

K - YAQIN QO'SHNI ALGORITMI

Onarkulov Maksadjon Karimberdiyevich

Amaliy matematika va informatika kafedrası dotsenti

maxmaqsad@gmail.com

Yusupov Mirsaid Abdulaziz o'g'li

Farg'ona davlat universiteti o'qituvchisi

mirsaidbeky@gmail.com

Ismoilova Mohinurbonu Is'hoqjon qizi

Farg'ona davlat universiteti 3-kurs talabasi

mohinur7479@gmail.com

Annotatsiya: *K - yaqin qo'shni (k -NN) algoritmi sun'iy intellekt va mashina o'rganish sohalarida keng tarqalgan va oddiy, ammo samarali tasniflash va regressiya metodlaridan biridir. Ushbu tezisdá, k - yaqin qo'shni algoritmning nazariy asoslari, amaliy qo'llanilishi, kuchli tomonlari va cheklovlari, shuningdek, takomillashtirish va optimallashtirish usullari ko'rib chiqiladi.*

Kalit so'zlar: *K-yaqin qo'shnilar algoritmi, mashinani o'rganish, tasniflash, regressiya, masofa o'lchovlari, modelni optimallashtirish.*

Аннотация: *Алгоритм k-Ближнего соседства (K-NN) широко используется в областях искусственного интеллекта и машинного обучения и является одним из простых, но эффективных методов классификации и регрессии. В данной диссертации будут рассмотрены теоретические основы, практическое применение, сильные стороны и ограничения алгоритма k - Ближнего соседства, а также методы улучшения и оптимизации.*

Ключевые слова: *K-алгоритм ближайших соседей, машинное обучение, классификация, регрессия, измерения расстояний, оптимизация модели.*

Annotation: *The K-close neighbor (k-NN) algorithm is common in the fields of artificial intelligence and machine learning, and is one of the simple but effective classification and regression techniques. In this thesis, the theoretical foundations, practical applications, strengths and limitations of the K - close neighbor algorithm are considered, as well as methods of refinement and optimization.*

Keywords: *K-near neighbors algorithm, machine learning, classification, regression, distance measurements, model optimization.*

Kirish

K-NN ning asosiy g'oyasi shundan iboratki, yangi ob'ekt uchun sinf yoki qiymatni bashorat qilish uchun o'quv ma'lumotlar to'plamida uning eng yaqin qo'shnilariga qarash va shu qo'shnilar asosida xulosa chiqarish kerak. K-NN algoritmi an'anaviy usulda o'qitishni talab qilmaydi, chunki unda aniq modelni yaratish bosqichi yo'q.

Buning o'rniga, barcha ma'lumotlar o'quv majmuasida saqlanadi va bashoratlar to'g'ridan-to'g'ri ushbu to'plamdan amalga oshiriladi.

Asosiy qisim

K-yaqin qo'shni algoritmi - parametrik bo'lmagan o'rganish usuli bo'lib, tasniflash va regressiya masalalarini hal qilish uchun eng yaqin qo'shni tamoyilidan foydalanadi. Oddiyli va intuitiv kontseptsiyasiga qaramay, k-NN juda kuchli vosita bo'lib, naqshni aniqlash, tavsiya qiluvchi tizimlar va kasalliklar diagnostikasi kabi turli sohalarda keng qo'llaniladi.

K-yaqin qo'shni algoritmi, asosan, qo'shni nuqtalarning masofasiga asoslanadi. Algoritmning ishlash prinsipi quyidagicha:

Ma'lumotlar tayyorlash: k-NN ishlatilishi uchun ma'lumotlar o'rganilgan va klassifikatsiyalangan bo'lishi kerak.

Masofa o'lchovlari: k-NN masofani o'lchashda, odatda, Evklid masofasi, Manhattan masofasi yoki boshqa o'lchovlar ishlatiladi.

K parametrini tanlash: k - yaqin qo'shnilar soni parametrini tanlash, algoritmning aniqligiga ta'sir qiladi. K ning kichik qiymatlari ortiqcha moslashuv, katta qiymatlari esa kam moslashuvga olib kelishi mumkin.

Tasniflash yoki regressiya: k-NN o'rganilgan ma'lumotlar asosida yangi nuqtaning sinfini yoki qiymatini aniqlaydi.

K-NN algoritmini takomillashtirish usullari

Mahsulotlar o'lchovining optimallashtirilishi: O'lchovlarning ahamiyatini optimallashtirish uchun xususiyatlarni tanlash va o'lchovlarni kamaytirish usullari ishlatiladi. Masalan, PCA (Principal Component Analysis) yordamida o'lchovlar sonini kamaytirish k-NNning samaradorligini oshirishi imkonini beradi.

K qiymatini tanlash: k-NN algoritmida k parametrini tanlash juda muhimdir. Kning juda kichik qiymati modelni ortiqcha moslashtirishi mumkin, katta k esa uning umumlashganligini kamaytirishi mumkin. Bu qiymatni tanlash uchun kross-validatsiya (cross-validation) usuli ishlatiladi.

Masofa o'lchovlarini o'zgartirish: ba'zida barcha masofa o'lchovlari ham teng bo'lmaydi. Shunday hollarda, masofa o'lchovlarini dinamik ravishda o'zgartirish (masalan, turli xususiyatlar uchun turli og'rliklarni qo'llash) samaraliroq bo'lishi mumkin.

K-NN algoritmining afzalliklari

Oddiy va tushunarli: k-NN juda oddiy algoritm bo'lib, uning ishlash prinsipi tushunilishi oson.

K-NN parametrlarni oldindan o'rganish yoki modelni o'qitish kerak emas, chunki barcha hisoblashlar test nuqtasi kiritilganida amalga oshiriladi.

Lokal qarorlar: k-NN algoritmi har bir test nuqtasini o'zining yaqin qo'shni nuqtalari asosida tasniflaydi, bu esa uni murakkab va o'zgaruvchan muhitlarga moslashtiradi.

K-NN algoritmining kamchiliklari

Hisoblash xarajatlari: k-NN algoritmi katta ma'lumotlar bazalarida juda sekin ishlashi mumkin, chunki har bir test nuqtasi uchun barcha trening nuqtalarini ko'rib chiqish zarur.

Yuqori o'lchovli ma'lumotlar: k-NN yuqori o'lchovli ma'lumotlar bilan ishlaganda samaradorligi pasayishi mumkin. O'lchovlar soni oshgan sari masofani aniqlash qiyinlashadi.

Normalizatsiya: k-NN masofaga asoslangan algoritm bo'lgani uchun, ma'lumotlar to'plamidagi o'zgaruvchilarni normalizatsiya qilish zarur, aks holda ba'zi o'zgaruvchilar boshqa o'zgaruvchilardan kattaroq ahamiyatga ega bo'lishi mumkin.

K-NN algoritmining amaliy tadbiqlari

K-NN algoritmi ko'plab amaliy sohalarida muvaffaqiyatli qo'llanilmoqda:

Tasniflash: k-NN tasniflashda ma'lumotlarni sinflarga ajratish uchun ishlatiladi. Misol uchun, elektron pochta xabarlarini spam yoki normal deb tasniflash.

Regressiya: k-NN regressiyada ma'lumotlarning ketma-ketligini prognoz qilish uchun ishlatiladi. Misol uchun, uy narxlarini yoki aktsiya narxlarini prognozlash.

Mahsulot tavsiyalar: k-NN algoritmi tavsiyalar yaratish uchun ham ishlatiladi, masalan, foydalanuvchilarga shunga o'xshash mahsulotlar yoki xizmatlar taklif qilish.

Sog'liqni saqlash va biotibbiyot: Tibbiy diagnostika tizimlarida k-NN diagnostik belgilarni tasniflashda samarali ishlatiladi.

Xulosa

K-yaqin qo'shni algoritmi, o'zining oddiyligi, moslashuvchanligi va samaradorligi bilan ko'plab real dunyo muammolarini hal qilishda qo'llaniladi. Biroq, uning samaradorligi ma'lumotlar to'plami va masofa o'lchovlariga bog'liq. Shu bois, k-NN algoritmining afzalliklarini to'liq amalga oshirish uchun mos parametrlarga va ma'lumotlar to'plamini optimal tarzda tanlashga e'tibor berish zarur.

FOYDALANILGAN ADABIYOTLAR:

1. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. "IEEE Transactions on Information Theory", 13(1), 21–27.
2. Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. "The American Statistician"; 46(3), 175–185.
3. Han, J., Kamber, M., & Pei, J. (2011). "Data Mining: Concepts and Techniques" (3rd ed.). Elsevier.
4. Bishop, C. M. (2006). "Pattern Recognition and Machine Learning". Springer.
5. Zhang, X., & Zhao, L. (2015). A review of the K-nearest neighbor algorithm in data mining. "International Journal of Computer Science and Technology", 6(2), 61–64.