

MA'LUMOTLARNI INTELLEKTUAL TAHLIL QILISHDA CART ALGORITMIDAN FOYDALANISH

Sattarov M.A., Mardanov A.K., Xaniyev N.

Muhammad al-Xorazmiy nomidagi TATU Samarqand filiali

ANNOTATSIYA

CART (Classification And Regression Tree) - bu tasniflash va regressiya vazifalari uchun ishlatiladigan yechimlar daraxti algoritmi. Bu ko'rinmaydigan ma'lumotlarni bashorat qilish uchun belgilangan ma'lumotlardan o'rganadigan nazorat ostida o'rganish algoritmidir. CART tugunlar va shoxlardan iborat daraxtga o'xshash tuzilmani quradi, bu erda tugunlar turli xil qaror nuqtalarini ifodalaydi va filiallar bu qarorlarning mumkin bo'lgan natijalarini ifodalaydi.

Sinflashtirish va regressiya daraxtlari (CART) algoritmi sinflashtirish va regressiya vazifalari uchun ishlatiladigan yechimlar daraxtini o'rganishning mashhur usullaridan biri hisoblanadi. Ishda sinflashtirish uchun yechimlar daraxtini yaratishda CART algoritmidan foydalanishning misolini keltirilgan.

Kalit so'zlar: Data Mining, yechimlar daraxti, CART, classification, decision tree.

Masalaning qo'yilishi. Ushbu maqolada, biz ob-havo sharoitlari va odamlar yugurishga borishi yoki bormasligi haqida ma'lumotlar to'plamidan foydalanamiz. Bizning vazifamiz – ob-havo sharoitidan kelib chiqib inson yugurishga borishini bashorat qilish.

1-jadvalda bizning misolimiz uchun kichik ma'lumotlar to'plami keltirilgan.

1-jadval. Ma'lumotlar to'plami

| Kun | Harorat | Namlik | Shamol | Yugurishga boradimi? |
|-----|---------|--------|--------|----------------------|
| 1 | Issiq | Yuqori | Yo'q | Yo'q |
| 2 | Issiq | Yuqori | Bor | Yo'q |
| 3 | Iliq | Yuqori | Yo'q | Ha |
| 4 | Sovuq | Normal | Yo'q | Ha |
| 5 | Sovuq | Normal | Bor | Yo'q |
| 6 | Iliq | Normal | Bor | Ha |
| 7 | Issiq | Normal | Yo'q | Ha |
| 8 | Iliq | Yuqori | Bor | Yo'q |
| 9 | Sovuq | Normal | Yo'q | Ha |
| 10 | Issiq | Normal | Bor | Yo'q |

1-qadam: Gini indeksini hisoblash. Ishni har bir belgi uchun Gini indeksini hisoblashdan boshlaymiz. Gini indeksi – bu ma’lumotlar ajralish tozaligini sonli aniqlash amaliyoti. Gining past qiymati eng yaxshi ajralishni ko’rsatadi.

Keling, har bir xususiyat uchun Gini indeksini hisoblaymiz:

1. Harorat:

• Issiq: 2 Yo‘q, 2 Ha ($Gini = 1 - (2/4)^2 - (2/4)^2 = 0.5$)

• Iliq: 2 Yo‘q, 2 Ha ($Gini = 0.5$)

• Issiq: 1 Yo‘q, 2 Ha ($Gini = 1 - (1/3)^2 - (2/3)^2 \approx 0.44$)

Harorat uchun mutlaq Gini indeksi = $(4/10) * 0.5 + (4/10) * 0.5 + (3/10) * 0.44 = 0.48$

2. Namlik:

• Yuqori: 4 Yo‘q, 1 Ha ($Gini = 1 - (4/5)^2 - (1/5)^2 = 0.32$)

• Normal: 1 Yo‘q, 4 Ha ($Gini = 1 - (1/5)^2 - (4/5)^2 = 0.32$)

Namlik uchun mutlaq Gini indeksi = $(5/10) * 0.32 + (5/10) * 0.32 = 0.32$

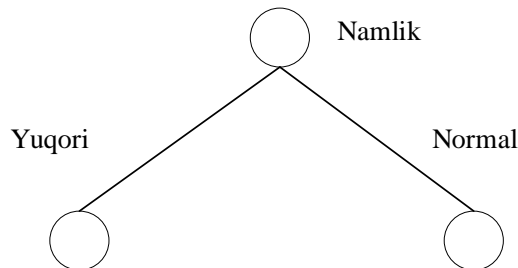
3. Shamolli:

• Yo‘q: 2 Yo‘q, 4 Ha ($Gini = 1 - (2/6)^2 - (4/6)^2 \approx 0.44$)

• Bor: 3 Yo‘q, 1 Ha ($Gini = 1 - (3/4)^2 - (1/4)^2 = 0.375$)

Shamol uchun mutlaq Gini indeksi = $(6/10) * 0.44 + (4/10) * 0.375 = 0.412$

2-qadam: Eng yaxshi ajralishni aniqlash. Gini indeksi hisob-kitoblaridan biz "Namlik" eng past Gini indeksiga ega ekanligini ko’rishimiz mumkin (0,32). Shunday qilib, biz ushbu xususiyat asosida ma’lumotlarni ajratamiz.



3-qadam. Har bir qism to‘plam uchun takrorlaymiz. Keyinchalik, har bir qism to‘plamdagi qolgan xususiyatlar uchun Gini indeksini hisoblab chiqamiz va eng yaxshi bo‘linishni tanlaymiz.

1-qism to‘plam (Yuqori namlik)

| Kun | Harorat | Shamol | Yugurishga boradimi? |
|-----|---------|--------|----------------------|
| 1 | Issiq | Yo‘q | Yo‘q |
| 2 | Issiq | Bor | Yo‘q |
| 3 | Iliq | Yo‘q | Ha |
| 7 | Issiq | Yo‘q | Ha |
| 8 | Iliq | Bor | Yo‘q |

2-qism to'plam (Normal namlik)

| Kun | Harorat | Shamol | Yugurishga boradimi? |
|-----|---------|--------|----------------------|
| 4 | Issiq | Yo'q | ha |
| 5 | Cool | Bor | Yo'q |
| 6 | Iliq | Bor | Ha |
| 10 | Issiq | Bor | Ha |

1-qism to'plam (yuqori namlik)

1. Harorat:

- Issiq: 2 Yo'q, 1 Ha (Gini = $1 - (2/3)^2 - (1/3)^2 \approx 0.44$)
- Iliq: 1 Yo'q, 1 Ha (Gini = 0.5)

Harorat uchun mutlaq Gini indeksi = $(3/5) * 0.44 + (2/5) * 0.5 = 0.464$

2. Shamol:

- Yo'q: 2 Yo'q, 2 Ha (Gini = 0.5)
- Bor: 2 Yo'q (Gini = 0)

Shamol uchun mutlaq Gini indeksi = $(4/5) * 0.5 + (1/5) * 0 = 0.4$

1-qism to'plam (Yuqori namlik) uchun eng yaxshi bo'linish "Shamolli" bo'lib, Jini indeksi 0,4.

2-qism to'plam (Normal namlik):

1. Harorat:

- Sovuq: 1 Yo'q, 2 Ha (Gini ≈ 0.44)
- Iliq: 1 Ha (Gini = 0)
- Issiq: 1 Ha (Gini = 0)

Harorat uchun mutlaq Gini indeksi = $(3/5) * 0.44 + (1/5) * 0 + (1/5) * 0 = 0.264$

2. Shamol:

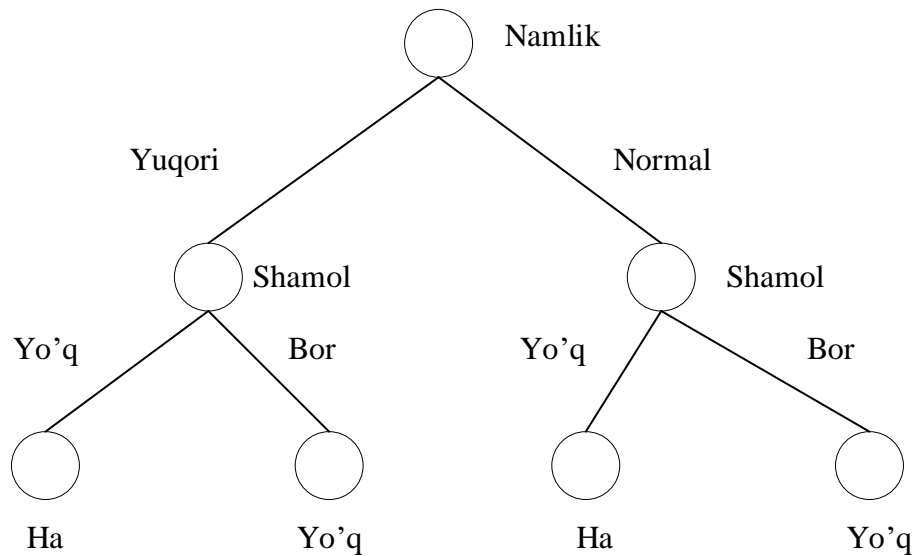
- Yo'q: 3 Ha (Gini = 0)
- Bor: 1 Yo'q, 2 Ha (Gini ≈ 0.44)

Shamol uchun mutlaq Gini indeksi = $(3/5) * 0 + (2/5) * 0.44 = 0.176$

2-qism to'plam uchun eng yaxshi bo'linish (normal namlik) "Shamolli" bo'lib, Gini indeksi 0,176.

4-qadam. To'xtash mezon. Biz to'xtash mezoniga yetgunimizcha bo'linishni davom ettiramiz. Bunday holda, agar bo'linadigan boshqa xususiyatlar qolmaganida yoki barg tugunida faqat bitta sinf mavjud bo'lsa to'xtaymiz.

Bizning yakuniy yechim daraxtimiz quyidagicha ko'rinadi:



Endi biz ushbu yechim daraxtidan yangi ma'lumotlar nuqtalari uchun bashorat qilish uchun foydalanishimiz mumkin. Misol uchun, agar bizda normal namlik va shamol bo'lmagan kun bo'lsa, yechim daraxti odamning yugurish uchun ketishini taxmin qiladi.

FOYDALANILGAN ADABIYOTLAR RO'YXATI: (REFERENCES)

1. **James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013).** An Introduction to Statistical Learning: with Applications in R. Springer Science & Business Media. Chapter 8 of this book provides an accessible introduction to decision trees, including the CART algorithm, with examples in the R programming language.
2. **Lantz, B. (2013).** Machine Learning with R. Packt Publishing Ltd. Chapter 6 of this book provides a practical introduction to the CART algorithm, with examples in the R programming language.
3. **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011).** Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. This article introduces the scikit-learn library, which includes an implementation of the CART algorithm, with examples in Python.
4. **Raschka, S., & Mirjalili, V. (2017).** Python Machine Learning. Packt Publishing Ltd. Chapter 6 of this book provides an introduction to decision tree learning, including the CART algorithm, with examples in Python.
5. **Brownlee, J. (2020).** How to Implement the CART Algorithm from Scratch in Python. *Machine Learning Mastery*. This online tutorial provides a step-by-step guide for implementing the CART algorithm in Python, with a practical example.