

10-dars uchun Report

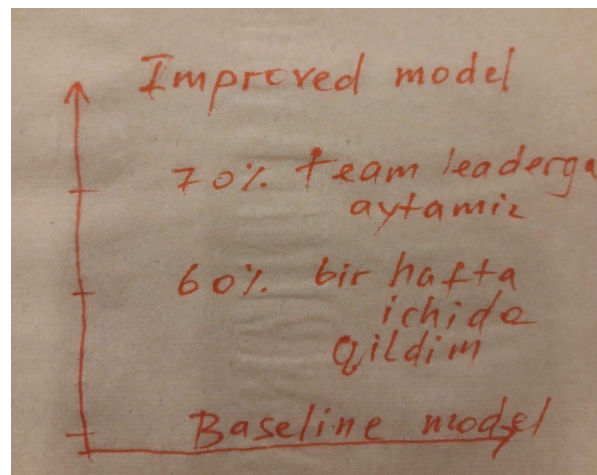
Avvalo takrorlash bo'ldi bunda biz: **Logistic Regression** ko'proq **binary classification** uchunligi haqida gaplashdik va **multiclasslar** uchun esa **Decision Tree** ishlatish yaxshiroqligi haqida ham gaplashdik.

Training jaroyoniga ham to'hdaldik bunda; **Preprocessing** qilingan datani biror algoritim yordamida kompyuterning miyyasiga quyish yoki o'rgatish hisoblanadi.

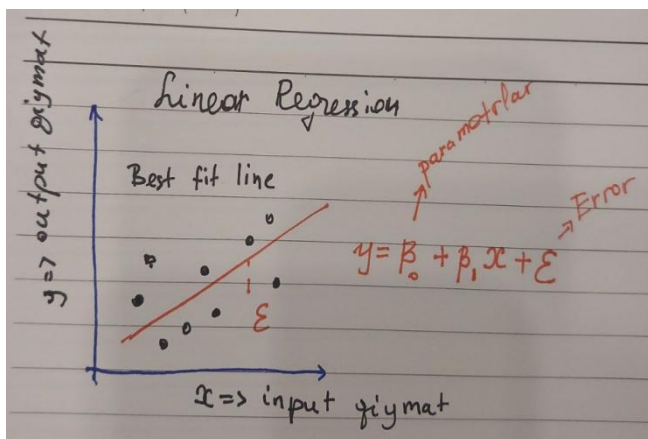
Kompyuter datalar yordamida o'rganadi ma'lumotlar oladi, **patternlarni** o'rganadi shunga qarab turib qaror qabul qiladi.

Linear Regressionga quyidagilar kiradi:

- Linear family,
- Baseline model
- Regression (continues)
 - * Players goals
 - * Net injury
- Linear equation ($y = kx + l$)



Linear Regressionning ko'rinishi bunday:



Masalan bunga misol o'laroq **osh** tayyorlashni olib unda; osh ---> u output, qolganlari ya'ni masalliqalar o'lchovi esa bu inputlardir.

Parametrlar input qiymatni muhimligini o'lchovchi o'lcham hisoblanadi.

Linear Regressionda demak; **x** --> **input**, **y** --> **output**, **beta** --> **parametr**, va **epsilon** --> **Error**

Linear Regression maqsadli yoki Linear Regression ishlatayotkanimizda natija zo'r chiqayotkan bo'lsa qachonki liniya chiqiz paydo bo'lsa (input va output uchun mos chiqiz topib berish). Linear Regression o'z navbatida: **Simple Linear Regression** va **Multiple Linear Regression** ga bo'linadi. Bunda featuralar ko'p, juda ko'plab input qiymat uchun parametr qo'shiladi.

Qo'chimcha qiladigan bo'lsak --> **Epsilon** bu **Error** hech qachon 0ga teng bo'lmaydi ($\neq 0$). Sababi, **Error** kamaytirishga harakat qilinadi lekin albatta **epsilon**imiz **0** ga teng bo'lmaydi. Buni ham hisobga olish kerak.

Va **Linear Regressionni pros** va **cons**lari mavjuddir.

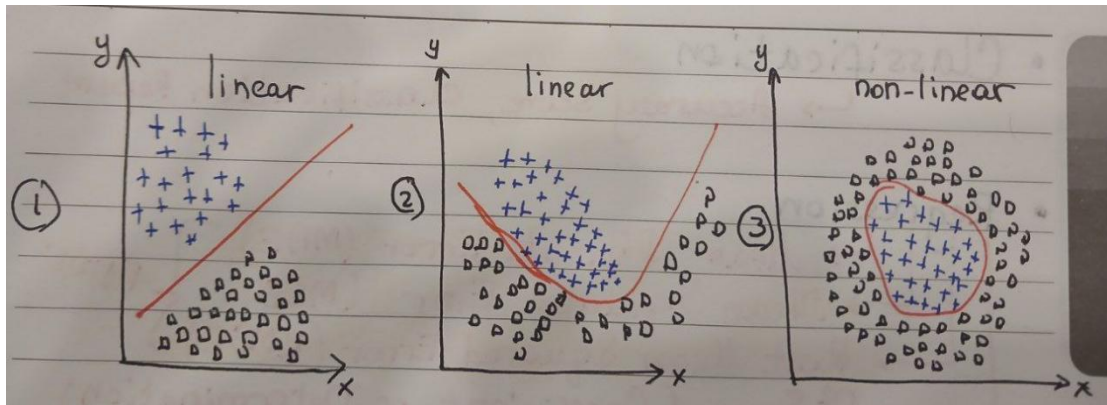
- **Pros:**

- * Tez va soddadir

- * Kichikroq datalar bilan ishlashi oson
- * Baseline model qurish uchun qulayligida
- * Kam chiqim (low cost) - oz material va harajat talab qiladi.

- Cons

- * Complex (non-linear) xolatlarni o'rgana olmaydi.
- * Yod elementlarda tez o'zgaruvchan (stabil emas)
- * Katta datasetlarda yaxshi ishlamaydi.



Yuqoridagi rasmda esa ko'rishimiz mumkin Linear Regression va Non-Linear Regression uchun misolni.

Evaluation (baholash) --> Generalization va Overfitting

Quyidagilarni baholaymiz: - Prediction qanchalik yaxshi? (raqam bilan o'chaydi masalan [0, 100] oraliqda), - Model ishinarlimi? (Biz qanchalik ishonishimiz mumkin?), - Generalization (Yaxshi umumlashtirdimi?)

Overfitting (bu yomon so'z ekan)- bu model 'train data'ni haddan tashqari yodlab olishdir. Masalan, talaba savolni tushunmay javoblarni yodlab olishi, ammo unga raqamlarni almashtirib berilsa javob topa olmasligi. **Generalization** (bu yaxshi so'z ekan) esa umumlashtirgan, mantiqan yondoshgani uchun buning javobi bu chiqadi deya aniq biladi talaba.

So'ngida esa **Classification** uchun biz --> - **Accuracy score** va **Classification Report**dan. - **Regression** uchun esa biz --> **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)** va **R² Score (Coefficient of Determination)** lardan foydalanishimiz mumkin.

Bunda asosiy e'tibor beradigan jihati shundagi birichi 3 tasi qancha kichik chiqsa shuncha yaxshi, sababi u origianl $y_{\text{test}}[0] = 1$ va $y_{\text{pred}}[0] = 0,1$ o'rtasidagi farqni topadi.

R² Score esa teskarisi, qancha yuqori chiqsa shuncha yaxshi.