

Gradient Boosting Modellari

- XGboost
- LightGBM
- CatBoost

Gradient Boosting - bu mashinani o'rganish texnikasi bo'lib, u bir necha zaif modellar (masalan, decision tree) laridan foydalanib kuchli modellar hosil qilishga asoslanagan. U xatoliklarni ketma ket kamaytiradi.

Asosiy Jarayonlar

1. Boshlang'ich Model:
2. Xatolikni aniqlash
3. Gradienni hisoblash
4. Yangi Model qo'shish
5. Takrorlash
6. Kombinatsiya

Kamchiliklari :

- Hisoblashlar ko'p vaqt talab qilishi mumkin
- Parametrlarni sozlash murakkab bo'lishi mumkin

Diagramma

1. Model 1 → Xato
2. Model 2 → Xatoni kamaytiradi
3. Model 3 → Yana xatoni kamaytiradi
4. Yakuniy natija: **Yaxshilangan model**

Quyida **XGBoost**, **LightGBM**, va **CatBoost** kutubxonalarini solishtirish uchun to'liq **jadval** keltirilgan:

Scenario	XGBoost	LightGBM	CatBoost
Uy narxini aniqlash	✓	✓	✓
Katta datasetlarda bashorat	⚠ Biroz sekinroq	✓ Juda tez	✓ Tez
Kategorik ma'lumotlar ko'p	⚠ Oldindan kodlash kerak	⚠ Qo'lda kodlash yoki ko'rsatish	✓ Avtomatik ishlaydi
Resurslar cheklangan	⚠ Ko'proq xotira talab qiladi	✓ Kam xotira talab qiladi	✓ Kam xotira talab qiladi
Overfittingni boshqarish	✓ Kuchli boshqaruv	⚠ Parametr sozlash talab qilinadi	✓ Dinamik boshqaruv
Yo'qolgan qiymatlar (Missing values)	⚠ Qo'lda to'ldirish kerak	✓ Avtomatik ishlaydi	✓ Avtomatik ishlaydi
Scaling (Masshtablash)	⚠ Oldindan talab qilinishi mumkin	✓ Talab qilinmaydi	✓ Talab qilinmaydi
O'rta datasetlar bilan ishlash	✓ Juda yaxshi	✓ Juda yaxshi	✓ Juda yaxshi
Kichik datasetlarda ishlash	✓ Juda yaxshi	✓ Juda yaxshi	✓ Juda yaxshi
GPU qo'llab-quvvatlash	✓ Juda samarali	✓ Yaxshi	✓ Yaxshi
Kodlash soddaligi	⚠ Biroz ko'proq sozlash talab qilinadi	✓ Oson sozlanadi	✓ Juda oddiy va avtomatik
Tezlik	✓ Tez	✓ Juda tez	✓ Juda tez

Xulosa:

- XGBoost:

- Yuqori aniqlik va nazorat kerak bo'lganda yaxshi.
- Kichik va o'rta datasetlar uchun juda samarali.
- Qo'shimcha kodlash talab qilinishi mumkin.
- **LightGBM:**
 - Juda katta datasetlar va resurs cheklovlari bo'lganda ideal.
 - Kategorik ma'lumotlar uchun qo'lda ishlov berish talab qilinadi.
- **CatBoost:**
 - **Kategorik ma'lumotlar ko'p bo'lsa yoki preprocessingni minimallashtirish kerak bo'lsa, eng qulay tanlov.**
 - Har xil datasetlar bilan ishlashga mos.

Boosting Algoritmilarini Solishtirish

Algoritm	Afzalliklari	Kamchiliklari	Qachon ishlatiladi?
Gradient Boosting	Yuqori aniqlik, moslashuvchan	Sekin, ba'zan overfitting	Oddiy holatlar va yaxshi aniqlik talab qilinganda
XGBoost	Tez, GPU qo'llab-quvvatlash, overfitting boshqaruvi	Ba'zan resurs talab qiladi	O'rta datasetlar va yuqori aniqlik talab qilinsa
LightGBM	Juda tez, katta datasetlar uchun ideal	Qo'lda kodlash talab qilinadi	Juda katta datasetlar va resurs cheklangan holatlar
CatBoost	Kategorik ma'lumotlar bilan avtomatik ishlaydi	Ba'zan sekinroq bo'lishi mumkin	Kategorik ma'lumotlar ko'p bo'lganda
HistGradientBoosting	Tezlik, katta datasetlar uchun samarali	GPU yo'qligi	Katta datasetlar va xotira samaradorligi muhim bo'lganda

Algoritm	Afzalliklari	Kamchiliklari	Qachon ishlatiladi?
AdaBoost	Oddiy, tushunarli, kichik datasetlarda yaxshi ishlaydi	Overfittingga moyillik	Oddiy va kichik datasetlar uchun

Xulosa

1. **Boosting algoritmlari** — ketma-ket zaif modellarni birlashtirib, kuchli model yaratadi.
2. Har bir algoritmning o'z afzallik va kamchiliklari bor:
 - **Katta datasetlar** uchun: LightGBM yoki HistGradientBoosting.
 - **Kategorik ma'lumotlar ko'p** bo'lsa: CatBoost.
 - **Oddiy va kichik datasetlar** uchun: AdaBoost.
3. **XGBoost** eng mashhur va ko'p maqsadli, lekin ba'zan resurs talab qiladi.