

6 – dars.Report

Encoding - bizda bor ma'lumotlarni bir trdan ikkinchi turga o'tkazish.

Ya'ni ma'lumotlarni raqamli, komputer tushunadigan tilga o'tkazish.

Encodingni turlari

1. One - Hot encoding

One - Hot Encoding : Kategoriya ma'lumotlarni raqam formatga aylantirish. Ular asosan kategoriyali (nominal) qiymatlarni model uchun tushunarli bo'lgan sonli ko'rinishga keltiradi.

Har bir qator uchun binarlashgan (0 va 1) vektor yaratadi.

Afzalliklari:

1. Tartibsiz kategoriyalar uchun ideal:

- Har bir kategoriya mustaqil vektor sifatida ifodalanadi.

2. Raqamlarni tartib sifatida tushunishdan saqlaydi.

Kamchiliklari:

1. Matritsa hajmi oshadi:

- Har bir yangi kategoriya uchun bitta yangi ustun yaratiladi.
- Agar kategoriyalar juda ko'p bo'lsa (masalan, 1000), matritsa hajmi katta bo'ladi.

2. Sparslik muammosi:

- Ko'p "0" qiymatlar bo'ladi, bu hisoblashni sekinlashtiradi.

2. Label encoding

Label Encoding — **kategoriya ma'lumotlarini raqamlarga aylantirish usuli**. Har bir kategoriya qiymatiga unikal raqam beriladi.

Misol:

mathematica

Copy code

Mevalar: ["Olma", "Banan", "Gilos"]

Label Encoding: [0, 1, 2]

Farqi :



Nominal va Ordinal ma'lumotlar farqi

Xususiyat	Nominal ma'lumotlar	Ordinal ma'lumotlar
Tartib	Yo'q	Bor
Masofa	Belgilanmagan	Belgilanmagan
Misollar	Ranglar, shaharlar	Baholar, qoniqish darajalari
Encoding usuli	One-Hot Encoding	Label Encoding

Nominal va Ordinal ma'lumotlarni qachon qanday ishlatish kerak?

1. Nominal ma'lumotlar:

- Ranglar, shaharlar kabi **tartibsiz** kategoriyalar uchun ishlatiladi.
- **One-Hot Encoding** ideal, chunki bu usul kategoriyalarni mustaqil ifodalaydi.

2. Ordinal ma'lumotlar:

- Baholar, kiyim o'lchami kabi **tartibli** kategoriyalar uchun ishlatiladi.
- **Label Encoding** yaxshi, chunki tartibni saqlab qoladi.

Bitta datasetda 4 ta ustun (ob'ekt) mavjud bo'lsa, Label Encoding yoki One-Hot Encodingdan qaysi birini tanlash ustun ma'lumotlarining turiga va ularning xususiyatlariga bog'liq. Har bir ustunni alohida baholash kerak.

Misol dataset:

java

Copy code

Ustunlar:

1. Rang (Nominal) → ["Qizil", "Ko'k", "Yashil"]
2. Kiyim o'lchami (Ordinal) → ["Kichik", "O'rta", "Katta"]
3. Narx (Numerik) → [100, 200, 300]
4. Material turi (Nominal) → ["Paxta", "Ipak", "Jun"]

Ma'lumotlarning turlari:

- **Rang:** Nominal → **One-Hot Encoding** ishlatiladi.
- **Kiyim o'lchami:** Ordinal → **Label Encoding** ishlatiladi.
- **Narx:** Numerik → Kodlash kerak emas, faqat normalizatsiya yoki standartizatsiya kerak.
- **Material turi:** Nominal → **One-Hot Encoding** ishlatiladi.

[Raqamli ma'lumotlarda ishlatiladigan Encoding usullari](#)