

سَقْرَمَةِ الْجَنَاحِيَّةِ

R جهات الرسم و أدوات

امتناع

لِـ ١٠٠ - لِـ ٦٣ - لِـ ٢٧

١٤٠٤

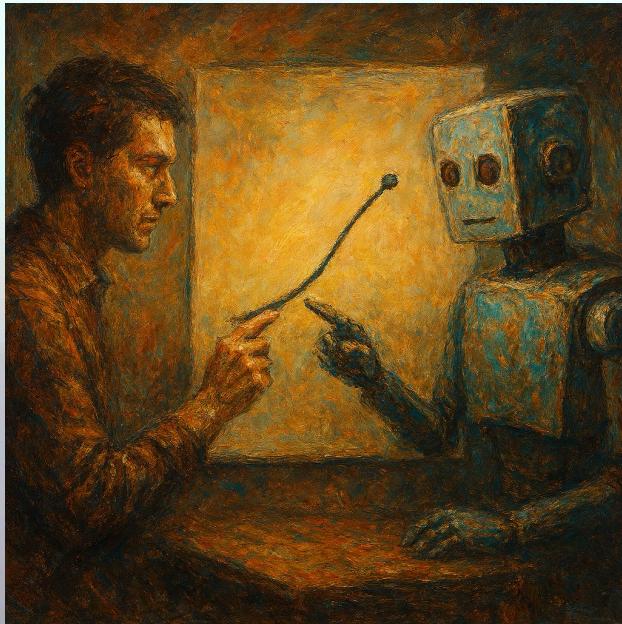
فصل اول: مقدمه

ضروری بہ یاد رکھیں گے اسکی آنکھیں :

دیکھنے والے بھی مصائب و سوء احوالوں سے درکھلے 1100 اسے دار.

Supervised Learning

ڈھانچہ سو (بیٹھارے)



Unsupervised Learning

یوناں خلائق

میں دستی اعماق

درادهای را فحی :

برای ارائه تصویری از رضی طبیعتی دارایی های مجموعه دارد و حافظی

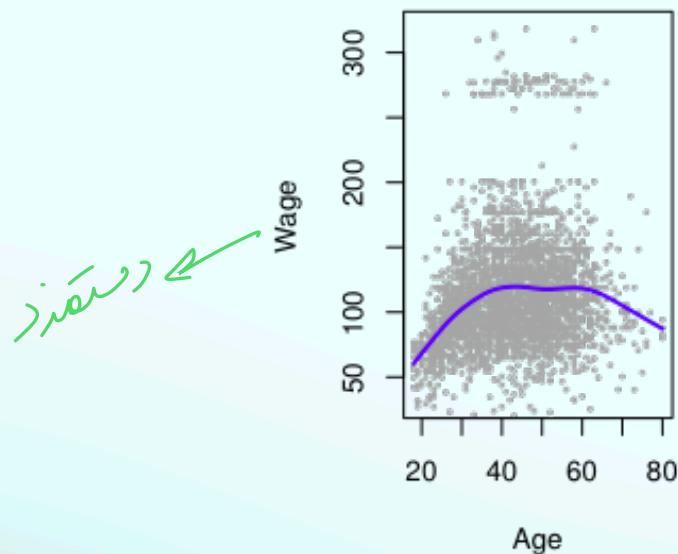
wage data

① دردهای دستمزد

درادهای مربوط به عامل مرغیت به دستمزد گزینی از مدخل سطحی اعیانوس اطلاق (ایالت تکه کری)

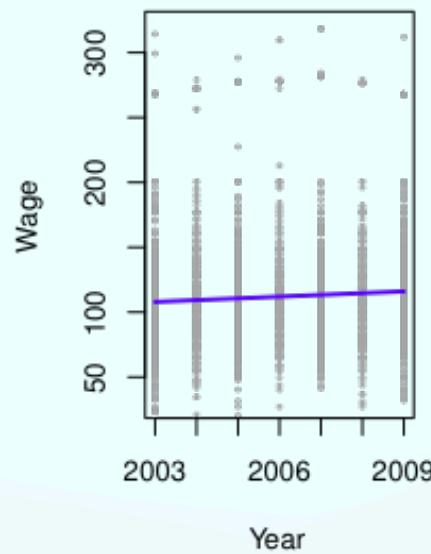
ارتباط بین سن، تحصیلات کی کارمند و همچنین سال تقویتی روی سران دستمزد

خطابی: پانزین درسته زیاد هر سن

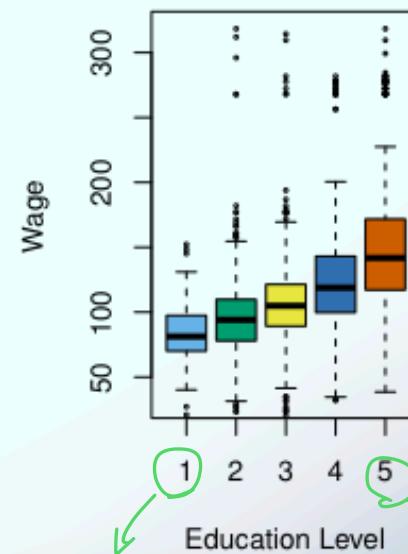


جذب (جذب)

سن



لمسات تعمیمی



جذب (جذب)

جذب (جذب)

عیواد نیزه عربت کد پیش بینی درسته به هر که این عملیات طبقه ایسے -

==

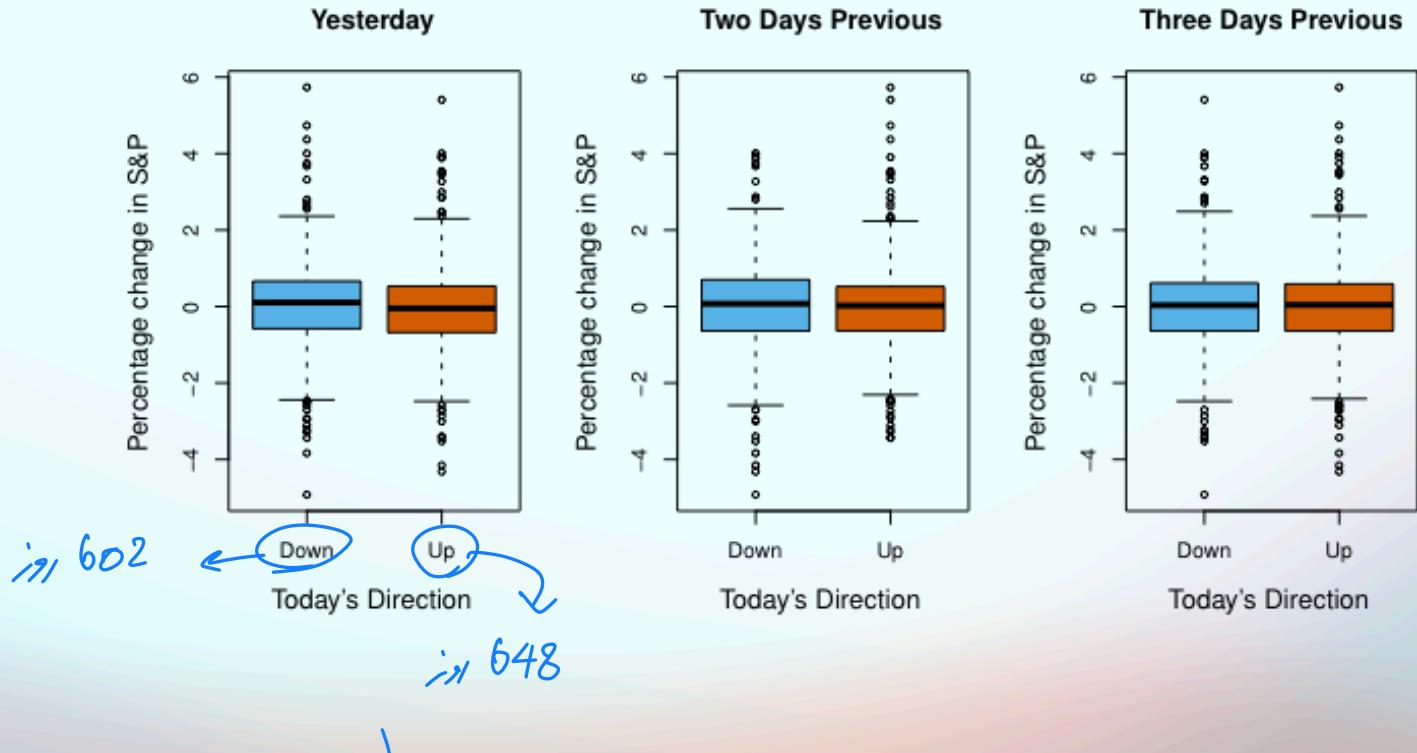
Stock market data

دارودهی بزار مکانی ②

تحتیات روزانه سهام S&P ۵۰۰ دریل دوره ۵ ساله بنی سلیمان

می خواهیم بینیم که انتشاره از درصد تغییرات روزگارنده کدام سهام در روز معین افزایش یا کاهش می خورد؟

در این حالت باید مقدار طبقه بندی روی روش تخمین زنی هدف پیشینی کسی مقدار عددی داشت



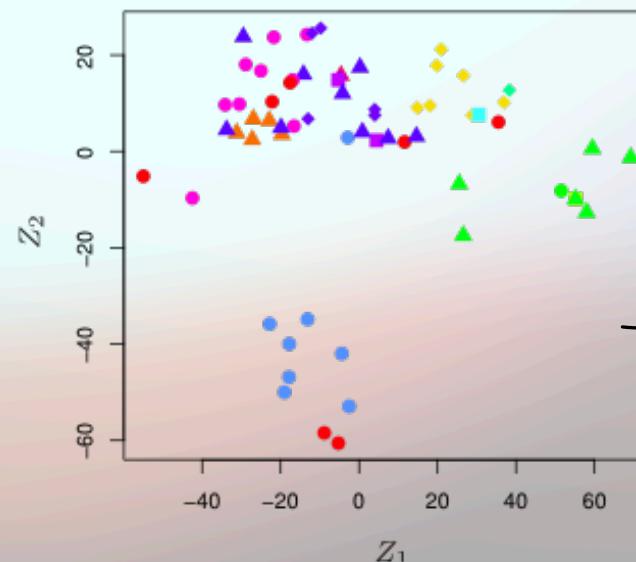
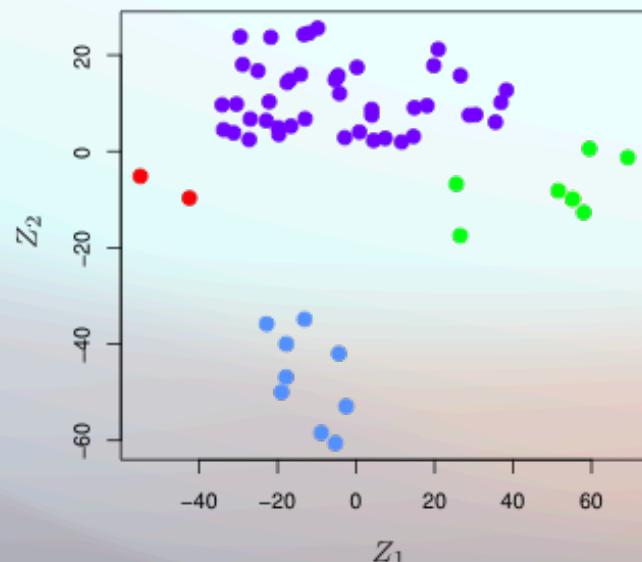
امروزه بازار اسلامی
میتواند در این سطح قرار گیرد

Gen expression data

داده‌های بیولوژی ③

اطراعات 6830 اندازه‌گیری بیولوژیکی برای 64 گروه سلول سرطانی است.

براساس اندماخته شدن داده‌ها بیولوژیکی، گروه‌ها به چه فرم و ترتیب دارند؟

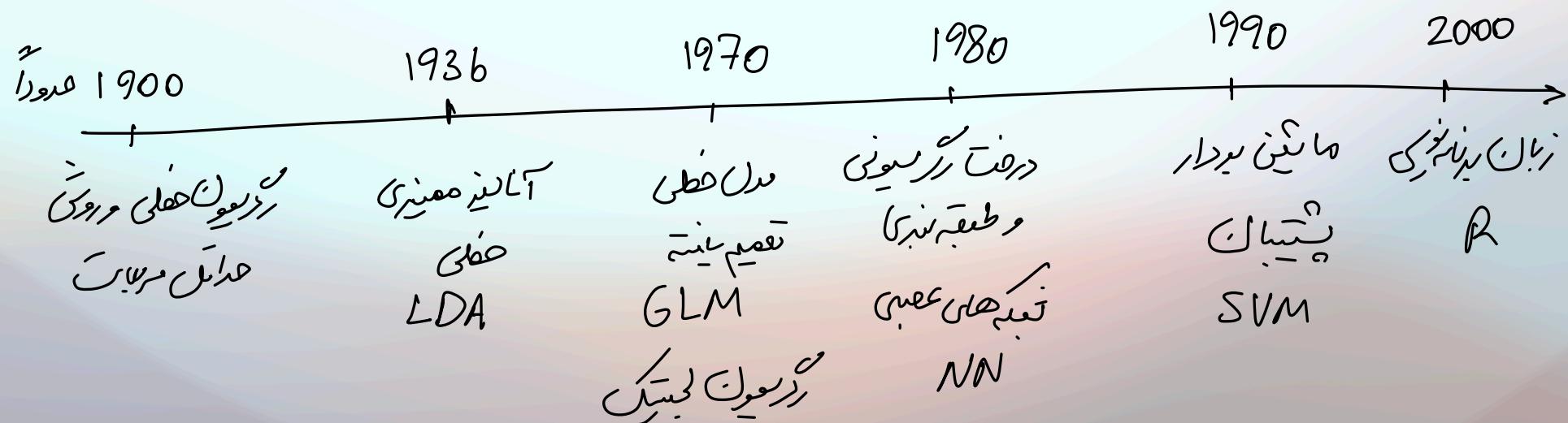


هر گروه ۱۴ نمونه دارد
نمایش تسلیم

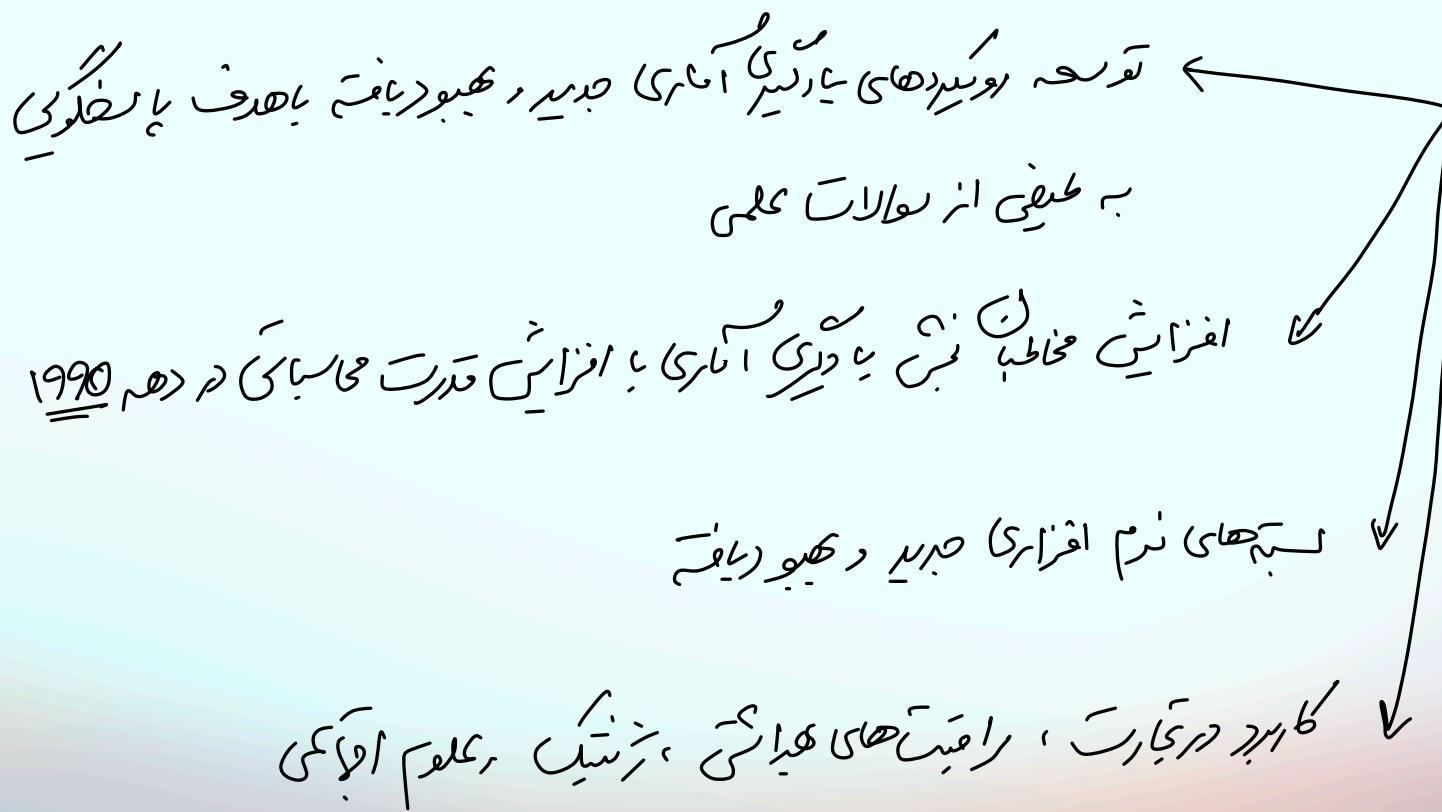
۱) ریکیه مختص (از تاریخ آماری):

آرچه / اصطلاح عده‌ای آماری نسبتی بودن از مفهومی به زیرینی آن مفهوم را که جمله‌ای

یافته بودند.



پیروزی دنیاگیری آغاز



لین دری چه افرادی صدایت؟

تمکنی علاقهمند به اتفاقات از روش‌های آماری برای برای مدل سازی و پیش‌بینی داده‌ها هستند

دانشمندان علم پژوهی، هنرمندان، تحلیلگران داده، دانشمندان علم داروی و کارشناسان علم آماری

علات تزاریخی : Notation

$n \leftarrow$ تعداد داده (ساخته)

$P \leftarrow$ تعداد متغیرها که برای پیش‌بینی در دلخواه هستند

wage $\begin{cases} n = 3000 \\ P = 11 \end{cases} \rightarrow \text{year, age, race, ...}$

در برخی از مطالعات، مانند داروهای بیولوژیکی پیشرفت به راردهای تسلیفاتی مبنی بر وابستگی مولفه‌است

P بیانگر بگو (هزاراً اینها می‌باشند !)

$$x_{ij} \quad i=1, 2, \dots, n \\ j=1, 2, \dots, p$$

مقدار متغیر زیبایی مساهه

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & & \ddots & \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

ن بیایی های اندازه (اندی) معنی های مساهه
ن بیایی های اندازه (اندی) معنی های مساهه

X کی ماتریس $n \times p$ ایت کے عینک (ز، ن) مساهه

گھی اوقات (لفی های) ماتریس X را بصورت x_1, x_2, \dots, x_n کے x_i میں نویس کر x_i میں نویس کر
بدر پ چیز کے عامل اندازه لیکھی P متغیر بیایی مساهه نام ایت

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}_{p \times 1}$$

نتیجه: بیانی داده های رسمت بر این فرضی سعی نماینده دارای عی رفته.

سل: برای داده های رسمت x_i که بیانی دار ب طول $P=11$ است که علی
من، سل، نر، در رفتاری دستبر برای فرد زاده ای است.

داتا موضعی \sim نعمتی داده های ماتریس X علاوه همچنین که به صورت سال در طول زمان

$$x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

و هر بیانی دار طول n دارد

سل: برای داده های رسمت x_i سل $n=3000$ ساله برای سه سال است.

با استفاده از این همایش ماتریس X را به صورت زیرنویس

$$X = (x_1 \ x_2 \ \dots \ x_p)$$

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$$

L

علامت ترکیبی ماتریس (transpose) را در همه جایی بخواهیم + علامت عبارت از عناصر ماتریس را برداشت.

$$X^T = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix}_{p \times n}$$

$$x_i^T = (x_{i1} \ x_{i2} \ \dots \ x_{ip})$$

$y_i \rightarrow$ نامن مساهه معمکن که می خواهیم پیدا بینی کنیم

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \rightarrow \text{محتمل نمودار برای مساهه به معرفت}$$

$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \rightarrow$ داده های مساهه

↓
برآورده

$P = \sum_{i=1}^n \alpha_i \underline{\underline{x}_i}$
باشد آن ماتریس x_i یک استاراست
↓
scalar

$\alpha \in \mathbb{R}$

$\alpha \in \mathbb{R}^n$

$A \in \mathbb{R}^{r \times s}$

برای فتح این در مورد اینجا، بردار، ماتریس $\underline{\underline{x}}$ $\underline{\underline{\alpha}}$

لی $B \in R^{d \times s}$, $A \in R^{r \times d}$ ضرب $(A \cdot B)$ ماتریسی است که $r \times s$ ماتریسی است و نتیجه حاصل از ضرب A و B است.

نمایش $(AB)_{ij}$ این ماتریس از راهله زیر میگیرد:

$$(AB)_{ij} = \sum_{k=1}^d a_{ik} b_{kj}.$$

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}.$$

Then

$$\mathbf{AB} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1 \times 5 + 2 \times 7 & 1 \times 6 + 2 \times 8 \\ 3 \times 5 + 4 \times 7 & 3 \times 6 + 4 \times 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}.$$

: پس

دراوهای اتحاده نوو در Lab (آزمایشگاه) و هرینک

مجموعه داده این داروهای کاربردی در حوزه بیماری‌ها، مانع، زیست شناسی، ... (ارزش

تعدادی R، همچنین بعضی از آنها در نتایج ای را، ISLR2 به

> install.packages ("ISLR2")

> library (ISLR2)

> data (Package = "ISLR2")

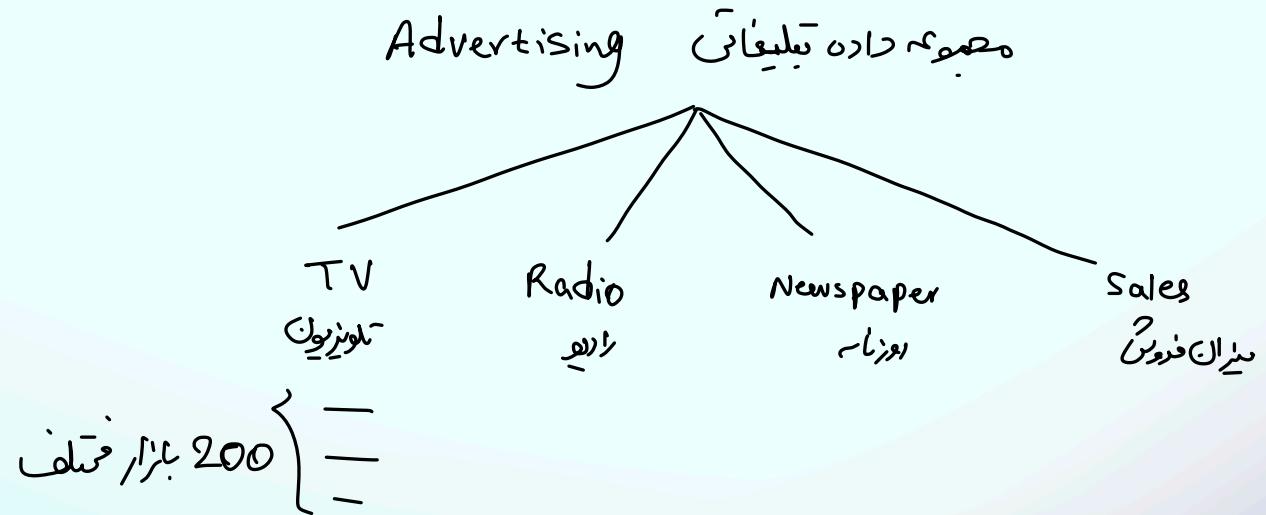
Name	Description
Auto	Gas mileage, horsepower, and other information for cars.
Bikeshare	Hourly usage of a bike sharing program in Washington, DC.
Boston	Housing values and other information about Boston census tracts.
BrainCancer	Survival times for patients diagnosed with brain cancer.
Caravan	Information about individuals offered caravan insurance.
Carseats	Information about car seat sales in 400 stores.
College	Demographic characteristics, tuition, and more for USA colleges.
Credit	Information about credit card debt for 400 customers.
Default	Customer default records for a credit card company.
Fund	Returns of 2,000 hedge fund managers over 50 months.
Hitters	Records and salaries for baseball players.
Khan	Gene expression measurements for four cancer types.
NCI60	Gene expression measurements for 64 cancer cell lines.
NYSE	Returns, volatility, and volume for the New York Stock Exchange.
OJ	Sales information for Citrus Hill and Minute Maid orange juice.
Portfolio	Past values of financial assets, for use in portfolio allocation.
Publication	Time to publication for 244 clinical trials.
Smarket	Daily percentage returns for S&P 500 over a 5-year period.
USArrests	Crime statistics per 100,000 residents in 50 states of USA. ← base R
Wage	Income survey data for men in central Atlantic region of USA.
Weekly	1,089 weekly stock market returns for 21 years.

www.statlearning.com

→ Book Website

فصل دوم: یادگیری آماری چیست؟

سال : فُوقِ کسی نہیں مساعر آماری ہتھیں کے توجہ کا فرمایا انتظام نہ ایم تا اربیط بین تبیغات و فروں کے محصول اپریل کی نیشن.



کارخانیوں کی تبلیغات را درج کر کر رکن نہیں تعریل کند.

نہ برائی اور سٹھنی دھمکیں بین تبلیغات و فروں ارکیٹ و صدور درمی توانیں کوئی تنظیم پورچہ تبلیغات را
بے کارخانی تھیں کیم تا حصہ من میزان فروش را داشتے ہیں۔

متغیرهای پیش‌بینی شده

Predictors Variables

متغیرهای مستقل

independent ~

ویژگی‌ها

features

متغیرها

Variables

متغیر پاسخ

response

متغیر وابسته

dependent

X_1

بروکم تلوزیون

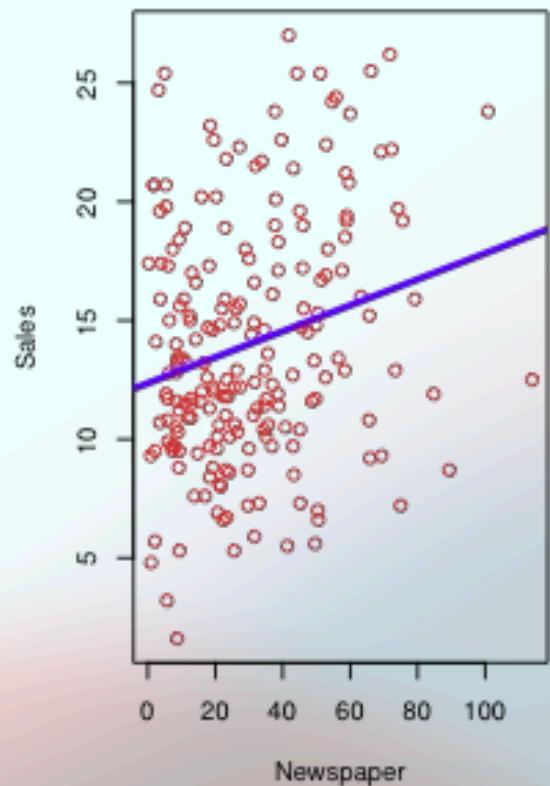
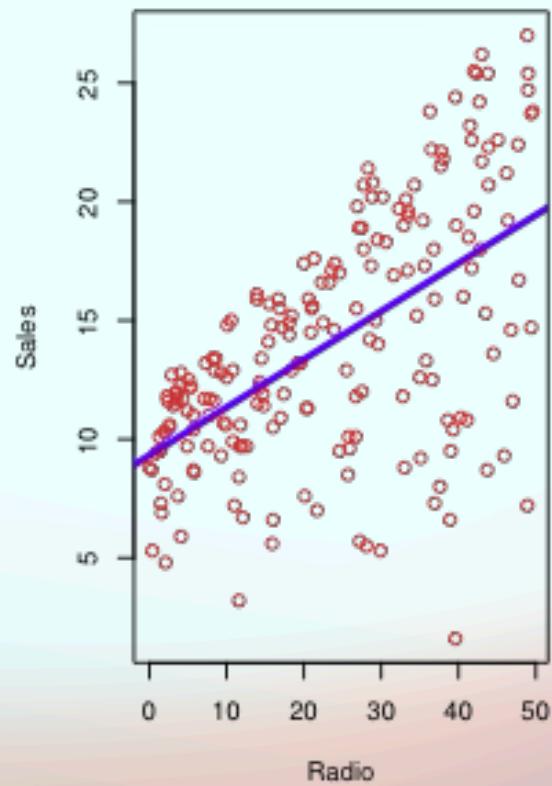
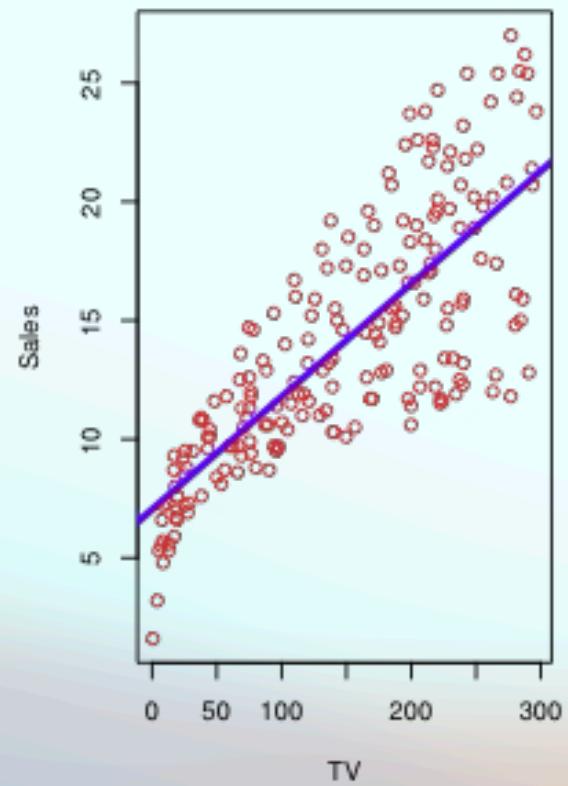
Y میزان فروخت

X_2

بودجه کارو

X_3

بودجه خانه



ب) معورت کلی تر :

فرضی کنیم که پاسخ کسی λ ، θ پیش بینی شده فنت x_1, x_2, \dots, x_p را تصور کنیم. بافرض وجود رابطه بین λ و $x = (x_1, \dots, x_p)$ را بتوان نوشت:

$$Y = f(x) + \varepsilon$$

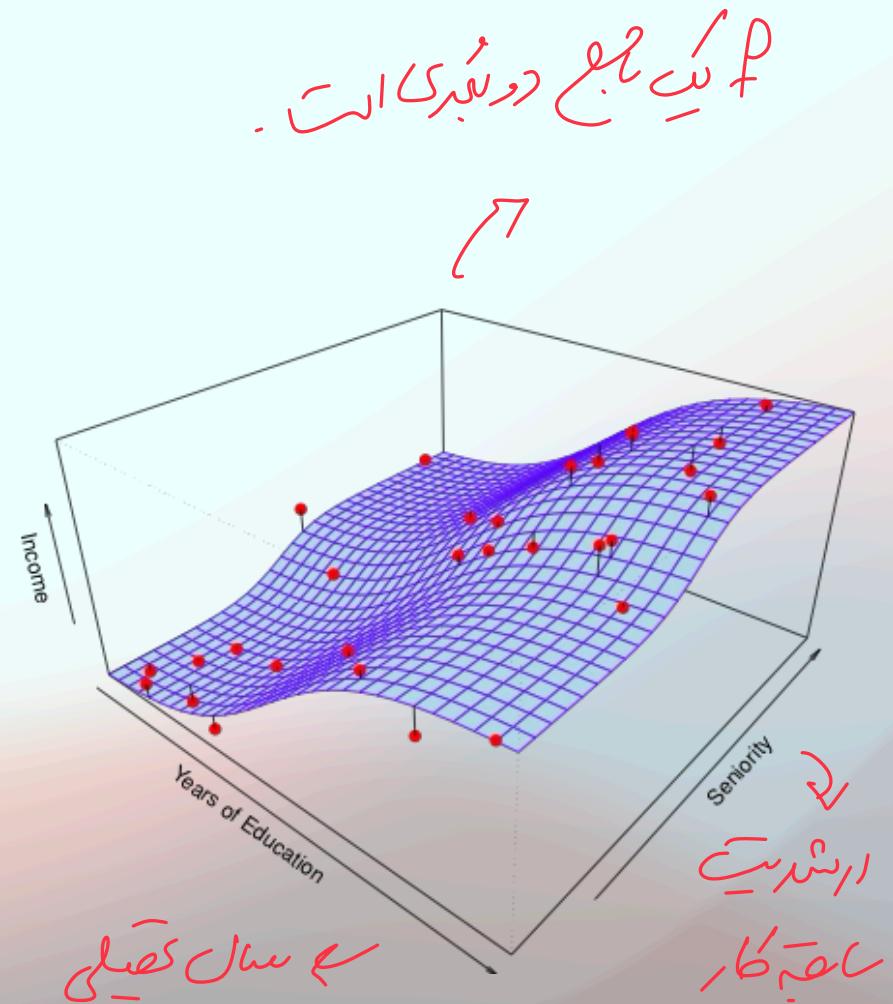
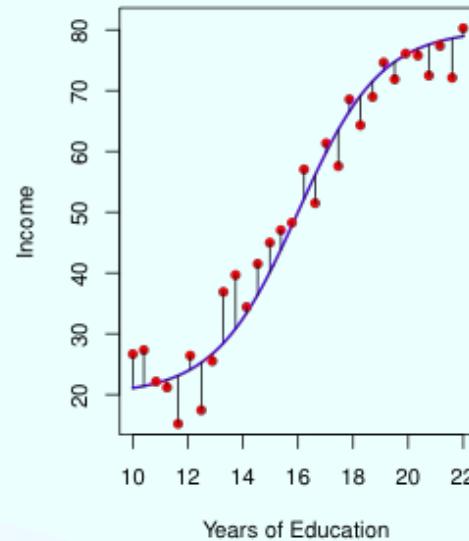
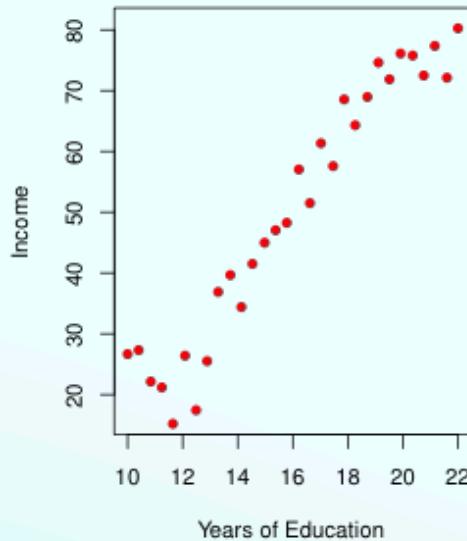
f : تابع ماب آما محول از x_1, \dots, x_p

تمام اطلاعات مبتنی بر اینکه x در مورد Y ارائه می‌گیرد.

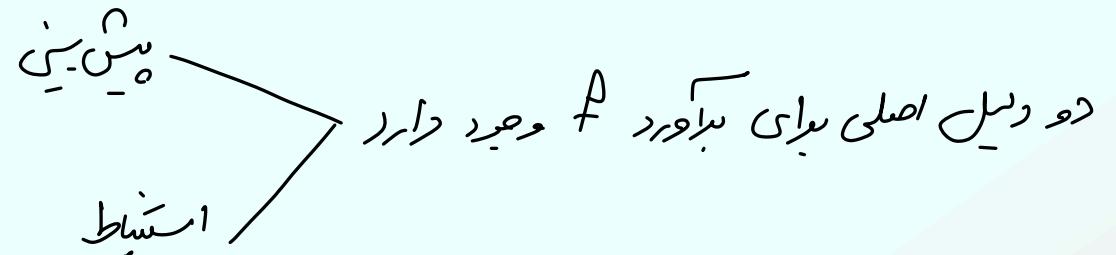
ε : خطای صادفی \leftarrow سفل از x و دلایی می‌باشد که نتیجه صفر است.

سازنده احتمال پنهانی ای از روش‌ها برای برآورد f اسماهه دارد

٥٠٢٣٥



چرا f را ببرآورده نمی‌خواهیم؟



Prediction پیش‌بینی

نهایی اوقات مجموعه‌ای از دیوای خود X به راهی درست می‌شوند اما حرفی Z را نتوان برآورده نداشت اورده.

عمل: فرضی کنید x_1, x_2, \dots, x_p مسیرهای مختلف بیان متن Z به سادگی در آرایه \mathbf{x} می‌باشد.

آخر Z خطر ماسنگ نمی‌بیند، بگویی دارویی خاص ندارد، ملکیتی ایست به دنبال پیش‌بینی Z ، انتشاره از X بینی تا از دارویی مرتضی به عینی ران در معرض خطر بیان و آنست نامطلوب، احتمال نمی‌گیریم.

از آنجایی که معنایلین عبارت خط (ع) به صفر می‌رسد، می‌توان ۲ را به معورت زیرین برآورد کرد

$$\hat{Y} = \hat{f}(X)$$

که پسین بینی \hat{Y}

برابر f

نکته: در این حالت \hat{f} به عنوان یک جعبه سیاه (Black Box) در نظر گرفته می‌شود، زیرا آنچه که پسین بینی های

دقیقی برای ۲ ارائه شود، بدلان خاصی در مواردشان دقیق \hat{f} وجود ندارد.

دقیقی محض مقدار دقیق پسین بینی است و احتمال ندارد که \hat{f}

بی هم تسلیل باشد.

دقت پیش‌بینی :

بطرکلی، \hat{f} برآوردهای بیانی f نت و این معن دقت، مقداری خطاب ایجاد می‌نماید.

آنچه این خطاب کاهش است؟ بل - متوالی دقت \hat{f} را با انتشاره از منابع ترین تلسیون یادگیری آنرا
بیانی برآورده f افزایش دهم

در همین حالت $f(x) = \hat{f}(x)$ باز هم سینی مادرای مقداری خطاب خواهد بود ای زیرا \hat{f} تابع از عوامل
و طبق تعریف نمی‌توان آن را با ارتفاع از x سینی نمی‌کرد.

نیز برای تغییرات مرتبه \hat{f} را روی دقت پیش‌بینی‌های ماتمای عرضه نماید

دقت $\hat{\wedge}$ به عنوان میکنیم برای Z به واسیع نگاه دارو

خطای تحلیل ناپذیر

خطای λ از سینه پذیری \Leftarrow به عنوان خطای تحلیل ناپذیر خاصه میگردد \rightarrow زیاد همیست همچو f را فرموده باشیم، ما من تعالیم خطای ناشی از λ را کاوش میکنیم.

حاوی متغیرهای اندازه سنجی سند طایی بگردید و میکنیم Z سینه نداشته باشد

حاوی تغییرات غیر مابین اندازه پذیری باشد

نهیت ∞

برآورده داره \hat{f} و مجموعه ای از پیش‌بینی‌سنهای X را در نظرم بگیرم

$$\hat{Y} = \hat{f}(X)$$

فرضی کنید که \hat{f} و X تابع باشد به طوری که تغییر نیزی فقط از ϵ حاصل نمود

$$\begin{aligned} E(Y - \hat{Y})^2 &= E[f(X) + \epsilon - \hat{f}(X)]^2 \\ &= \underbrace{[f(X) - \hat{f}(X)]^2}_{\text{Reducible}} + \underbrace{\text{Var}(\epsilon)}_{\text{Irreducible}}, \end{aligned} \quad (2.3)$$

تغییر نیزی

تغییر ناپذیر

لهرنگ این درس برای رسیدن اهداف خود f با هدف به حداقل رساندن خطای تغییر نیزی است