Logistic Regression

- Classification tasklar uchun effiktive algorithmlardan biri:
  - **Sodda**
  - **Kichik datasetlarda yaxshi ishlaydi**

Logistic regression(baseline model hisoblanadi yani biz shablon qilib olamiz baseline) nisbatan soda datasetlarda nisbatda complex bolmagan holatlarda va asosan input va output qiymatlar linearly munasobatda boladi chiziqli munosabatda boladi binaryda yaqshiroq ishlaydi. Multiclass boladigan bolsa demak Decision Tree ishlatishimiz kere

Pythonda biz Data typelarni, operation, conditionlarni va for loop ni otdik

Mana shu toralasidan yaqinda imtihon soddaroq otkan darajamizdagi savoldan bersam codingda ishlab beraolasizmi? Pythonan shu PROBLEM SOLVING nan shu toralasi tushadi. Toralasidan istalgan iktasidan sorashim mumkin shunga oxshagan savolar boladi.

from sklearn.linear_model import LogisticRegression

from sklearn.tree import DecisionTreeRegression, DecisionTreeClassifier_

Raw Data----------Data Preprocessing----------Algorithm(Method, oqituvchininign otish metodi)

Model Training bu Preprocessing bolib bolgan datani bironta algorithm(metod) yordamida kompyutir miyasiga quyamiz

ML Model Structure

Data Collection

Data Preprocessing

Model(Agorithm)selection

Model Training

Prediction (Predictioni tashlab ketsaniz ham bolmaydi sababi osha predictionga qanchalik ishonshimiz ligin tekshirishimiz kerak) Misoli: y[0] deganimizda birinchidagi turgan qiymatini = Yes(1) , No(0) yoki y[0] = 200 ming dollar. Endi bu qiyamatlarga(Yes(1) yoki No(0) yoki y[0] = 200 ming dollar) qanaqa qilib ishonamiz degan savol bolishi mumkin oshani biz EVALUATION qismida korib chiqamiz. Unan tashqari bizda osha EVALUATION qilganan kiyin biz tashqaridan data topib turib yoki ichidagi dataning malum bir qismi yordamida test qilishimiz mumkin boladi. Evaluation bilan Testing ning farqi bor. Evaluation qanchalik yaqshi ishlayapti yoki yomon ishlayapkanligini korsatuvchi metric hisoblanadi. Testing bevosita natija olish qisoblanadi. Mana misoli Evaluation qildik va bu qiymatlarga (Yes(1) yoki No(0) yoki y[0] = 200 ming dollar) ishondik. Ishonganimizdan kiyin ham yana bosqa qiymat berib koramiz misoli y[2] tekshirib ber deyishimiz mumkin

Evaluation (Model Improvement degan bir katta bolim bor) 50, 60% pas korsatkich bolib qolsa accuracy scori yoki boshaqa metric lari pas chiqadigan bolsa u qiymatga biz ishona olmaymiz chunki buning aniqligi juda pas uning orniga model ni rivojlandirishimiz mumkin yangi Feature larni qoshamiz, malum birlarini olib tashlaymiz, transformation qilamiz, hyperparametr tuning qilamiz va hakozo juda koplab bolimlari bor. MODEL EVALUATION qismida yomon boladigan bolsa Model improvent qilamiz. Model improvement qilib u qaytib keladida va yana EVALUATION qilinadi. Keyin yana testing va Deployment ga otib ketishi mumkin. Yani bu cycle hisoblanadi. Boldi manashu ketmaketlik qilsak boldi model qilib oldik degan joyi yoq. Albatta buni rivojlantirishimiz kerak va yana turli qil boshqa qismlari boladi va bu

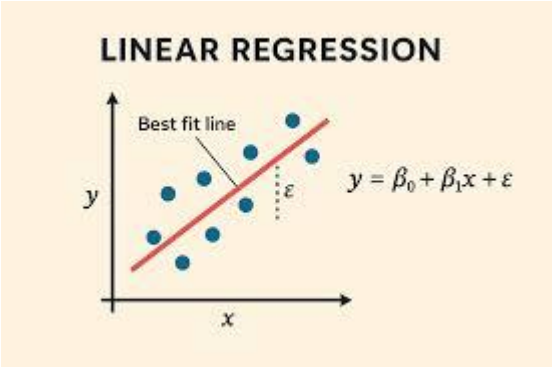
faqat shu biz organiyotgan Supervised Machine Learningga uchun chunki boshqa AI modelarida bu ketma ketliklarda ozgina farqli jihatlari yam bor lekin umumiy struktura oxshash.

Testing

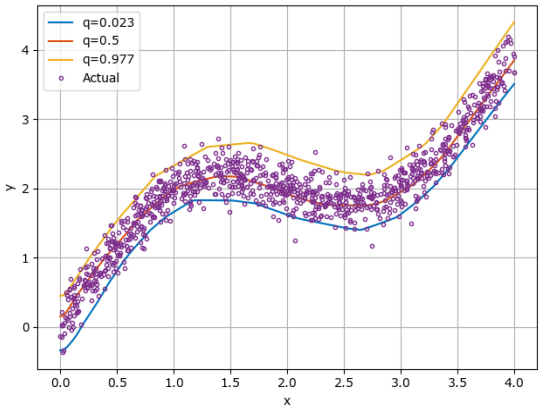
Deployment Endi hammanarasadan otdik deylik yaqshi testing qildik yaqshi yuqori foyz chiqti deylik 90-95% foyz chiqti deylik. Unan tashqari hammasi yaqshi biz xoxlagandek bolyapti va u xolatlardi keyin biz Deployment qilishimiz mumkin. Yani Deployment model ning osha haqiqiy ishga tushirish jarayoni. Yani faqatgina biror bir ipynb fayl korinishida emas va product, mahsulot sifatida chiqrish boladi. Yaniy oylanglar testing bosqichgacha biz nima qildik. Restoranimiz bor biz ovqat qildik uni ozimiz tartip kordik tuzini yoki boshqa holatlarini yaqshi sifatli lekin shu holatda qolib ketsa nima boladi? Biz bankrot bolamiz chunki biz ovqatni sotishimiz kerak shu jarayonda ovqatni bironta eshik qoyip restoraning korinishida mijozlarni chaqirib. Mijozlarga yoki eltib berishimiz kerak yoki ozlari kelishi kerak osha jarayoni DEPLOYMENT deyiladi. Yani bitta modellimzdan bitta ozimiz emas butunlay boshqalar ham ishlatish boshqalar ham foydala olish holati deployment hisoblanadi (Xudo xolasa keying oy).

LINEAR REGRESSION farqi Logistic regressionan bu regressionda ishlatiladi Logistic regression classification da ishlatilsa

Ozi asli ikta bolimnan iborat linearli va non-linearli yani misol uchun input output qiymatlarin oladigan bolsaniz x,y



mana bunaqa korinishida bolishi mumkin



yoki mana bunaqa tekis emas korinishida mumkin

Chunki biz bilmaymiz gu bizga qanaqa qiymat beriladi datalarimiz qanaqa boladi chuning uchun yoki linear yoki non linear holati bolishi mumkin. Yo bolmasa ikkiga bolinadi shu holatda yani biz nimasiga qarab bularni oilalariga ajratamzi ekan yani osha Input va Output qiymat orasidagi munosabatga kora yaniy input qiymatlari osha y ham oshiyaptimi. Yaniy odamning Yoshi osha deylik bu tomanda oyligi ham oshadimi. Shu holatlardan kelip chiqan holatda biz bularni oilalarga ajratishimiz mumkin yaniy bu orqa jarayonda. Yaniy input va output qiymatlar orasidagi munosabatdan kelib chiqan qolatda biz buni ajratishimiz mumkin hisoblanadi. Linear ham huddi shunaqa. Va unnan tashqari yana bir holat borki bu baseline model hisoblanadi

Improved Model

↑

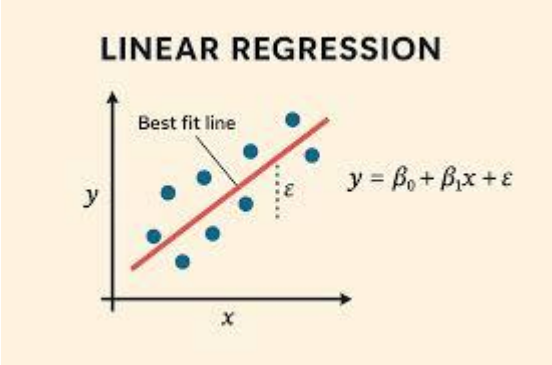
---------- Baseline Model

Birinchi manashunaqa kuchizroq modelar bilan tekshirib olamizda. Yani nimaga dur qarab baholaydiku.

Misol Team da ishlatoygan bolsanizlar lider kelip sizdan sorashi mumkin nima qilyapsan dep va siz aytasiz men baseline model qurdim va baseline model 60% accuracy berayotgan bolsa shu hafta ichida man uni 70% ga kotardim endi Keyingi hafta 85 % ga kotarmoqchiman deysiz va model improved ga olib kelasiz.

Aytkanimizdek Regression (Continues). Output qiymatlar davomiy bolsa biron discrete oraliqa tushmasa continues boladigan bolsa biza Regressiondan yaniy Linear Regressionan foydalanamiz.

LINEAR EQUATION biza maktabta otkanmiz y=kx+l korinishida bolsa chiziqli holat hisoblanadi yaniy bir olchamli chiziqli holat hisoblanadi demak Mana bu Linear Regression va Logistica Regression orqasida aynan input va output qiymatlar orasida huddi shunday munosabat bor.



y tomanimizda bizda output qiymat x tomonimizda input qiymatlar boladigan bolsa va y=B0+B1x+e chiziqning korinishi. B0, B1 parametrlar hisoblanadi yaniy input qiymatimiz qanchalik kuchga egaligini qanchalik muhimligini korsatadi parametrlar hisoblanadi. Va epsilon e bizda error hisoblanadi hatolik hisoblanadi. Misol uchun osh qilyapsizlar OSH=0.5kg gosht + 1kg guruch+0,001 kg duz. 0.5, 1, 0.001 parametrlar hisoblanadi. Gosht,guruch va duz input hisoblanadi va oshimiz output hisoblanadi. Parametrlarning muhimlik jihati kattada chunki sizlar ozroq qoshaniz yoki koproq qoshanlar bevosita ovqatni tamiga mazasiga tasir qilib qoyasizlar shuning uchun meyorini bilish kere. Shuning uchun ham bu parametrlar hammada har qil boladi. Hamma dataseta hamma loyihada har qil boladi. Sizning oshiniz boshqacha bolishi mumkin. Kimdur koproq yog qoshadi kimdur oziroq qoshadi yaniy bironta standard yoqku togrimi. Bir kg osh uchun shunaqa qilishimiz kere degan joyi yoqku togrimi. Togi oxshash standard lar bor lekin malum bir ozgarib qoladigan bolsa balki u oshpazdan yaqshi qilishiniz mumkin va aksincha yomonroq qilishiniz ham mumkin. Demak Parametr bu input qiymatlarini muhimligini olshovchi olcham hisoblanadi. Linear Regression maqsadi iloji borincha hamma nuqta ga yaqin bolgan best fit yaniy eng zor mos tushadigan chiziqni topishdan iborat boladi shuning uchun ham nomi linear da buni. Lekin linear chiziq chizish imkoni bolmasa complex bolib ketadi. Demak mana bu parametrlarni errorlarni turli qil usullar orqali ozgartirish qiymatlarini oshirish orqali biz linearli holatga qaytarishimiz yaqinlashtirishimiz mumkin boladi

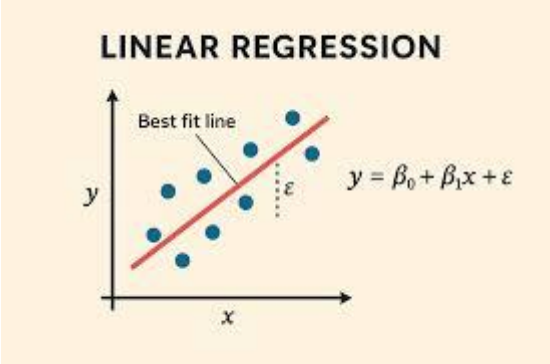
Linear Equation (chiziqli tenglama). Demak linear regression maqsadi yoki linear regression ishlatiyotkanizda natija yaqshi chiqyaptimi. Qanaqa holatda natijasi yaqshi chiqarkan qachonki shunaqa linear chiziq paydo bolsa iloji boricha tekis aynan shunaqa 100% tekis otishi shart emas

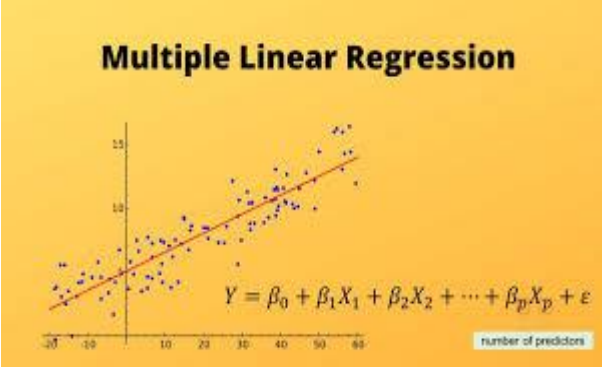
Minimum holat otkan narsalarni aytip bera olishlik maksimum holat bularni kutibxonasi bor shunchaki Linear Regression docs deb yozib Linear Regression ga oid juda kop malumotlarni chiqarib beradilar bu sickitlearning da ikta kutibxona orqali.

Linear equation Linear Regression maqsadi osha input va output qiymatlari uchun best fit line toppish.

Linear Regression

Simple LR Multiple LR (kop holatlarda Multiple Linear Regression ishlatilinadi, deyarli kop holatlarda bu aloqada aloqada codlar, aloqada aloqada yana bironarsa kere emas)





Qaranglar Multiple Linear Regressionda input qiymatlari judayam koyab ketti input qiymat kopayganan keyin oz ozidan xammasinin oldida turuvchi bettalar yaniy parametrlar kopayadi. Yaniy ovqatga qancha kop mahsulot qoshsanglar albatta mahsulotning hajmi olchami boladiyu ular ham oz navbatida kopayadi . Simple LR va Multiple LR regression orasida farq Simple LR bitta feature input.

Error bolsa har doim osha errorni kamaytirishga harakat qilasak. Error,Epsilon!=0 Umuman mukamal bolgan data mukamal bolgan algorithm yoq. Lekin error maksimal darajada kamaytiramiz degan holat bu tabiy holat

Linear Regression

- Kuchli tomonlari
    - Tez va soda(jarayon osonligida)
    - Kichikroq datalar bn ishlashda oson
    - Baseline model (Shuning uchun birdan gina kuchli model(birdangina murakkab holatni olib ketadigan bolsaniz unda yaqshi natijaga erishib bolmaydi) Baseline nan boshlab osha qilib atirgan ishlarimiz oxshayaptimi rivojlantirish metodlarimiz rol oynayaptimi yoki yoqmi degan savolrga javob berishimiz mumkin boladi
    - Kam chiqim (low cost hisoblanadi yaniy osha training bolayotgan jarayonda oz material oz mablag’ talab qilinadi)

Intuitive explanation

Imagine many points scattered on a graph.

Linear regression asks:

“Which single straight line stays, on average, as close as possible to all points?”

- If the line moves up → errors increase for many points
- If it tilts more → errors increase elsewhere

The final line is a **global compromise**.

Step 1: Look at all points

The machine sees many pairs:

(x1,y1),(x2,y2),…(x_1, y_1), (x_2, y_2), \dots(x1,y1),(x2,y2),…

It does **not** pick one point.

Step 2: Guess a line

It starts with a random line.

Step 3: Measure how bad the line is

For **each point**:

- Go vertically to the line
- Measure the distance (error)

Then:

- Square each error
- Add them all

This gives **one number**: total error.

Step 4: Adjust the line

- Tilt the line a little
- Move it up or down
- Recalculate total error

If the error is **smaller**, keep the change.

Step 5: Repeat

It keeps adjusting until **no small change improves the error**.

That line is the answer.

Same idea with very simple math intuition

- Too steep? → large errors on one side
- Too flat? → large errors on the other side
- Just right? → smallest total error

One key fact (important)

The final line:

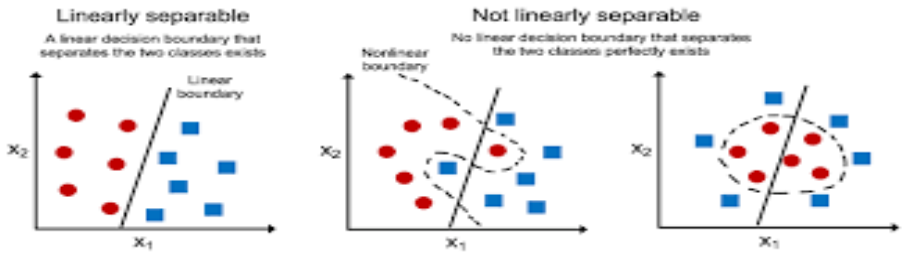
- Is influenced by **every point**
- Is closest **on average** to all points
- Passes through the **mean point** ($\bar{x}$, $\bar{y}$)

One-sentence mental model

Linear regression keeps nudging a line until the **sum of squared vertical distances to all points is as small as possible**.

Agar sizlarda maqsatiniz kichiroqmi agar datasetiniz ozroq bolsa kickinaroq bolsa yaniy juda katta boladigan bolmasa va boshida modelni qanaqa ishlayotganligini baseline holatida tekshirib olish holati bolsa va albatta Regression uchun boladigan bolsa Linear regression ishlatib ketish maqul boladi

- Kuchsiz tomonlari
    - Compex(non-linear) xolarlarni o’rgana olmaydi
    - Yot elementlarda tez o’zgaruvchan(stabil emas)
    - Katta datasetlarda yaxshi ishlamaydi



Hozirgi holatda social media tasiridan malumotlar kopayapti(oldin raqamlashtirish bolmagan hozir qog’oz bozliktan ketip raqamlashtirishga otiliyapti) va malumotlar kopaygani uchun linearli va complex(non-linear holatlar juda kop boladi)

Evaluation

- Quyidagilarni baxolaydi:
  - Prediction qanchalik yaxshi?
  - Model ishonalarmi?
  - Generilization? Overfitting? Bolib bolganan keyin
  - Generilization umumiylashtira oldimi? Overfitting bu yodlab olish misol uchun oquvchilarni oladigan bolsak test varaqalarini yodlab olgan boladi lekin tushinib yechmagan boladi agar javobdagi variantlarni almashtirib qoysak unda natija yomon chiqadi oquvchi yodlab olganda. Demak EVALUATION da ham biza generalization yoki overfitting ligin aniqlashtirib olishimiz kere boladi

Evaluationda

- Quyidagi bo’limlari bor:
- Classification
  - Accuracy score, Classification Report
- Regression
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - R^2 Score (Coefficient of Determination)

y[0]test=1
3 Error 0.9 errorni hisoblaydi
y[0]pre=0.1

Demak hulosa qiladigan bolsak hachon bizning model yaqshi boladi qachonki shu errorlar kichkina bolsa

R2Score esa accuracy ga oxshab yuqori chiqishi kerak va osha qanchalik model yaqshi organdi yaqshi organdimi yoqmi kabi savolargajavob beradi. Error lar esa hammasi kichkina bolishi kerak.

RMSE = ildiz ostida MSE