

## 10-dars uchun Report

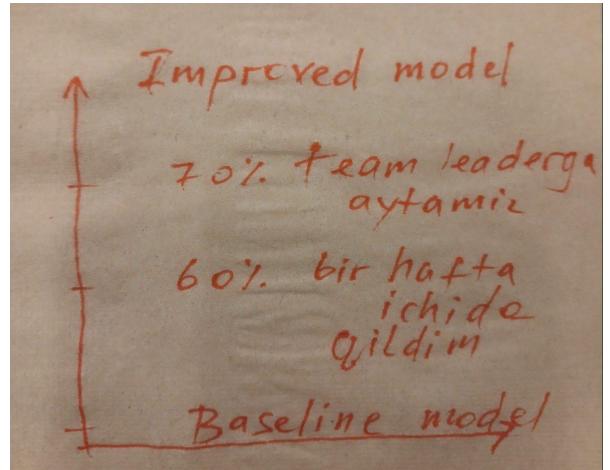
Avvalo takrorlash bo'ldi bunda biz: **Logistic Regression** ko'proq **binary classification** uchunligi haqida gaplashdik va **multiclasslar** uchun esa **Decision Tree** ishlatalish yaxshiroqligi haqida ham gaplashdik.

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

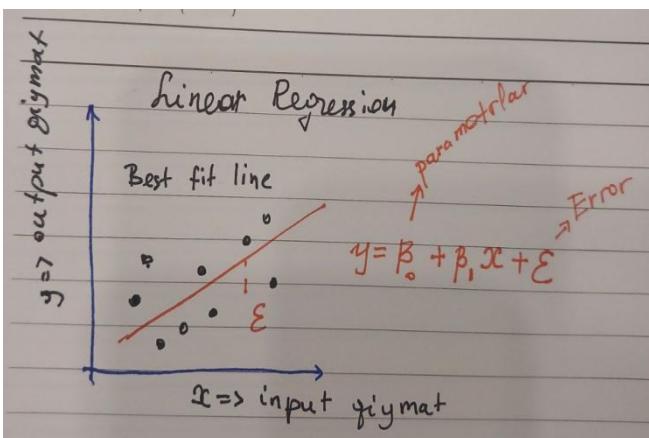
Kompyuter dastalar yordamida o'rganadi ma'lumotlar oladi, **patternlarni** o'rganadi shunga qarab turib qaror qabul qiladi.

**Linear Regressioninga** quydagilar 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 masalliqlar o'ichovi esa bu inputlardir.

**Parametrlar** input qiymatni muhimligini o'ichovchi o'icham hisoblanadi.

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

**Linear Regression** maqsadli yoki Linear Regression ishlatalayotkanimizda 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 featuralr 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 **epsilonimiz** 0 ga teng bo'lmaydi. Buni ham hisobga olish kerak.

Va **Linear Regressionni** pros va conslari mavjuddir.

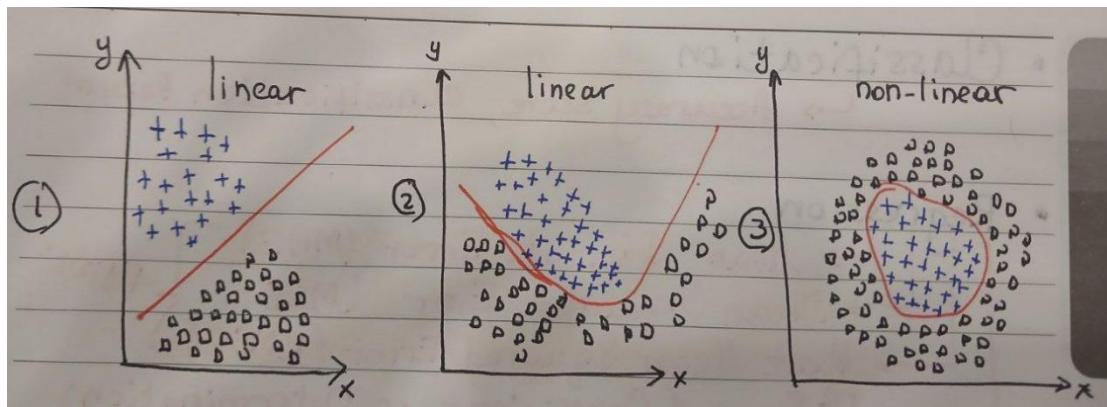
- **Pros:**

- \* 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'rishmiz mumkin Linear Regression va Non-Linear Regression uchun misolni.

### Evaluation (baholash) --> Generalization va Overfitting

Quyidagilarni baholaymiz: - Prediction qancahlilik yaxshi? (raqam bilan o'chaydi masalan [0, 100] oraliqda), - **Model** ishinarlimi? (Biz qancahlilik 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 Reportdan**. - **Regression** uchun esa biz --> **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)** va **R<sup>2</sup> Score (Coefficient of Determination)** lardan foydalanishimiz mumkin.

Bunda asosiy e'tibor beradigan jihat shundagi birichi 3 tasi qancha kichik chiqsa shuncha yaxshi, sababi u originall y<sub>test</sub>[0] = 1 va y<sub>pred</sub>[0] = 0,1 o'rtasidagi farqni topadi.

**R<sup>2</sup> Score** esa teskarisi, qancha yuqori chiqsa shunchah yaxshi.