

REGRESSION MODELLARIDA COST FUNCTIONLAR TAHLILI: MSE, MAE, HUBER LOSS TAQQOSLANISHI.

Xujayarov Ilyos

Toshkent axborot texnologiyalari universiteti

Samarqand filiali dotsenti;

i.khujayorov@tuit.uz

Maxmadiyov Faxriddin

makhmadiyovfakhriddin@gmail.com

+998882782006

<https://doi.org/10.5281/zenodo.17383311>

Annototsiya: Mashina o'rganish (Machine Learning) sohasida regression modellari sonli qiymatlarni bashorat qilish uchun keng qo'llaniladi. Masalan, uy narxini, havo haroratini yoki ishlab chiqarish hajmini oldindan aniqlashda regressiya modellari asosiy vositadir. Bunday modellarni o'qitish jarayonida cost function (yoki loss function) juda muhim rol o'ynaydi, chunki u modelning qanchalik "xato" bashorat qilayotganini o'lchaydi.

Kirish

Cost function yordamida modelning og'irliklari (parametrlari) optimization algoritmlari (masalan, Gradient Descent) orqali yangilanadi. Shunday qilib, cost function model o'rganish jarayonining markazida turadi. Turli cost functionlar modelni turlicha yo'naltiradi: ba'zilar outlierlarga sezgir, ba'zilar esa bardoshli (robust) bo'ladi. Shu sababli, regression modellar uchun to'g'ri cost function tanlash modelning natijasi, barqarorligi va umumlashish (generalization) qobiliyatiga bevosita ta'sir qiladi.

Ushbu maqolada regression modellar uchun eng ko'p qo'llaniladigan uchta cost function — Mean Squared Error (MSE), Mean Absolute Error (MAE), va Huber Loss— nazariy va amaliy jihatdan tahlil qilinadi hamda ularning o'zaro farqlari muhokama qilinadi.

2. Regressionda Cost Functionlar nazariyasi

2.1. Mean Squared Error (MSE)

Mean Squared Error (MSE) regressionda eng keng tarqalgan cost function hisoblanadi. U bashorat qilingan qiymat $\hat{y}_i$ va haqiqiy qiymat y_i o'rtasidagi kvadrat farqlarni o'rtacha hisoblaydi:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Bu funksiya **differentiable** va **analitik gradient**ga ega bo'lgani uchun Gradient Descent yoki hatto oddiy Linear Regressionda **yopiq yechim (closed-form solution)** olish imkonini beradi.

Afzalliklari:

Hisoblash sodda va tez.

Matematik jihatdan silliq va gradient mavjud.

Kichik xatoliklarni kuchli penalizatsiya qiladi, bu esa modelni aniq o'qitishga yordam beradi.

Kamchiliklari:

Outlierlarga juda sezgir: bitta katta xatolik natijani kuchli buzadi.

Data'da g'ayrioddiy nuqtalar bo'lsa, model noto'g'ri yo'nalishga ketadi.

2.2. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) — xatoliklarni kvadrat emas, balki mutlaq qiymat bo'yicha o'lchaydi:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i|$$

U MSE ga qaraganda **outlierlarga kamroq sezgir**, chunki kvadrat emas, mutlaq farqni oladi. Ammo MAE ning muhim kamchiligi — u **differentiable emas** nol nuqtada ($y_i - \tilde{y}_i$), bu gradient asosidagi optimizatsiyani biroz qiyinlashtiradi.

Afzalliklari:

- Outlierlarga bardoshli (robust).

Bashoratdagi katta xatoliklarni haddan tashqari og'irlashtirmaydi.

Kamchiliklari:

- Gradient silliq emas.
- O'qitish jarayoni MSE ga nisbatan sekinroq bo'lishi mumkin.

2.3. Huber Loss

Huber Loss – bu **MSE** va **MAE** orasidagi muvozanatni saqlovchi cost function. U kichik xatoliklarda MSE kabi, katta xatoliklarda esa MAE kabi ishlaydi. U quyidagicha aniqlanadi:

$$J(\theta) = \begin{cases} \frac{1}{2}(y_i - \tilde{y}_i)^2, & \text{agar } |y_i - \tilde{y}_i| \leq \delta \\ \delta \left(|y_i - \tilde{y}_i| - \frac{1}{2}\delta \right), & \text{aks holda} \end{cases}$$

Bu yerda δ — **threshold** qiymat bo'lib, u MSE dan MAE ga o'tish chegarasini belgilaydi.

Afzalliklari:

- Silliq gradientga ega.

Outlierlarga nisbatan bardoshli.

MSE va MAEning eng yaxshi tomonlarini birlashtiradi.

Kamchiliklari:

δ parametrni to'g'ri tanlash tajribaga bog'liq.

Hisoblash biroz murakkabroq.

3. Cost Functionlarning amaliy taqqoslanishi

Turli cost functionlar turli xususiyatlarga ega. Quyidagi jadvalda ularning asosiy parametrlar bo'yicha taqqoslanishi keltirilgan:

Cost Function	Outlierlarga sezgirlik	Gradient silliqligi	Hisoblash murakkabligi	Robustlik
MSE	Yuqori	Juda silliq	Past	Past
MAE	Past	Diskontinuuous	Past	Yuqori
Huber	O'rtacha	Silliq	O'rta	Yuqori

Bu cost functionlar kaggle platformasida "House Prices" datasetida tekshirilgan (<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>):

Regression modellarni muvaffaqiyatli o'qitish uchun **cost function tanlovi** hal qiluvchi ahamiyatga ega.

- **MSE** — data toza bo'lganda eng tez va samarali natija beradi.
- **MAE** — outlierlarga nisbatan mustahkamlikni oshiradi.
- **Huber Loss** — bu ikki yondashuv o'rtasida optimal balansni ta'minlaydi.

Demak, **cost functionni tanlash** data tabiatiga, outlierlar mavjudligiga va modelning o'qitilish usuliga qarab amalga oshirilishi kerak. Kelajakdagi tadqiqotlarda bu funksiyalarni **neural tarmoqlar, regularization** va **optimizerlar** bilan birgalikda tahlil qilish regression modellar barqarorligini yanada oshiradi.