

# GMM egzaminas 2023

## 1 Testas (1 b.)

1. Rinkinio (angl. *batch*) normalizavimas:

- (a) Turi fiksuotus hiperparametrus – vidurkjį ir dispersiją
- (b) Turi parametrų, kuriuos reikia įvertinti/mokyti iš duomenų**
- (c)\* Parametrai kiekvieną kartą apskaičiuojami vykdymo metu
- (d) Neturi mokomų parametrų

2. Ar mokymosi žingsnį (angl. *learning rate*) galima keisti parametrų vertinimo (angl. *training*) metu (pvz. padidinti)?

- (a) žingsnis pasirenkamas pradžioje ir yra pastovus
- (b) žingsnis skaičiuojamas iš duomenų ir gali būti padidintas ir sumažintas
- (c) žingsnį galima keisti, galima ir didinti, ir mažinti**
- (d) žingsnį galima keisti, bet galima tik sumažinti

3. Naudojant kryžminės entropijos nuostolių funkciją, kartais gražinamos modelio reikšmės pateikiamos kaip „*logit*“ išraiška. Pvz.: Kryžminė entropija (`from_logits = True`) Ką tai reiškia?

- (a) Nuostolių funkcija tikisi, kad prognozės bus bet kokie realūs skaičiai**
- (b) Nuostolių funkcija tikisi, kad prognozės bus tikimybės
- (c) Nuostolių funkcija tikisi, kad prognozės bus tikimybės, kurių suma vienetas
- (d) Nuostolių funkcija tikisi, kad prognozės bus sveikieji skaičiai, atitinkantys prognozių klases

4. Ką daryti, jei reikia išspręsti daugiatislį uždavinį (pvz: regresijos ir klasifikavimo), o tik dalis duomenų pažymėta?

- (a) Jei žymenų duomenų skaičius nesutampa, sukurti modelio nepavyks
- (b) Atrenkame stebėjimus, kurie turi visas žymes, ir sukuriame kelių tikslų užduotį su vienu modeliu
- (c) Naudojame visus duomenis ir sukuriame kelių tikslų užduotį su vienu modeliu, užmaskuodami trūkstamas žymes**
- (d) Kuriamo du nepriklausomus modelius

5. Kam žetonų dekoderis (angl. *token decoder*) naudojamas klausimų/atsakymų kalbos modelyje?

- (a) Žetonų dekoderis nenaudojamas**
- (b) Modelis numato atsakymo pradžią ir pabaigą – dekoderis pasirinktus žetonus paverčia tekstu
- (c) Modelis numato teisingo atsakymo žetoną – žetonų dekoderis paverčia jį tekstu
- (d) Modelis numato teisingus žetonų atsakymus – žetonų dekoderis paverčia juos tekstu

6. Kokio tipo užduotys yra klausimų-atsakymų modeliuose, naudojant kontekstinę informaciją?

- (a) Regresijos – numatoma atsakymo vieta
- (b) Atsakymo užduotis gaunama nuspėjus teisingą atsakymą teksto forma
- (c) Du vieno žymens klasifikavimo modeliai – prognozuojama atsakymo pradžia ir pabaiga**
- (d) Daugelio klasių klasifikacija, kai modelis prognozuoja atsakymą

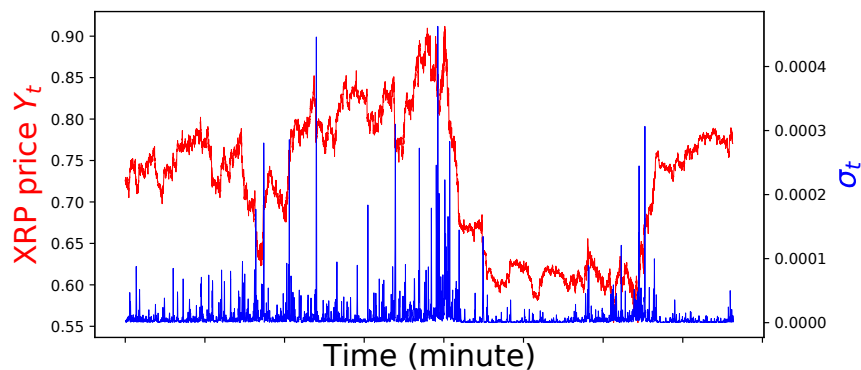
7. Kuri aktyvacijos funkcija nėra apibrėžta iš viršaus ir apačios?

- (a) Dalinai tiesinis vienetas (*ReLU*)
- (b) Mastelio padidintas eksponentinis tiesinis vienetas (*SELU*)
- (c) *Softplus*
- (d) Parametrinis tiesinis vienetas (*Parametric ReLU*)**

8. Kuris iš žemiau pateiktų metodų naudojamas norint rasti „patikimą“ prognozuojamą stačiakampį iš visų įvertintų prognozių?
- Mes paimame didžiausią stačiakampį
  - Imame stačiakampių sąjungą
  - (c) Paimkite stačiakampį, gautą ne maksimalaus slopinimo algoritmu**
  - Imame stačiakampių sankirtą
9. Sukuriate neuroninį tinklą. Norite taikyti rinkinio normalizavimo transformaciją. Taikote *Adam* optimizavimo priemonę, kryžminę entropiją kaip tikslo funkciją, ir rinkinį su dydžiu *batch size* = 2, mokymo žingsnis  $\alpha = 0.001$ . Modelio transformacijos neveikia, todėl modelis nesimoko. Kodėl?
- Adam* neturėtų būti naudojamas
  - Kryžminės entropijos nuostolių funkcija neturėtų būti naudojama
  - (c) Problema – toks mažas rinkinio dydžio naudojimas**
  - Transformacijos nėra problema
10. Turite daugelio klasių daugelio žymenų klasifikavimo problemą. Kokią duomenų transformaciją naudosite paskutiniame sluoksnyje?
- ReLU*
  - (b) Sigmoidas**
  - Softmax*
  - Linear*

## 2 Praktinė/modelio kūrimo užduotis (1 b.)

Tarkime sprendžiame daugiatakslį uždavinį, prognozuoti kainos volatilumą (kainos išsibarstymą), pačią kainą ir klasifikuoti ar kaina didės ar ne.



- Kiek išėjimų modelyje bus?  
Modelyje bus 3 išėjimai – 1 volatilumui, 1 kainai, 1 didėjimo/mažėjimo klasifikatoriui.
- Parašykite įvesties ir išvesties duomenų tenzorius dimensijas.  
 $T$  – praeities informacijos žingsnių skaičius (parenkamas rankiniu būdu). Modelio įvestis susidės iš  $T$  porų iš kainos ir jos volatilumo, t.y., įvesties dimensija bus  $T \times 2$ . Išvesties dimensija bus 3 (po 1 kiekvienam uždaviniui).
- Suformuluokite, kokia problema yra kiekvienai iš trijų užduočių nagrinėjama:
  - Prognozuoti kainai – regresijos uždavinys;
  - Prognozuoti išsibarstymui – regresijos uždavinys;
  - Prognozuoti pirkimui – klasifikavimo uždavinys.

4. Pasiūlykite ir pateikite, kokias nuostolių funkcijas siūlytumėte naudoti:

- (a) Prognozuoti kainai – L1 arba L2 nuostolių funkciją;
- (b) Prognozuoti išsibarstymui – L1 arba L2;
- (c) Prognozuoti pirkimui – binarinės kryžminės entropijos (angl. *Binary Cross-Entropy*).

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_i^N |y_i - \hat{y}_i|, \quad \mathcal{L}_{L2} = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2, \quad \mathcal{L}_{BCE} = -\frac{1}{N} \sum_i^N y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)$$

5. Pasiūlykite galutinę nuostolių funkciją

$$\mathcal{L} = \alpha \mathcal{L}_{L2, \text{ kainai}} + \beta \mathcal{L}_{L2, \text{ volatilumui}} + \gamma \mathcal{L}_{BCE, \text{ klasifikatoriui}},$$

čia  $\alpha, \beta, \gamma$  – hiperparametrai, parenkami rankiniu būdu.

6. Kokios galimos problemos gali kilti numatant kainų išsibarstymą? Kaip tai išspręsti?

7. Kokį galimą modelį pasiūlytumėte tokiai užduočiai?

Pasiūlyčiau laiko eilučių prognozavimo modelį, pvz. transformerį, *RNN*, ir pan.

8. Aprašykite, kaip sukurtumėte duomenų rinkinio padalijimus parametru vertinimui ir testavimui?

Visų pirma susiformatuočiau duomenų rinkinį pagal 2 užd. atsakymo formatą; tada, jei turime daug sekų, imčiau 70% sekų mokymui (parametru įvertinimui iš gradientų), 15% validacijai (modelio įsiverstinimui ir hiperparametru parinkimui), ir 15% testavimui po mokymo. Jei turime tik vieną ilgą seką, šią seką pagal tokias pat proporcijas padalinčiau.

9. Tarkime, kad pradėsite mokymą ir pastebėjote, kad pirkimo indikatorius daro daug klaidingų neigiamų (angl. *false negative*) klaidų. Ką tai reiškia praktiškai?

*False negative* reiškia, kad modelis spėja klasę nepirkti, tačiau *ground truth* klasė yra pirkti. Kadangi tik pradėjome mokymą, tai natūralu, jog dauguma spėjimų bus netikslūs. Jei modelis po mokymo vis dar daro daug klaidingų neigiamų sprendimų, praktikoje tai gali reikšti, kad mūsų modelis neaptinka potencialių kainos šuolių ir (jei turime automatinį sprendimų priėmimo variklį) priimame prastus sprendimus (t.y., prarandame galimybę pasipelnyti).

10. Gavote prognozę apie išsibarstymą  $\hat{\sigma}^2 = 1.3$ . Kaip interpretuoti rezultatus, kokias išvadas galime padaryti?

Išsibarstymo prognozė  $\hat{\sigma}^2 = 1.3$  reikštų, kad kaina gali vidutiniškai būti nutolusi per  $\sqrt{1.3}$  nuo paskutinės stebėtos kainos. Jei ši reikšmė didelė lyginant su pačios kainos reikšme, galima interpretuoti, kad modelis nėra itin užtikrintas, kainos prognozė gali būti netikslė, yra daug neapibrėžtumo. Jei reikšmė maža – modelis užtikrintas, galime labiau pasitikėti kainos prognoze.

### 3 Teorinė užduotis (1 b.)

1. Tarkime, kad sprendžiate klasifikavimo užduotį. Turite modelio prognozes  $\hat{y} = f(X|\hat{\theta})$  ir žymėtus duomenis  $y$ :

$$\hat{y} = [2.1 \quad 2.2 \quad 2.7 \quad -1.3 \quad -2.4 \quad 0.3] \quad y = [0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$$

Skaičiuokite klasifikavimo lentelę vieno žymens ir daugelio žymenų klasifikacijos atveju. Jei reikia slenksčio, naudokite  $T = 0.2$ .

Tarkime, kad duotos ne galutinės modelio reikšmės, tačiau reikšmės prieš galutinę aktyvacijos funkciją. Klasifikavimo lentelės:

$\hat{y} =$	2.1	2.2	2.7	-1.3	-2.4	0.3
Kl. vieno ž. a.	0	0	1	0	0	0
$\text{sigmoid}(\hat{y}) =$	0.89	0.9	0.94	0.21	0.08	0.57
Kl. daugelio ž. a.	1	1	1	1	0	1

*Kl. vieno ž. a.* lentelėje išgauname išrinkę maksimalią vektoriaus  $\hat{y}$  reikšmę. *Kl. daugelio ž. a.* gauname pritaikę *sigmoid* funkciją ir slenkstinę reikšmę.

2. Selu aktyvacijos funkcija apibrėžiama kaip:

$$\text{selu}(x) = \lambda \begin{cases} x & \text{jei } x > 0 \\ \alpha e^x - \alpha & \text{jei } x \leq 0 \end{cases}$$

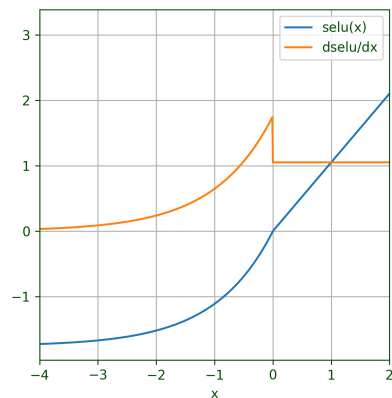
kur  $\lambda = 1.0507009873554804934193349852946$  ir  $\alpha = 1.6732632423543772848170429916717$ .

(a) Apskaičiuokite Selu aktyvacijos funkcijos išvestinę.

$$\begin{aligned} \frac{\partial \text{selu}}{\partial x} &= \lambda \begin{cases} \frac{\partial}{\partial x} x & \text{jei } x > 0 \\ \frac{\partial}{\partial x} (\alpha e^x - \alpha) & \text{jei } x \leq 0 \end{cases} \\ &= \lambda \begin{cases} 1 & \text{jei } x > 0 \\ \alpha e^x & \text{jei } x \leq 0 \end{cases} \end{aligned}$$

(b) Nubrėžkite Selu funkciją ir jos išvestinę.

Grafikas:



(c) Kokia yra funkcijos viršutinė ir apatinė riba?

Viršutinė riba  $+\infty$

Apatinė riba  $-\lambda\alpha = -1.758$