klasifikavimo tikslumas. Turi būti apskaičiuotas tikslumas ir mokymo, ir testavimo duomenims.

Klasifikavimo tikslumo matas gali būti santykis tarp teisingai klasifikuotų ir visų duomenų. Norint jį gauti, reikia kiekvienam duomenų įrašui paskaičiuoti klasę pagal gautus neuronų svorius. Naudojant slenkstinę funkciją, neurono išėjime gaunamos reikšmės 0 arba 1, tad duomenų priskyrimas klasėms yra akivaizdus. Naudojant sigmoidinę funkciją, neurono išėjimo reikšmės yra intervale (0; 1), tad vertinant klasifikavimo rezultatą, šias reikšmes reiktų suapvalinti iki artimiausio sveiko skaičiaus (0 arba 1).

- Gautos paklaidos po kiekvienos epochos mokymo duomenims ir paklaida testavimo duomenims.

Paklaidą reikia skaičiuoti, skirtumą tarp perceptrono gautų ir trokštamų reikšmių pakėlus kvadratu. Skaičiuojant paklaidą, sigmoidinės funkcijos reikšmių apvalinti nereikia.

3. Atlikti tyrimus ir vienai, ir kitai duomenų aibei (Irisų ir Krūties vėžio). Kiekvieno tyrimo rezultatus (klasifikavimo tikslumas ir paklaida) pateikti lentelėse arba grafikuose su atitinkamais komentarais.

Tie rezultatai, kuriuos lengviau ir greičiau galima suprasti iš grafikų, jie turi būti pateikti grafikuose, pvz., rezultatų priklausomybė nuo tiriamų hiperparametrų.

Tyrimų metu būtina nustatyti ir aprašyti:

- Kaip klasifikavimo tikslumas priklauso nuo epochų skaičiaus? Pateikti grafiką, kurio x ašyje yra atidėdamos epochos, o y ašyje klasifikavimo tikslumas mokymo duomenims.
- Kaip paklaidos reikšmės priklauso nuo epochų skaičiaus? Pateikti grafiką, kurio x ašyje yra atidėdamos epochos, o y ašyje paklaidų reikšmės mokymo duomenims.
- Kaip rezultatai priklauso nuo skirtingų mokymosi greičio reikšmių? Vaizdavimo būdą (lentelės, grafikai) galima pasirinkti patiems.
- Kaip rezultatai priklauso nuo to, kuri aktyvacijos funkcija yra naudojama? Vaizdavimo būdą (lentelės, grafikai) galima pasirinkti patiems.

Užduoties ataskaitoje:

- Aprašyti, kokie duomenys buvo naudojami, kiek yra duomenų įrašų (eilučių), požymių (stulpelių), kaip duomenys buvo padalinti į mokymo ir testavimo aibes. Pateikti kitą, jūsų manymu, svarbią informaciją;
- Paaiškinti kas yra epocha.
- Paaiškinti, kaip buvo parenkamos pradinio svorių reikšmės.
- Pateikti programos tekstą su išsamiais komentarais.
- Detaliai aprašyti atlikto tyrimo rezultatus (žr. 3 punktą). Aprašykite, kaip rezultatai skiriasi skirtingoms duomenų aibėms.
- Nustačius variantą, kada gaunami tiksliausi klasifikavimo rezultatai ir mažiausia paklaida, pateikti gautus rezultatus: svorius, paklaidas, epochų skaičius, klasifikavimo tikslumo įverčius, paklaidą. Kiekvienam testavimo aibės įrašui nurodyti, kokias klases nustatė neuronas ir kokia turėjo būti.
- Pateikti atliktų tyrimų išvadas. Išvados turi būti aiškios ir lakoniškos, tiksliai atspindinčios gautus rezultatus. Išvadų apimtis – 4–10 sakinių.

P. S. Ataskaitoje turi būti aprašytas kiekvienas atliekamas veiksmas, pateikti žymėjimų aprašymai ir kita, jūsų manymu, svarbi informacija.