

ML Definition

→ (Informal) algorithms that automatically improve experience.

الgoritم کسی کہ بصرت اور خالک تجربی وجود را ہدایت کرے۔

+ ML relying on pattern and inference rather than explicit-instructions.

درستاری تینی الگوریتم کے بعد سری کوئی با استنتاج کی واسیہ ہستہ وہ دنالی سری۔
دستور العمل کی صریح نہیں۔

ex.1: Sort a list of numbers.

e.g.: 5 7 9 4 7
اعمال الگوریتم
1 4 5 7 9

ex.2: classification of tumor type based on gene expression data.

ذخیرے می خواہیم tumor type را بآسانی سلسلی داری سیان ٹن منطبق کیتم۔
مشتھ خزن کیڈ درا سن 20000 ٹن داریم وھ کوڈم کیڈ Value داری دستور منطبق وجود دنالا داری
کہ واسیہ بیان ٹن کہ داریم منطبق کند تو تو ختن Breast اسٹ لے lung۔
ایضاً حلقہ بندی بآسانی الگوریتمی اسٹ کہ ازدواجی تبلیغہ ایم وہ آسان آہنا استنتاج می کیتم۔

این دو مثال، عنوان کی ای زیارگری مائن میں میں



Applications of ML:

+ spam identification

براساس ایمیل وجود در ایمیل مخصوص گویی کردن

- مخففی است ایمیل هست یا نه (نمایه ایمیل براساس

(کارهای طراحی وجود spam ایمیل ۱۰۰۰۰)

+ Handwritten digit classification

سیستم خواص از این دستوری است و آن

بین دو دسته کنترل

prediction task



supervised learning

+ prediction of tumor type

+ prediction of HIV drug resistance,

from genotype data

پتانسیل ترجیحی دارایی داده هاست باشند و مخفف این

عزم مخفر هست یا نه.

+ identifying the risk factors for prostate cancer

Inference task

supervised learning

+ motif discovery

متدهای تابعی میان متن

+ Molecular subtyping

Unsupervised learning

(knowledge discovery task)

سری داده از عبارات رسمی و مخفف این بینم از پنهان

ایمیل ایمیل (بدون اسناده از چیز اطلاعاتی)



قرآن کریم دوڑن را یعنی دو خواص هم جایگزین کنیم. Cancer type (انواع سرطان) است.

gene	G ₁	G ₂	Cancer type
	100	70	Lung
	4	2	Breast
	94	107	Lung
$x[4] \rightarrow$	92	103	Lung
	7	4	Breast

+ Simple pattern

در اینجا اینکه داده‌ها را با یک مجموعه آنرا در 2 از دسته بزرگتر باید بدانیم.

• اول Breast Cancer و Lung Cancer

• ایجاد مجموعه داده برای ML

Three classes of ML problems:

① Supervised learning

map داده‌ها را به output که input کرده‌اند.

* learning a function that map some input variables to an output based on a set of examples.

$$(x[1], y[1]), \dots, (x[n], y[n])$$

n: # observations

x: input variables $\xrightarrow{\text{other names}}$ features, predictor, independent variables

مثال: چندین ایمیل را ایجاد کنید تا مخفف را ایجاد کنیم -

کارکرده خاص و بودن را نشان دهیم

y: output $\xrightarrow{\text{other names}}$ target, Response, dependent

$$x[4] = \begin{bmatrix} x_1[4] \\ x_2[4] \end{bmatrix} = \begin{bmatrix} 92 \\ 103 \end{bmatrix} \quad y[4] = 1$$

↓
Lung = 1

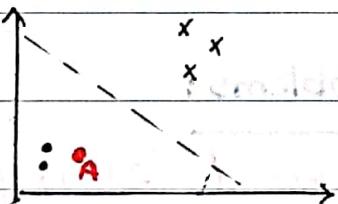
: مُعْنَى التعلم المراقب بالإنجليزية \rightarrow supervised learning

Response Variable is a discrete Var (qualitative) \leftarrow classification (I)

" " " Continues (quantitative) ← Regression (2)

متاً جای سین ارزش بی کانه براساس هست آن از کرسون استفاده نمی کنم.

ex1: classification

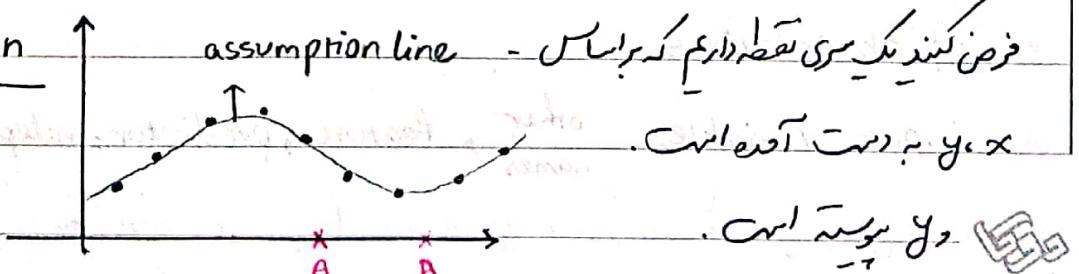


طیور میتوانند خرمنیات تغذیه A را خود نمایانند و بروجابت Breast در پستان خود فرضنیه -

حُمَاجِنْ كَارْبُونْ جَرْجَنْ / Breast Cancer مُوْجَهَةٌ تَرْدِيْكَ رَاسِ سِنْ دَرَرُودَ آخْرَاسِ وَ

لابل لاین نامه سرویس

ex02: Regression



اگر y میتواند از x_1, x_2, \dots, x_n بسته باشد و $y = f(x_1, x_2, \dots, x_n)$ باشد، آنها را متغیرهای مستقل می‌نامیم و y را متغیری مستقل می‌نامیم.

با این فرض y را میتوانیم با x_1, x_2, \dots, x_n مدل کرد.

با این فرض y را میتوانیم با x_1, x_2, \dots, x_n مدل کرد.

با این فرض y را میتوانیم با x_1, x_2, \dots, x_n مدل کرد.

ex 3: (ESL book) \rightarrow prostate cancer

multivariable regression

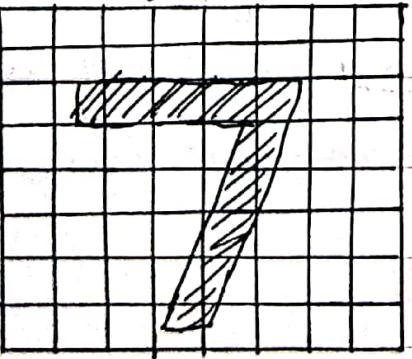
<input type="checkbox"/> input Variable : age <input type="checkbox"/> output " : PSA level <input type="checkbox"/> other inputs : prostate weight	univariable regression
---	------------------------

Gleason factors

point: input variable can be categorical or continues in either classification or regression.

ex 4: classification

\hookrightarrow digit classification



ابعاد سیمبل بینی 8x8

هر خانه یک عدد است

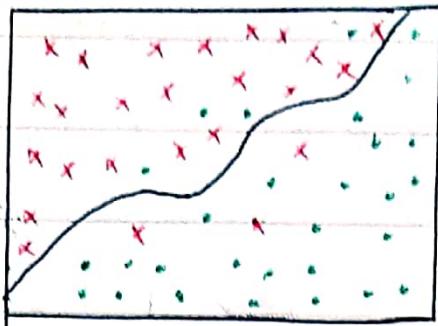
مثلاً ۰۱۰۰۰۰۰۰

Image feature $\leftarrow x$

عدی نماینده موجود است $\leftarrow y$

$$y \in \{0-9\}$$


ex 5: classification



میکنی سری نتایجا با x مشخص داریم
و y آن را نگیری و میز آهناست.
Boundary اما (یا پل آن)
بنی تغیروه تنها مشخص نهادی
دستن ایست که همچویر باشد.

② unsupervised learning

نیاد داریم که بعضی از این اور

هدف این ایست که برای سیم (بینی) بعضی از data \rightarrow pattern بسازیم (بینی) بعضی از جویبار

$(x[1], \dots, x[n]) \rightarrow$ find pattern in the data

↓
interesting things

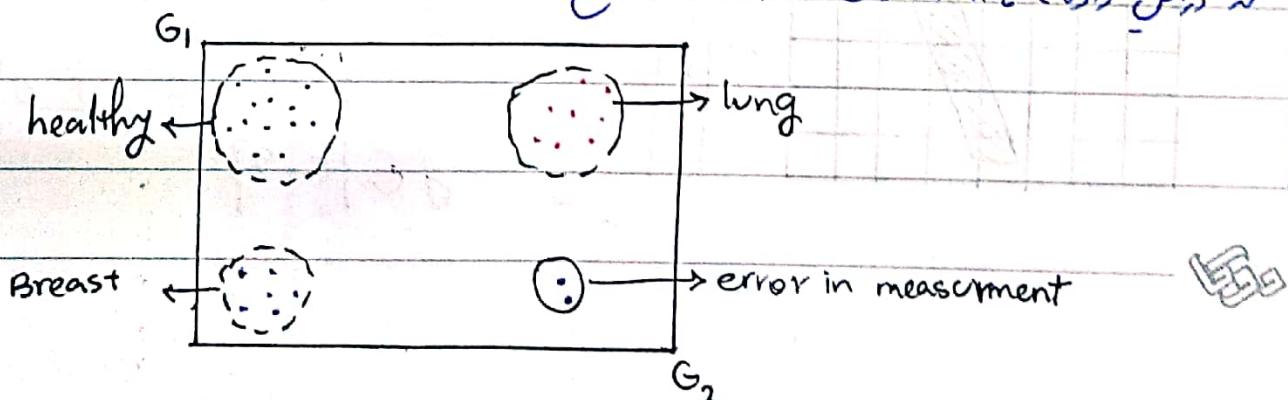
- Civil clustering \rightarrow unsupervised learning وجود دارد

* خوش شنیدی که زادی بنیان نشان گرفته ایم و دریس کرد ایم به مردم زیره عکس رو (ستنبی) چی خوند

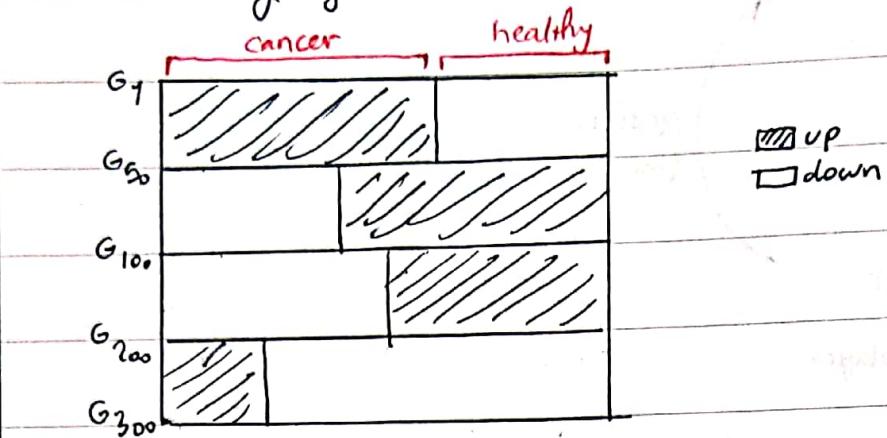
Lung Cancer, Civil Breast Cancer و ... درینم که بینی کروه

بعد این بینتر در رس کی سیم که درینم که دو مورد هم از سایر کروه های پرست هستند، با این بینتر مشخص کی کروه

- Civil \Rightarrow error in measurement



ex: interesting gene sets from an expression profile



- سینه کردن آنها با الگویی، مسح جی کم کروه clustering! \rightarrow sort
- . down سینه و up sample از ۱-۵۰ دریک سری از
- بعد با طاله ای دستی ترجیح می کنی \rightarrow DNA replication
- . موندانه سینه pathway enrichment \rightarrow این نو

HW: learn about blind source separation & cancer signatures.

(unsupervised) \rightarrow دستیار و متادادهای دستی از

(real)

③ Reinforcement learning (RL)

روش دادنی تعلیمی در ریاضیاتی کارکردی داد و بسته باید

Robotics - } ۱۵, ۱۶ ← مفهای

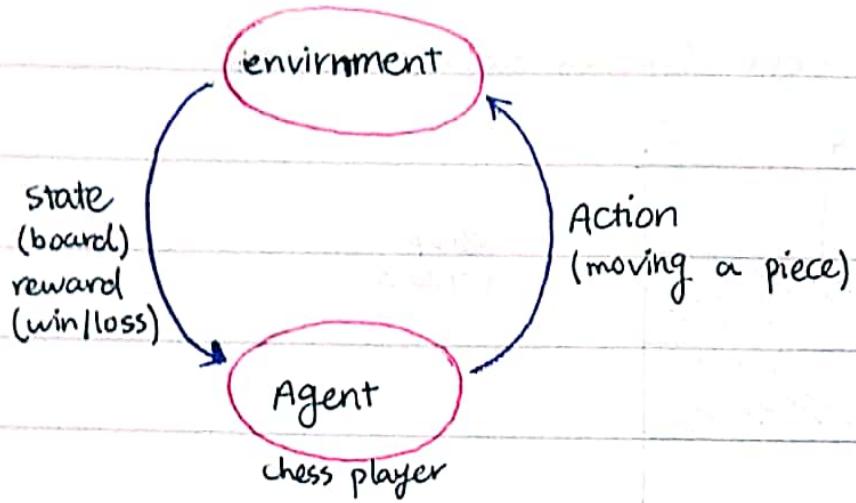
self-driving car - }

(like chess) playing games - }

learner - Agent در محیط وی environment



بررسی کن طبعاً learner می خواهد



ریانه هایی که از این طبقه هستند را **reinforcement learning** می نامند.

شما شخصی کنید و بازی طبقه شخص بروید با ساخته ای ایم.

ازین طبقه **supervised learning** نیز نوشتی دارد:

درینجا عامل سروع بازی صریح باطراف مقابل می کند (که حی تو انداسان باشد programming باشد).

اینها وصفت board را که کرده و چون برای باخت نیست همچنان reward به آن را داده شود.

حالا خود و چنین طور State2 خود را داده تا اینجا که در و دو همچنان بازد.

$$S_1, a_1, S_2, a_2, S_3, a_3, S_4, a_4, S_5, a_5 \Rightarrow R = -1$$

(loss)

درینجا حالت S_n action کلی جمل از a_n نیز نباشد.

حی دهد و آنرا را به عنوان استخبار کنی که منزبه باخت من شود در نظر می کرد.

$$S_1, a_1, S_2, a_2, \dots, S_n, a_n \Rightarrow R = +1$$

↑ (win)

به اینها اعیان خوبی می دهد

در سال ۱۹۹۲ با چنین مفاسدیم، برای بازی کردن تخته نبرد، اجازه دادند دو learner بازی کنند.

و درینست پس از خود را شدن الگوریتم ایجاد شد که برای مسابقات توانستند از آن استفاده کنند.



مثال: $\text{لذت حزون} \leftarrow \text{دوافع حبوب و ملائكة} \leftarrow \text{دوافع حبوب و ملائكة} \leftarrow \text{دوافع حبوب و ملائكة}$

— supervised \rightarrow اس کو supervised درج کریں Reinforcement پر سے تقدیر کریں

$S \rightarrow A$ هر دسته ای خوب است یا نه.

، هر action c state s، را عن داشتم (Reinforcement)، این را عن داشتم (action reinforcement)،

ادسیز، Crl reward و $\text{Crl reward}_{\text{new}}$ ، $\text{جور دنلدر} \rightarrow \text{جور دنلدر}$ training data

• Class 1 unsupervised, supervised, reinforcement

Other Variations: (4) semi-supervised learning

در اینجا کتابخانه‌ی label، unlabel، در دارم، عجیب نیست سری داده‌ی هم دارم.

Given $(x[1], y[1]), \dots, (x[k], y[k]) \rightarrow$ label $\cup \{b_j\}$ (expensive)

$x[k+1], \dots, x[n] \rightarrow \text{unlabel}_{\{\varepsilon_0\}}(\text{cheap})$

predict: $y[k+1], \dots, y[n]$

- در اینجا با استفاده از مخل (داده‌ی) label (دار و unlabel) حق توانیم عدل سازیم. (درواقعه با برکلیب)

کردن این درسته از درده سبیل حسوم کر دست learner که حس از عزم بسترسود



+ one-class classification \rightarrow semi-supervised learning

مکالمہ کا حصہ میں negative data، سخت داریم، مساویں حالتیں سنیں۔

لک سری دارو، target، میانگین وجود دارد اما در عورت میانگین نداشتن اطلاعی نداریم.

* two drug targeting interactions

اینها با استفاده از راهی **positive** که داریم، در عورده اینه آیا داریم، میتوانیم (اردیابیز) ،

جیٹ جی ستم۔

other variations → ⑤ Active learning

* the same settings as semi-supervised learning

نواتیت با semi-supervised سمت generate label این اسکر نهایا جو توانم باشد، با اینکه هنوز طوری جو کنم، نهان Random label کناری جو تواند باشد،

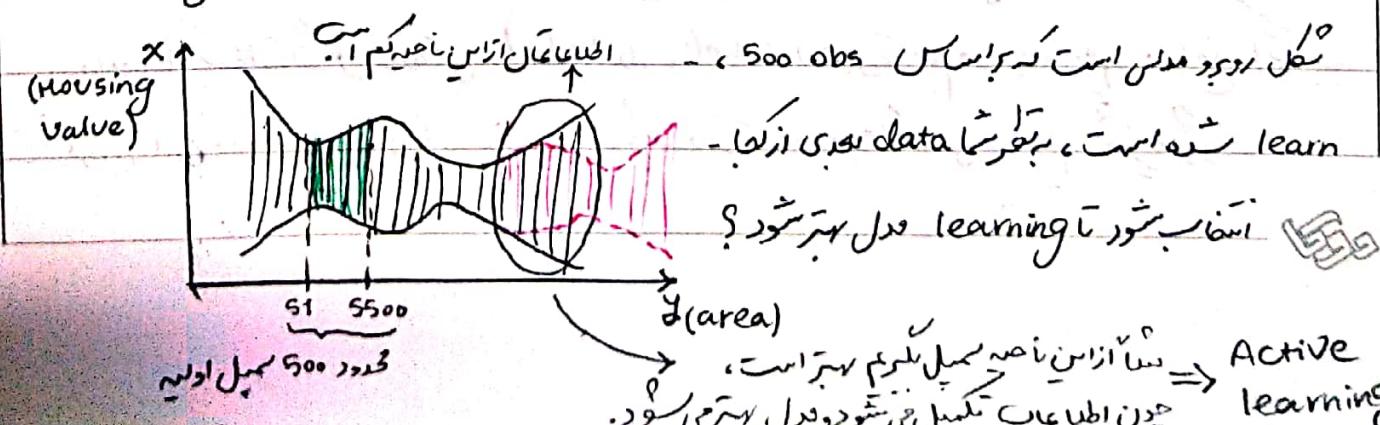
• Over active learning, \rightarrow الحمد لله رب العالمين, رسان label

حال: صد نرض ایتم سری بصوری داریم که label دوچیخ خود را اند، حالا این سری یعنی مارسین و

دوجه خواهی و خواص آنها در درس را با سفر نظری active learning می‌دانیم.

پرچم بـ label (ارجعی چنی، سـ 500 unlabeled) دو چهارم اسـ تـ استخـابـ هـ سـورـ

ex: a regression with 500 obs.



کاری کوچک است و محدود نیست
 اما باز هم محدود است و محدود است

K-Nearest Neighbor (KNN)

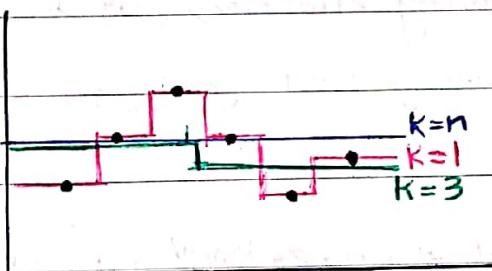
↳ + a simple method for both classification & regression.

Given \underline{x} and \underline{X} , we find $N_k(x)$ ($k=1, 2, \dots, n$)

KNN Regression

نحوه این است که x را با k نزدیک ترین مجموعه از X را پیدا کرده و این مجموعه را برای $f(x)$ استimation کنیم.

$N_k(x) \rightarrow k$ closest observation \Rightarrow distance measure



$$\hat{f}(x) = \frac{1}{k} \sum_{i \in N_k(x)} y[i]$$

hat: estimate

$$\hat{f}(x) = \frac{1}{3} (y[1] + y[2] + y[3])$$

$k=3$

نحوه این است که $k=1$ ، میانه ای است و $k=n$ میانگین است و $k=3$ میانگین سه عدد است.

- training error \rightarrow بزرگتر است که k کوچک است

$$d(x, x[i]) = \|x - x[i]\|^2 = \sum_{j=1}^d (x_j - x_{j[i]})^2$$

distance اندیسی برای

$k=1 \rightarrow$ training error = zero

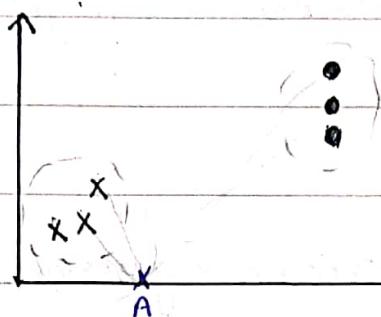
$k=n \rightarrow$ simple model

مدل بسیار ساده است و درینجا
کاری کوچک است.

Car min loss

learning in memorization

- KNN classification → majority voting among neighbors.



اگر خواهیم یافته طبقه A را بوجه بخواهیم

مقابل مخفف کنیم، با این رأی ترجیحی میان K

محسایی می‌شود و بدست حی آوریم. اینها

که درین مخفف کنیم آن نظریه می‌شوند.

چنانچه خروجی دادم دسته قرار گردید. (با استفاده از فاصله نظریه ای برآمده)

- training error vs test error

وقتی دل را روی training error و گزینه K=1 برخیزیم جواب را می‌شود، بنابراین

وقتی را روی دادی test می‌شوند. (داده‌ای که می‌خواهیم شنید)

مسئلای که در KNN وجود دارد این است که تابع نامهواری به دست مع آمد.

بنابراین KNN را بقیه دسته‌های دوام حل نمی‌شود و همچنان دسته‌های دیگر نیز نمی‌شود.

- Nadarayan-Watson kernel regression

+ one dimension

+ smooth estimator →

تفاوت این دل با KNN داشت که می‌توانیم کراوفون (ارجی) کنیم، یعنی نطااط که بنتظری خود را در نظر می‌گیریم و آنرا که در برآورد نهادن کنیم. که تابع وابسته به X است.

$$\hat{f}(x) = \sum_{i=1}^n K(x, x[i]) y[i]$$

← صفر باید KNN, 2K

وی ایجاد وسیع تر

$$\sum_{i=1}^n K(x, x[i]) \rightarrow$$

داره.

$$k_\lambda(x, z) = D\left(\frac{|x-z|}{\lambda}\right) \rightarrow D(t) = \begin{cases} \frac{3}{4}(1-t^2) & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}$$

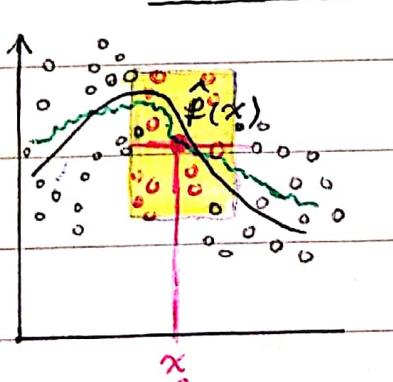
خوب ببین (تابع نویسته شده) در واقع این راشخص می کند که اگر فاصله ای نهایی روی محدود است. فقط می خورد تظر لکراز ۱ باشد، تا سر دار دلی اگر فاصله بین از ۱ باشد همچو اگر نباشد.

$\boxed{S = \lambda}$ باشد، در چه حالتی می سایه (برای مرتفعه چی می شود؟)

آنکه تویر که تابع حد میانی را در تصریح می کنی که بنده می خورد تظر برای لکراز ۱ تا سر بیشتری دارند، دیگر نیست که دورتر شاید کسر. (رابطه λ و K می تضمین)

هر دو روش KNN و KNN خوب روشی هستند. معنی از روی - وجود تجربه ای از data

(ESL Book) \rightarrow Figure 6.1

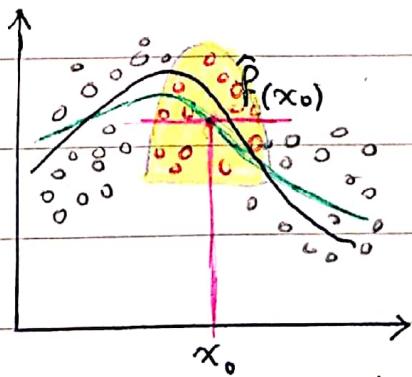


nearest-neighbor kernel

\downarrow
RNN

دستینا همچو وزنی بر تغطیه از

و تابع تجربه زده شده هموار است



Epanechnikov kernel

استیا براساس فاصله ای هسته از

فقط می خورد تظر، به تغطیه از

و تابع تجربه زده شده هموار است.

ML - 2

Linear regression (LR) \rightarrow a parametric model

از خود چیزی نهاده و دست نماینده آنها باشند) جزو تظریه ای Nadaya-watson, KNN

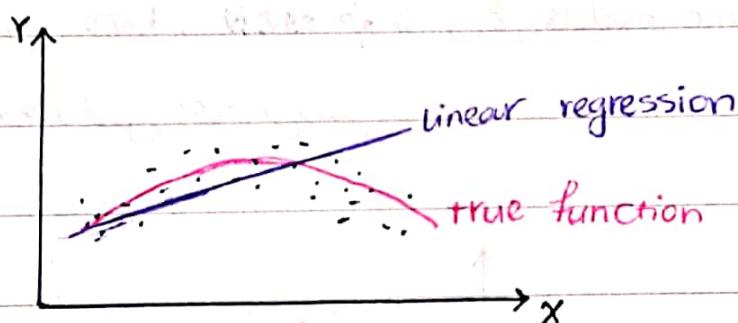
نیز معمولی خود استفاده می کردم و بنابراین predict

ما در اینجا می خواهیم از data، اسارتگری رسم، پارامتریک داشت
پارامتریک رای استفاده کنیم.

هر دوی از y ، KNN \rightarrow y assumption

+ LR assumption \rightarrow dependence of y and x_1, \dots, x_d is linear

+ true function is always nonlinear



- در اینجا چند دستی نموده داریم اسارتگری تابع واقعی داشت و می خواهیم

استفاده کنیم، اما با این حال از عمل linear regression به دلیل احتیت و کاربردی نماید

- برای استفاده کنیم، هنوز اینکه چیزی که در حالت هندی سیستم وضعیت

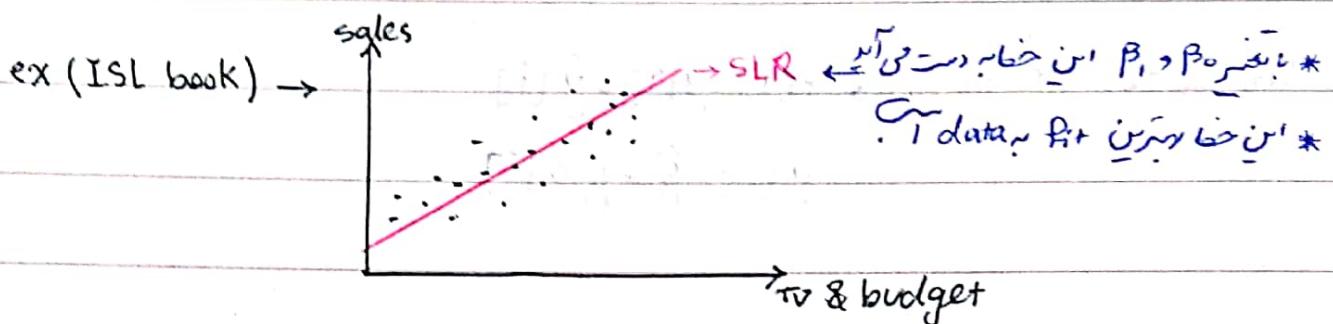
دارد و چیزی که محدودیت را تحدی Handle کرد

① Simple Linear regression (SLR)

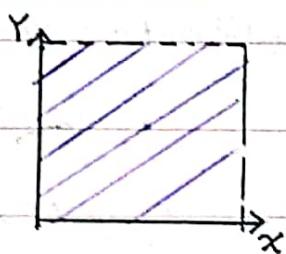
↳ + predicting Y based on only a single predictor(X)

$$Y \approx \beta_0 + \beta_1 X$$

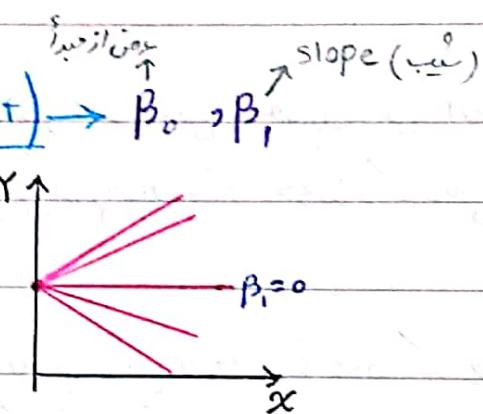
approximately is modelled as



+ parameter (model coefficient) → β_0, β_1



changing β_0



changing β_1

+ estimating β_0, β_1 → given training data → $(x[1], Y[1]), \dots, (x[n], Y[n])$

estimate $\hat{\beta}_0, \hat{\beta}_1 \rightarrow \hat{Y} \approx \hat{\beta}_0 + \hat{\beta}_1 x$

مقدار دلخواهی ← $\hat{\beta}_1, \hat{\beta}_0 \rightarrow \beta_0, \beta_1$

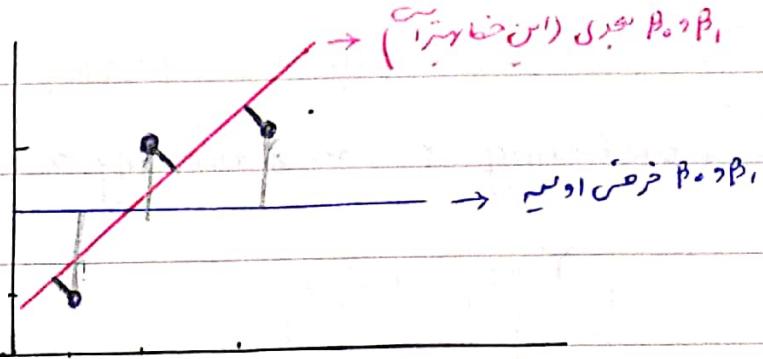
aim: Find parameters values that fits the data better

[chap1, FCML بـ $\hat{\beta}_0, \hat{\beta}_1$]



(نمونه) اولیه داده های خطابی

ex.	$x[i]$	$y[i]$
1	1	5
2	3	13
3	5	17



residual \rightarrow for i th obs. $\rightarrow r[i] = y[i] - \hat{y}[i]$

$$\hat{y}[i] = \hat{\beta}_0 + \hat{\beta}_1 x[i]$$

-loss function

در شال اول بطور تقریبی مساحت کمینه کننده خطابی است و در بمحضی که نیاز داریم در واقع بدلی $\hat{\beta}_0, \hat{\beta}_1$ و loss function نیز داریم، پس سه نوع سینه از $\hat{\beta}_0, \hat{\beta}_1$ و loss function در optimization بحث کنیم.

minimize

در طبقه دو نوع تابع برای optimization داریم که loss function و likelihood تابعی که سمت دیدار را در خواهد داشت. loss function و likelihood با خواسته پارامتر $(\hat{\beta}_0, \hat{\beta}_1)$ را تجربه کنیم.

① squared loss function:

$$L(\beta) = \frac{1}{2} \sum_{i=1}^n (y[i] - \beta_0 - \beta_1 x[i])^2$$

گام منتهی تری و ساده تری در روش optional

برای minimize کردن استفاده کنیم.



② Absolute loss function

$$L(\beta) = \frac{1}{2} \sum_{i=1}^n |y[i] - \beta_0 - \beta_1 x[i]|$$

+ optimization problem $\rightarrow \beta_0, \beta_1 = \arg \min_{\beta_0, \beta_1} L(\beta)$

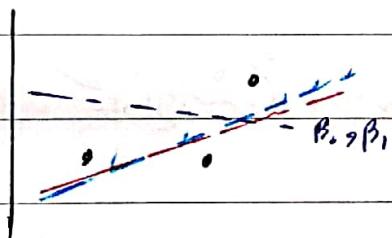
$\rightarrow L(\beta)$ در $\beta_1 > \beta_0$ که C_w نام دارد، loss function می‌باشد. همچنان که β_1 را minimize کنند، $L(\beta)$ را $\beta_1 > \beta_0$ که minimize کنند.

- analytical برای بروش کمتر کمتر کردن loss function minimize

حلیم کرد، اما اینها بروش ساده تری را مطرح کردند

+ gradient descent algorithm

\rightarrow an iterative algorithm \rightarrow یعنی مسأله اینه ۳ خط داریم استراحت اول با-



- آنرا بخود (تصادفی)، سپس

- این روند تکرار می‌گردد و β_0, β_1 بهتر می‌شوند

- minimize loss function کرده تا بخود

* Algorithm

+ start with an initial guess for β_0, β_1

+ iterate (stop، تحدید شود، تکرار، تکرار شود)

$$\beta_0 := \beta_0 - \alpha \frac{\partial L(\beta)}{\partial \beta_0} \rightarrow \text{partial derivative}$$

هزینه
هزینه
هزینه
هزینه

learning rate

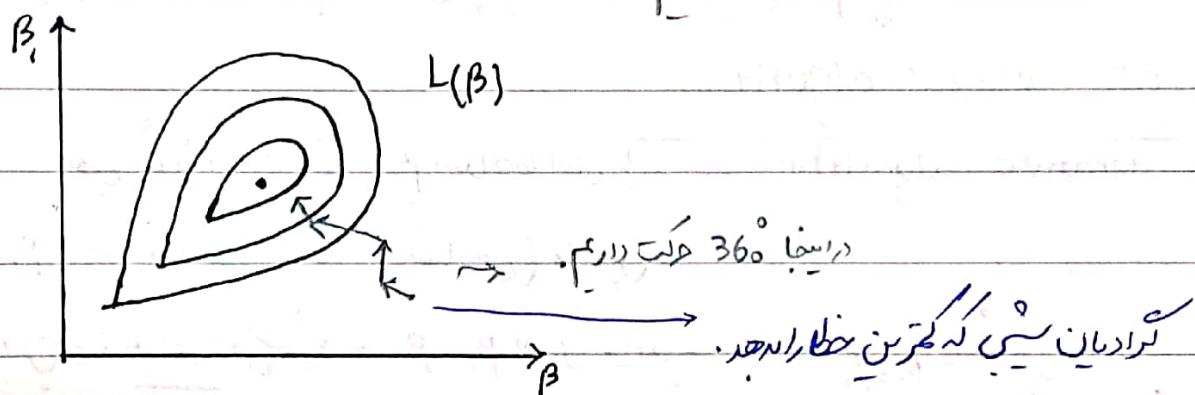
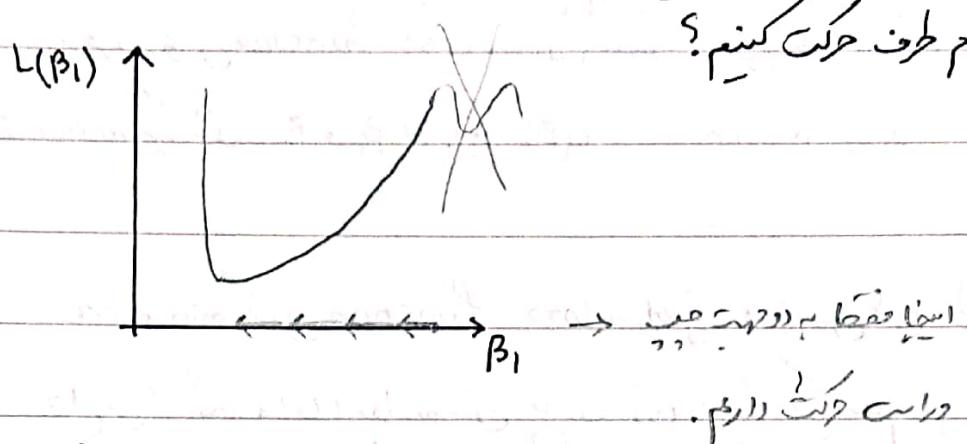
- $\beta_1 > \beta_0$ همچنان

$$\beta_1 := \beta_1 - \alpha \frac{\partial L(\beta)}{\partial \beta_1}$$

- update

اگر منظق اول در حیث بعد برسی کرد، سفرم گرادیان حاصل می‌کرد. مسأله اگر منظق را هم نسبت به β gradient برسی کنیم ← دھم β برسی کنیم

ex. initial guess β_0 را در عالم و حالات $L(\beta)$ را با خواصی بررسی کردیم. فرضن کنید β_0 نقطه A است. بکدام طرف حرکت کنیم؟



Gradient descent for squared loss function

$$L(\beta) = \frac{1}{2} \sum_{i=1}^n (y[i] - \beta_0 - \beta_1 x[i])^2 \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$\frac{\partial L}{\partial \beta_0} = \frac{1}{2} \sum_{i=1}^n -2 (y[i] - \beta_0 - \beta_1 x[i])$$

$$\frac{\partial L}{\partial \beta_1} = \frac{1}{2} \sum_{i=1}^n -2 x[i] (y[i] - \beta_0 - \beta_1 x[i])$$

updated formula:

$$\beta_0 := \beta_0 + \alpha \sum_{i=1}^n (y[i] - \beta_0 - \beta_1 x[i])$$

batch gradient descent

$$\beta_1 := \beta_1 + \alpha \sum_{i=1}^n x[i] (y[i] - \beta_0 - \beta_1 x[i])$$

all observations

stochastic gradient descent

- جرایمی این روش برخلاف روش gradient descent میباشد که در هر مرحله فقط یک observe را در نظر میگیرد و با این روش سرعت update بسیار کمتر است.

$$(x[i], y[i])$$

$$\beta_0 := \beta_0 + \alpha (y[i] - \beta_0 - \beta_1 x[i])$$

$$\beta_1 := \beta_1 + \alpha x[i] (y[i] - \beta_0 - \beta_1 x[i])$$

این روش که این روش دارد این این بعنوان stochastic gradient D یا SGD نامیده شده و دوستی دارد. این روش که این روش دارد این این بعنوان stochastic gradient D یا SGD نامیده شده و دوستی دارد.

Analytical solution (Differentiating the loss function)

$$\frac{\partial L(\beta)}{\partial \beta_0} = 0, \quad \frac{\partial L(\beta)}{\partial \beta_1} = 0$$

$$\frac{\partial L(\beta)}{\partial \beta_0} = -\frac{1}{n} \sum_{i=1}^n (y[i] - \beta_0 - \beta_1 x[i]) = 0 \rightarrow$$

$$\Rightarrow \frac{1}{n} \sum_{i=0}^n y[i] + \left(\frac{1}{n} \sum_{i=0}^n \beta_0 \right) + \left(\frac{1}{n} \sum_{i=0}^n \beta_1 x[i] \right) = 0$$

$$\downarrow \quad \downarrow \quad \downarrow$$

$$-\bar{y} + \beta_0 + \beta_1 \bar{x} = 0 \quad (\bar{x}, \bar{y} \text{ is on regression line})$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \frac{1}{n} \sum_{i=0}^n x[i] (y[i] - \beta_0 - \beta_1 x[i]) = 0$$

$$\Rightarrow \bar{xy} - \beta_0 \bar{x} - \beta_1 \bar{x^2} = 0$$

$$\Rightarrow \bar{xy} - (\bar{y} - \beta_1 \bar{x}) \bar{x} - \beta_1 \bar{x^2} = 0 \Rightarrow \hat{\beta}_1 = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2}$$

نحوه دعیاً، این در مسیر دوام می‌گذرد، با این نتایج باشد، loss function، $\hat{\beta}_1$ ، $\hat{\beta}_0$ را بدستور minimize کنید.

شود، در فرم ماتریسی به صورت $(\text{معنی مسقی دوم} + \text{معنی دوم})$ توجه شوید.

ex:	i	$x[i]$	$y[i]$	$x[i]y[i]$	$x[i]^2$	$\hat{\beta}_1 = 3.1$
	1	1	4.8	4.8	1	$\hat{\beta}_0 = 11.1 - 3.1 \times 3 = 1.8$
	2	3	11.3	33.9	9	\Downarrow
	3	5	17.2	86	25	$\text{predict } \hat{y} \text{ for } [x=4]$
	$\frac{1}{n} \sum_{i=1}^n \rightarrow$		\bar{x}	\bar{y}	\bar{xy}	$1.8 + 3.1 \times 4 = 14.2$

Brief review

Linear Algebra & multivariate calculus

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \rightarrow \boxed{x \in \mathbb{R}^n}$$

i = i-th column

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} a^1 & a^2 & \cdots & a^n \end{bmatrix}$$

↓

$$\boxed{A \in \mathbb{R}^{m \times n}}$$

matrix multiplication

$$A \in \mathbb{R}^{m \times n} \quad B \in \mathbb{R}^{n \times p} \quad \Rightarrow C = AB \in \mathbb{R}^{m \times p}$$

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

$$K^T y = [x_1 \cdots x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i$$

$$Ky^T = \dots = \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & \ddots & \vdots \\ x_n y_1 & \cdots & x_n y_m \end{bmatrix}_{n \times n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{1 \times n} = \dots$$

$$y = Ax = \boxed{\begin{array}{c} A \in \mathbb{R}^{m \times n} \\ x \in \mathbb{R}^n \end{array}} \quad \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

• *Cari A gives linear combination of y*

↓
ان فرم LR بجزء دلوقت



• *Cari under ماتریس*

$$y = Ax = \begin{bmatrix} \text{values} \\ a_1 \dots a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots + a_n x_n$$

Identity matrix (جذور عددي)

$$I_n \in \mathbb{R}^{n \times n} \rightarrow I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$AI = A = IA$$

Transpose \rightarrow

$$(A^T)_{ij} = A_{ji}$$

$$(A^T)^T = A$$

$$(AB)^T = B^T A^T$$

$$(A+B)^T = A^T + B^T$$

$$\text{symmetric matrix} \rightarrow A = A^T$$

Trace

$$\text{tr } A = \sum_{i=1}^n A_{ii}$$

خط العجز

$$\text{tr}(A) = \text{tr}(A^T) \quad A \in \mathbb{R}^{n \times n}$$

$$\underline{\text{HW:}} \quad \text{tr}(AB) = \text{tr}(BA) \quad AB \text{ is square}$$

$$\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB) \quad \text{if } ABC \text{ is square}$$



Norms → measures the length of a vector

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \rightarrow \text{طول دیگر رسانید}$$

$$\begin{aligned} \|x\|_2 &= \sqrt{\sum_{i=1}^n x_i^2} \\ \|x\|_1 &= \sum_{i=1}^n |x_i| \end{aligned} \quad \left. \begin{array}{l} \text{for the outlier } \sim \text{exp} \\ \text{least absolute loss function,} \end{array} \right.$$

$$\|x\|_\infty = \max|x_i|$$

inverse : $\alpha_i x_i$

$$(AB)^{-1} = B^{-1} A^{-1} \quad A, B \in \mathbb{R}^{n \times n}$$

$$(A^{-1})^T = (A^T)^{-1}$$

برای سری دو

Given a set of vectors $(x_1, \dots, x_n, x_i \in \mathbb{R}^m)$ is said-

to be linearly dependent if for $(\alpha_1, \dots, \alpha_n) \neq 0$

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \alpha_n x_n = 0$$

ex:

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix} \quad x_3 + 2x_1 - x_2 = 0$$

لئے کہیں دو تا لفایت فی لند

rank of matrix A :

the number of linearly independent columns of A (or rows)

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \text{full rank} \rightarrow \text{rank} = 3$$

$A \in \mathbb{R}^{m \times m}$, $x \in \mathbb{R}^m$: $(\cup) \rightarrow + 6 \rightarrow 5$ Category

A is positive definite if $x^T A x > 0$ for $x \neq 0$

" " semi " " " " ≥ 0 " "

" " negative " " " " ≤ 0 " "

" " semi " " " " ≤ 0 " "

A is indefinite if x_1 and x_2 $\left\{ \begin{array}{l} x_1^T A x_1 > 0 \\ x_2^T A x_2 < 0 \end{array} \right.$

ex: $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ is PD.

$$x^T A x = [x_1 \ x_2] \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = x_1^2 + 2x_2^2 > 0 \rightarrow \text{باید } x_1 \neq 0, x_2 \neq 0 *$$

و سبیل

- در حقیقت loss function کمینه کردن یعنی $f(x)$ را برابر با 0 نمودن
- معمولاً این کار را با استفاده از gradient descent یا backpropagation
- به همراه backpropagation یا gradient descent برای آنرا تجربه کرد



Quadratic form ز.و.پ

$A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$

* The scalar value $x^T A x$ is called a quadratic form.

• $x^T A x$ is transpose

$$x^T A x = \dots = \left(\sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j \right) \quad \text{(جائز حجم، مرضي)} \quad \text{f(x) لـ SVM}$$

$$x^T A x = (x^T A x)^T = x^T A^T x = x^T \left(\frac{1}{2} A + \frac{1}{2} A^T \right) x$$

Gradient & Hessian

$f(x_1, x_2, \dots, x_n)$ is function of x_1, x_2, \dots, x_n and f is function of x

[$\beta_0, \beta_1, \dots, \beta_m$ are parameters]

suppose $F: \mathbb{R}^m \rightarrow \mathbb{R}$ (one variable function)

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix} \quad \text{مُنْقَبَةُ اولَى}$$

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \dots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix} \quad \text{مُنْقَبَةُ ثانِيَّةٍ}$$

ex: $f(x) = x_1^2 + 3x_2^2$



$$\nabla_x f(x) = \begin{bmatrix} 2x_1 \\ 6x_2 \end{bmatrix}$$

$$\nabla_x^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$$

ex 2: $f(\beta) = \mathbf{x}^T \beta = \sum_{i=1}^n x_i \beta_i = x_1 \beta_1 + x_2 \beta_2 + \dots + x_d \beta_d$

linear regression \downarrow
مُرْجِع از ج

$$\nabla_{\beta} f(\beta) = \begin{bmatrix} \frac{\partial F}{\partial \beta_1} \\ \vdots \\ \frac{\partial F}{\partial \beta_d} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} = \mathbf{x}$$

$$\nabla_{\beta}^2 f(\beta) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & - & - \\ 0 & - & 0 & - \\ 0 & - & - & 0 \end{bmatrix} = \mathbf{I}$$

ex 3: $f(\beta) = \beta^T \beta = \sum_{i=1}^2 \beta_i^2$

$$\nabla_{\beta} f = 2\beta \quad , \quad \nabla_{\beta}^2 f = 2\mathbf{I} \quad ?? \quad \text{اینست کیم؟ : HW}$$

$$\begin{bmatrix} 2\beta_1 \\ 2\beta_2 \end{bmatrix}$$

$f(\beta)$	$\nabla_{\beta} f$
$\beta^T \mathbf{x}$	\mathbf{x}
$\mathbf{x}^T \beta$	\mathbf{x}
$\beta^T \beta$	2β
$\beta^T C \beta$	$2C\beta$ (symmetric C)

آنچه در تئوری سوچی
که $\nabla_{\beta} f$ میباشد

first + second part of the