

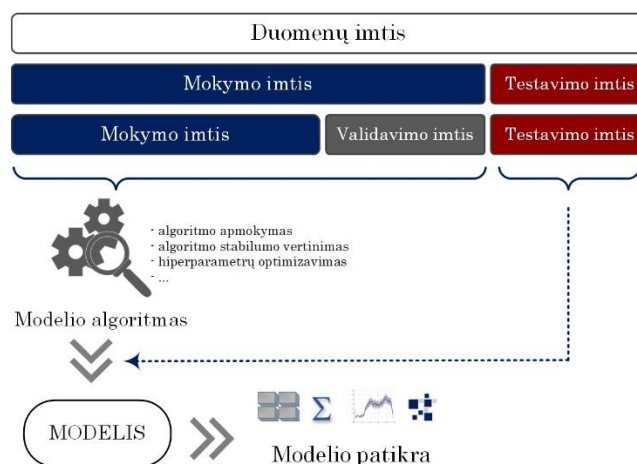
Prižiūrimojo mašininio mokymosi modelio patikra

I tema. Modelio patikros bendrieji principai

Tarkime, kad modelis kuriamas tam tikrai sprendimų priėmimo sistemai automatizuoti ar tiesiog rekomendacijoms modelio pagrindu gauti. Įprastai išbandomas ne vienas modelis ir pasirenkamas būtent tas, kuris testavimo metu turėjo aukščiausią įvertinimą. Žinoma, dar prieš tai reikia nuspręsti, kurie modeliai tinka esamai problemai spręsti ir kiek skirtingų modelių bus išbandoma. Kitas svarbus klausimas – kokia modelio (-ių) testavimo strategija bus taikoma bei kokios klaidų metrikos bus skaičiuojamos. Šioje temos dalyje susipažinsime su dažniausiai taikoma prižiūrimojo mašininio mokymosi modelio testavimo strategija ir keliomis jų alternatyvomis.

Natūralu, jog visais atvejais tikėtis tobulai veikiančio mašininio mokymo modelio neverta. Faktas tas, kad kiekvienas modelis generuoja tam tikrą klaidų kiekį (angl. *error rate*). Pavyzdžiui, dvireikšmio kintamojo (arba tiesiog dviejų klasių) klasifikavimo uždavinyje modelis spėja naujo stebėjimo klasę. Jei teisingai, tai sėkmė; jei neteisingai, tai klaida. Klaidų kiekis gali būti vertinamas mokymo imtyje esantiems stebėjimams, bet tai daryti yra netikslinga dėl to, kad būtent naudojant mokymo imtį yra mokomas modelis ir jo tinkamumo nematytiems stebėjimams tokiu būdu neįvertinsime. Svarbu tai, kad atliekant modelio patikrą klaidų kiekis turi būti skaičiuojamas testavimo imtyje esantiems stebėjimams. Šią imtį sudaro modelio mokymo metu nenaudoti stebėjimai. Be abejo, modelio mokymo imties ir testavimo imties kintamieji X turi būti suderinti, t. y. testuojant naudojami tokie patys kintamieji X , kaip ir modelį mokant.

1 pav. pateikta modelio patikros principinė schema.



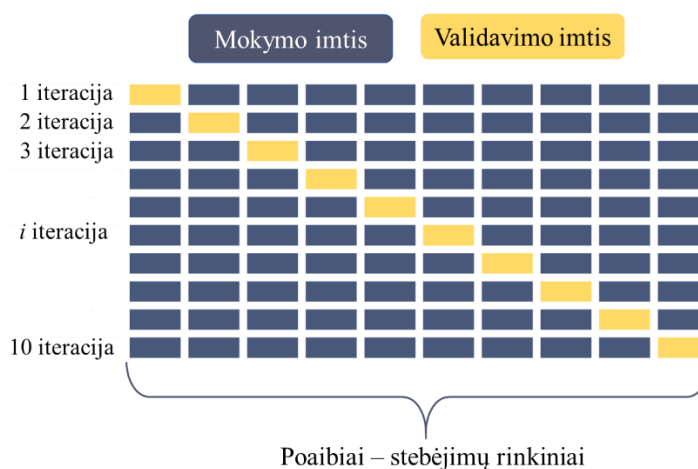
1 pav. Modelio patikros schema

Pagal pateiktą schemą, visas modelio kūrimo procesas prasideda nuo to, jog paruošiama tam tikra duomenų imtis. Atsitiktiniu būdu ši imtis yra dalijama į mokymo ir testavimo imtis. Gan primygtinai yra rekomenduojama mokymo imtį skelti į dar atskiras imtis: mokymo ir validavimo. Tiek validavimo, tiek testavimo imtyje saugomi modeliui nematyti stebėjimai. Tačiau skirtumas yra tas, kad mokymo ir validavimo imtyse stebėjimai yra maišomi tam tikru būdu ir daug kartų yra vykdoma modelio patikra, optimizuojami tam tikrų algoritmų hiperparametrai ar stebimas modelio stabilumas. Testavimo imtis naudojama tik vieną kartą ir tada galiausiai apibendrinami modelio patikros

rezultatai lentelėmis, grafikais ar tam tikromis išvalgomis. Tyrime naudojant daug modelių, ši principinė schema taikoma kiekvienam iš jų, ir tokiu būdu išrenkamas geriausias modelis.

Bendru atveju vienas iš esminių dalykų yra tyrimui paruoštos duomenų imties skaidymas į mokymo, validavimo ir testavimo imtis. Taigi, keliam klausimą, kaip dažniausiai yra skaidoma duomenų imtis. Jei imtis dalijama į mokymo ir testavimo, tai siūloma išlaikyti proporcijas 60:40, 70:30, 80:20. Jei duomenų imtis dalijama į mokymo, validavimo ir testavimo imtis, tai galimi padalijimo būdai pagal proporcijas 60:20:20, 70:20:10.

Esu paminėjusi, jog mokymo ir validavimo imtyse stebėjimai yra maišomi tam tikru būdu ir daug kartų yra vykdoma modelio patikra. Vienas iš populiariausių šios idėjos realizavimo būdų yra kryžminė patikra (2 pav.).



2 pav. Kryžminės patikros realizavimo schema

Pagal šį metodą mokymo imtis suskaidoma į panašaus dydžio poaibius (angl. *fold*) ir mokymui naudojami visi poaibiai, išskyrus vieną, kuris rezervuojamas kryžminei patikrai. Kiekvienos iteracijos metu testavimui skirtas poaibis – validavimo imtis – yra keičiama, todėl ir vadinama kryžmine patikra. Analitikas pats sprendžia, į kiek poaibių turi būti padalijama imtis. Tai nulemia iteracijų skaičių. Dažniausiai nusistovėjusi praktika yra duomenų imtį skaidyti į 10 poaibių.

Tačiau dalijant mokymo imtį į poaibius bendruoju atveju gali būti neužtikrinamas klasių reprezentatyvumas gautose mažesnėse imtyse. Pavyzdžiui, dviejų reikšmių klasifikavimo uždavinyje negalime būti visiškai įsitikinę, jog kiekviename po padalijimo gautame poaibyje yra abiejų klasių stebėjimai. Dar daugiau, keliam klausimą, koks tų klasių proporcijų santykis kiekviename iš tų poaibių. Todėl rekomenduojama vykdyti stratifikuotą kryžminę patikrą, kuri leidžia užtikrinti maždaug tokią pačią kiekvienos klasės stebėjimų proporciją kiekviename poaibyje.

Kryžminės patikros metu yra skaičiuojamos klaidų metrikos. Tačiau tos pačios metrikos įverčių gaunama tiek, kiek yra vykdoma iteracijų. Pavyzdžiui, skaidant imtį į 10 poaibių gaunama 10 įverčių. Rezultatams apibendrinti gali būti apskaičiuojama klaidų metrikų vidurkis ir standartinis nuokrypis ar braižoma stačiakampė diagrama. Pavyzdžiui, gauta didelė klaidų metrikos įverčio sklaida gali signalizuoti modelio jautrumą tam tikroms kintamųjų kombinacijoms ar stebėjimų reikšmėms ir

Masinis atvirasis internetinis kursas „Dirbtinis intelektas“

sąlygoti nestabilų tolimesnį modelio veikimą. Tokiu būdu kryžminės patikros metu yra įvertinamas modelio veikimas skirtingoms mokymo ir validavimo imtims.

Taigi, apibendrinkime tai, kad buvo išsakyta šioje temoje:

- modelio validavimo ar testavimo metu apskaičiuoti klaidų metrikų įverčiai leidžia nujausti, kiek patikimai veikia sudarytas modelis ir ar galima pasikliauti jo gaunamomis prognozėmis ar spėjimais;
- yra siūloma keletas modelio validavimo ar testavimo strategijų;
- išbandant keletą mašininio mokymosi modelių, prioritetas teikiamas tikslesniam, mažiau klaidų generuojančiam ir stabiliau veikiančiam modeliui;
- pastebėsime tai, jog jei dauguma išbandytų modelių rodo prastus modelio patikros rezultatus, tai gali reikšti, jog galimai duomenyse nėra pakankamai informacijos, iš kurios modelis galėtų mokytis, arba reikia išbandyti pažangesnius mokymosi algoritmus. Taip pat vertėtų apsvarstyti duomenų kokybės klausimą ar paieškoti naudingesnių (naujų) kintamųjų modeliams mokytis;
- be to, būtent kryžminės patikros metu gali būti optimizuojami mašininio mokymosi algoritmo hiperparametrai ar aptinkamas persimokymas. Šie klausimai nėra įtraukti į šį kursą.

Kitose šios temos dalyse susipažinsime su prižiūrimojo mašininio mokymosi modelio patikros metu skaičiuojamomis klaidų metrikomis. Pastebėsime tai, jog metrikų naudojimas skiriasi atsižvelgiant į tai, ar tikslinis kintamasis yra kategorinis (klasifikavimo modelis) ar vis dėlto tolydusis (regresijos modelis).