

کنیز حسنه است → سیان آزم ۲۵-۲۰

۷-۳۰-۴۰ بیان رسم

۷-۲۰-۲۰ غریب

۷-۱۵-۲۰ پروژه

«دانش و تفاضل اینه است» ← حضورا درست

در بخش اولیه ها → یادگیری عین توصیه می شود (یادگیری تئوری نظریه های در کسی شون)

پیشواز ← بیان پایتوں ، سایس (برانی پیشواز)

پیغمه ← ساده سازی کی تعلیم (ایمیل - نویسه و ...)

Yaser S. Abu-Mostafa et al. →

منیم های آموزشی حسنه دارد

ارائه های خوبی دارد

آن خوبی دارد. دلیل بیان است ، نظریه notation ، خاص خود را دارد.

Kevin Murphy →

عیت های آماری خوبی دارد

Bradley Efron, Trevor Hastie

استاد اند

Bishop (یادگیری دار)

آن خوب است

Tom Mitchell

دانشگاه بیان کرده . آن بکار گیری ، بیان ساده ، خوب دلیل آموزت شده

امکانلا روش تفہیم از این دست - پرداخته می شود.

«Machine learning is nothing but prediction»

(inductive) learning → یادگیری استقرای

deductive یادگیری

International Conference Machine Learning

ICML

از سه بزرگترین نشریه های یادگیری ماشین → ICML ← UCI

داده های test و train و ... تا می توانیم برخورد کنیم .



بایگانی می‌شوند و این داده‌ها را برای استخراج می‌کنند.

## Spark + Hadoop

در داده طریق ارزش‌های آماری با یارکوبی ماسن (ستواره سیزده

think map تکنیک

Alice → chat bot → Turing Test ↗ JGML

## Turing Test

LaTeX editor  $\infty$   $\rightarrow$  Turing Award

Herbert Simon 1970      Tom Mitchell 1990      Ether Alpaydin 2010      Shai Ben-David 2014 }      ML isler

$$9V, 4, 2V : \underline{4} \text{ mls}$$

برین، DNA ریز، عقیق  $\leftarrow$  Skinner's lab

IBM Watson ← سہیل بروئے یادیو مائن ← ہنل دسیار نیٹک ہم استناد سی سردار

کے درمیان پرتو : Deep Blue : نسبت انسان در رازی سمع



(بررسی) regression  $\leftarrow$  ۴۰ : age  $\leftarrow$  seeing AI  
 classification  $\leftarrow$  female / male  $\leftarrow$   
 "  $\leftarrow$  happy / surprised / ...  $\leftarrow$

یاد رفته IBM Watson  $\leftarrow$  NAO  $\leftarrow$

سایر ربات ها  $\leftarrow$  CHEETAH , PEPPER , ASIMO , ROBI  
 بینی اسب

مساعده کی علیم آن ترمینی - یا شور

shakthydoss.com

stats.stackexchange.com

stackoverflow.com

اللوریسم های پادلریکی (پلات)  $\leftarrow$  ... / real / خوب یا نیزه ...  
 درست راهنمایی می کرد .

peter b. (پلر) supervised

discretization

classification / Graviton : خوب نامه

ولماز  $\leftarrow$  سوسته

تمار آلتندل  $\leftarrow$  دسته

6AM 12AM 6PM

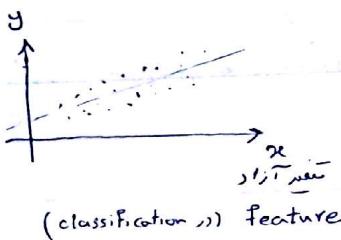
on target

early morning

شیخیت مسیریت : میان ، آمد ، ذرا ، حذینه های بخارت ها ، از ریز سیزه  
 سود ، هفت خانه

regression toward the mean : میانگین برگشتی  $\leftarrow$  دنواه قد شده است .  
 کسی در دست کم در تحریک (شدید) در دور بعدی خفف شد ، ایست

Francis Galton  $\leftarrow$  میانگین روزیون



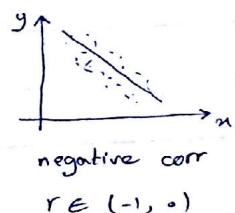
positive correlation

$r \in (0, 1)$

ضریب همبستگی (r) : ضریب همبستگی (r)

$$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

افزون سیار

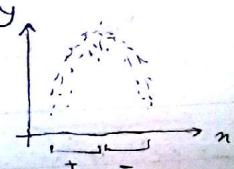


negative corr  
 $r \in (-1, 0)$

strong corr.



弱یزی خوب  
 $r \approx 0$



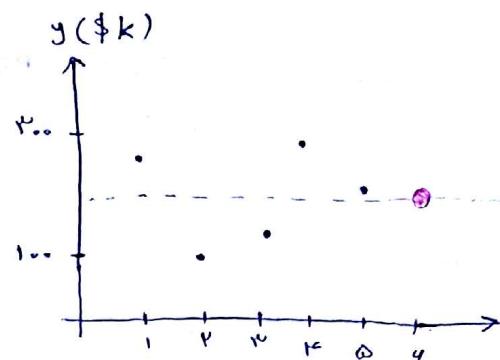
جلسہ ۳ ۹۷، نومبر

مثال: پیش نمایتیت خانہ  $\rightarrow$  سندہ ریزیں  $\rightarrow$  بانتظارت  $\rightarrow$  نئی طبیب سندہ داری داریں (دادھاکی آنریز)

برحسب دار داریں

id	price (\$k)
۱	۲۴۰
۲	۱۰۰
۳	۱۶۰
۴	۲۹۰
۵	۲۱۰

$$F_{\text{all}} = ?$$



id دیگری سنت صفتیں تسلیم کرے اسے

$$\begin{aligned} \text{avg} &= 200 \\ \rightarrow y_4 &= \frac{1}{5} \times 200 = 200 \end{aligned}$$

اعمال در جیسے باری دیگری shuffle کی جائے۔ دادھاکی آنریز، یعنی bias یا error یا شرود (تکمیلی) کو سرد کرنا (کوئی ایجاد نہ کرنا)

$$\text{residual}_i = \bar{y} - y_i$$

خطا در سندہ نویسی میں residual error کوئی شرود

عہد حاکمی آنریز  $\leftarrow$  statistical machine learning

متغیر مستقل (آثار)  $\leftarrow$  attribute or feature

متغیر معمول (دادھاکی)  $\leftarrow$  (training set)

	۰ (حکم فتنہ) $\times$ (متغیر)
۲۰۰	۱۰۸
۱۰۴۰	۲۴۰
۵۸۲	۱۲۰
۸۲۰	۱۴۰
:	:

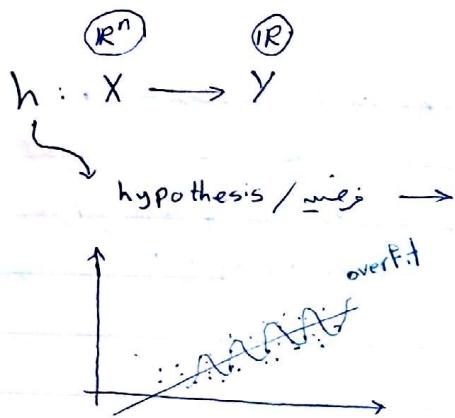
$$m : \text{قدور دادھاکی آنریز}$$

univariate  $\rightarrow$  حرداہ دی آنریز ہے جس کے

( $(\bar{x}, y)$ )  $\rightarrow$  multivariate  
 feature vector

کوئی متغیر مستقل داری





پریسی داده های داریم  $\rightarrow$  پریسی داده های داریم  $\rightarrow$  regression

که چنین است از نتایج داده های  $X$  و  $Y$   $\rightarrow$  map  $\rightarrow$  (بریسی خطا، رخت تعمیم، SVM, ...)

ماتریس  $h$  در منظمه پیش بینی نسبت خانه: (بریسی خطا)

اصل اوکام کا راز: Occam's razor

کے حوصلہ، جنگلی (فناوری)  $\rightarrow$  دری بر زیری (مشکل)

Simplicity is the ultimate sophistication

از دنونادر داده هایی

پریسی ماشین: مدل های ساده تر، مطابق تر هستند.

$$h(x) = \theta_0 + \theta_1 x \rightarrow$$

هدف پارامتری پارامتری  $\theta_i$  است

$\hookrightarrow$  نسبت ساده پارامتری است (بنیان گذری داده ها را بسته است) (بریسی بیان نهاد)

هدف یافتن تابع است که همین کرازن را به داده ها بدهد.

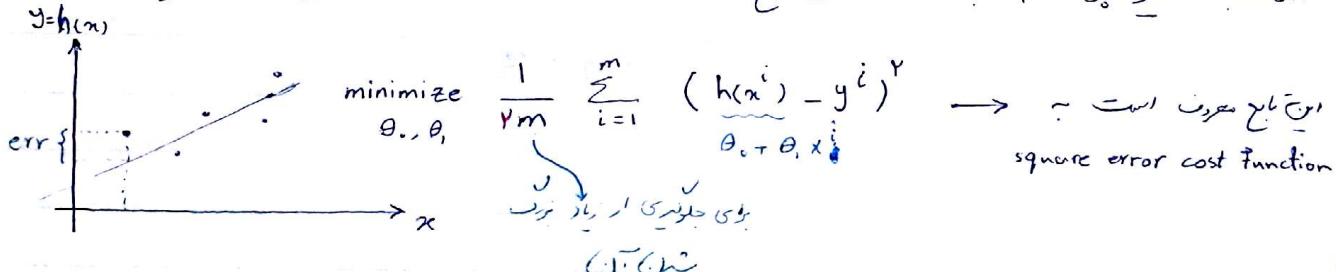
پریسی خطا  $\rightarrow$  تدریج تعمیم دهنده باشد

تدریج پیچیده تر عمل است خطای آشنای ترین پارامترها را باز بگیرد. داده ها را خنثی کند

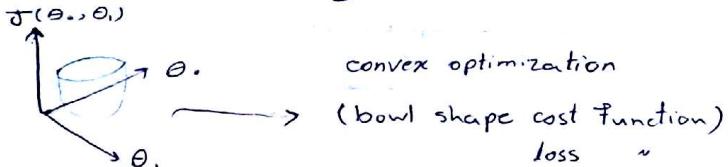
learning vs. memorization

$\hookrightarrow$  over F.T

آنما تعاریف نیز هست که cost Function  $\rightarrow$  objective function  $\rightarrow$   $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

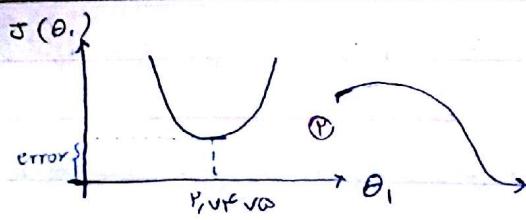


اصلیه در  $\rightarrow$  دریج برای ساده شدن دریج در اینجا  $\leftarrow$  mathematical convenience



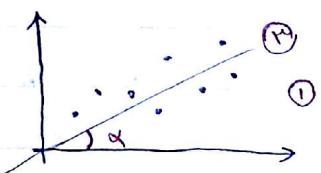
convex optimization

(bowl shape cost Function)  
loss ~



\* میتوانیم  $\theta_0 = 0$  را می‌خواهیم تصور کنیم

برای تقریبی تابع  $J(\theta)$  حدوداً همچنین است  
(تاج درجه ۲،  $\theta > 0$ ) احتمال غفارشان آن بسیار است  
چون  $\theta$ -های متساوی توزیع کامل ندارند برای  $\theta_0$  کمی خط است  
و overfitting نیز می‌شود



$$h(m) = \theta_0 + \theta_1 x \\ \Rightarrow \theta_1 = \tan \alpha$$

$$y = \theta_0 + \theta_1 x \rightarrow$$

خط با سرعتی باریش (ترین خط)

جلسه ۱۴ : ۷، ۸، ۹

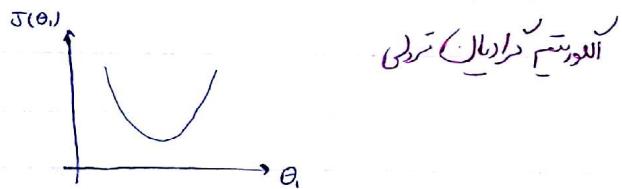
سوال: حدیده پارامترهاست  $\theta_i$  (از تابع خروجی) را سیاست دینم.

computer science  $\theta_i \leftarrow$  (Gradient descent)  $\rightarrow$  درایل تردی  
فرینتیاری: حل معادله معادله جریح ساخته (یعنی)  
دو تیرکی حریف سوال نظر

- Initialization

- repeat until convergence

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \dots, \theta_n) \quad j=0 \dots n$$



$$\boxed{\theta_0 = 0 \text{ با خوبی}} \quad \theta_1 \leftarrow \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1) = \theta_1 - \alpha \frac{d}{d\theta_1} \left( \frac{1}{m} \sum_{i=1}^m (h(x^i) - y^i)^2 \right) \\ = \theta_1 - \alpha \left( \frac{1}{m} \sum_{i=1}^m (h(x^i) - y^i) x^i \right) \quad (\alpha > 0) \\ \downarrow \text{learning rate} \quad \text{خوبی از دیگری}$$

نکات و نتایج:

- آیا تعداد ایتمی تائیری در حیث نهایی طرزی با خوبی؟

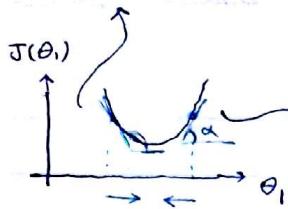
معلم است  $\alpha$  را  $\min$  علی سمعی  $\rightarrow$  نیز نوروسی باشد  
randomization and T-norm بازگشتی

استفاده از تجربیات دسته است  
randomization exploration  
کنترل فعالیت برای خوار از  $\min$   $\rightarrow$   $\max$   
randomization exploration or S-norm



نمایه از داده را نماین تا بزرگ آر است در جایب های ناین ندارد.

$$\tan < 0 \Rightarrow \theta'_1 > \theta_1$$



heuristic

خوب هست

$$\tan > 0 \Rightarrow \theta'_1 < \theta_1$$

برگشت

نماین داده کار نمایند.

نماین داده کار نمایند.

تغییر (نحوی): (آنچه می خواهیم)

که سطح - مناسن

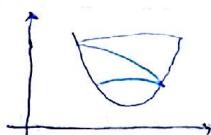
جیش انسان حدود ۲ بیلیون عصر در ۳۰ سال اول زندگانی داشتند. پس از ۲۰ سال < جیش > نیز

نماین داده بزرگ و خوب به طبقه تریک کرد، نماین داده بزرگ شود.  $\Delta = 0.01 \Delta$

ضد و درست: طبل

نحوی

خوب



با همان حس بجزب می رسیم وی لذت را است.

آخر و خوب بزرگ باشد میان است بجزب نیز. (جزب نیز miss یعنی)

دروزه خوب نیز - خوب میگوید (یعنی بخوبیست)

چه ترین از مدل استاده میگردد؟ univariate linear Regression

$|r| > 0.8$ : مابین بزرگ ← قدر رانه باشد ← قدر رانه باشد

نماین داده میگردد

داده های آموزشی Training Data

داده های آزمون sort: بزرگ داده ها است

Percentage split (44%)

برای این داده های میانه shuffling داشته باشیم  
قبل از زدن

size	price
0.00	118
1.04	133
9.88	148
10.90	247
...	...
15.84	390
19.82	403

Training data 44-50%.

test data 30-35%

برای این داده های میانه shuffling داشته باشیم  
قبل از زدن

دش خواهد بود (منا دری) است. ← خوش برخواند (یعنی بزرگ شود با هم) نیز است.

n-fold cross validation

10

در جایه ای است

### پارسیون خود چهارم (Polynomial Regression ۱) Univariate Nonlinear Regression

$$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_d x^d = \sum_{i=0}^d \theta_i x^i$$

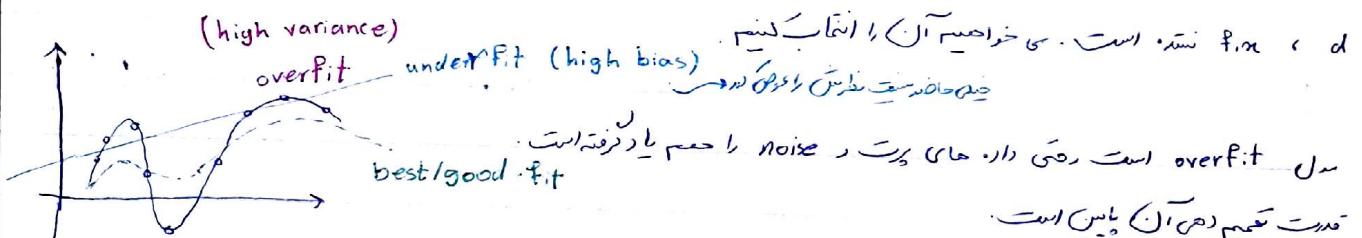
size | Price

10yo	¥45
8yo	¥32
10yo	¥90
6yo	¥11
14yo	¥0.5
:	:
10yo	¥15
9yo	¥18

عدد جانشاد ای اس ت  
training (yo.) →  $\rightarrow \theta_i$  پارامترین (Parameter learning)

validation (yo.) →  $\rightarrow$  پارامترین (چهارم خود چهارم) (model selection)  
(d)

test (yo.) →  $\rightarrow$  سنت سنت سنت overfit نشود → بازیابی نهادن / test / overfit - مارکی سنت



learning vs. memorization

اگر داده های یادی داشتیم در مقایسه ای underfit، overfit و training و validation را بتوانیم ایجاد کنیم

خطای بیان training و test

	training	validation	model selection
$d=1$	$h(x) = \theta_0 + \theta_1 x \rightarrow \bar{\theta}_1 = [\theta_0, \theta_1]$	$\rightarrow J_v(\bar{\theta}_1)$	$\left. \begin{array}{l} \text{linear} \\ \text{quadratic} \\ \text{cubic} \end{array} \right\}$
$d=2$	$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2 \rightarrow \bar{\theta}_2 = [\theta_0, \theta_1, \theta_2] \rightarrow J_v(\bar{\theta}_2)$		
$d=3$	$\rightarrow \bar{\theta}_3 = [\theta_0, \theta_1, \theta_2, \theta_3] \rightarrow J_v(\bar{\theta}_3)$		
$d=4$			

$$d = \arg \min_d (J_v(\bar{\theta}_d))$$



عیلر مدل از  $d=0$  خوب است با توجه به داده های training می شود بسیار خطای آن کم باشد و  $d=1$  خوب نیست با توجه به داده های validation می شود.

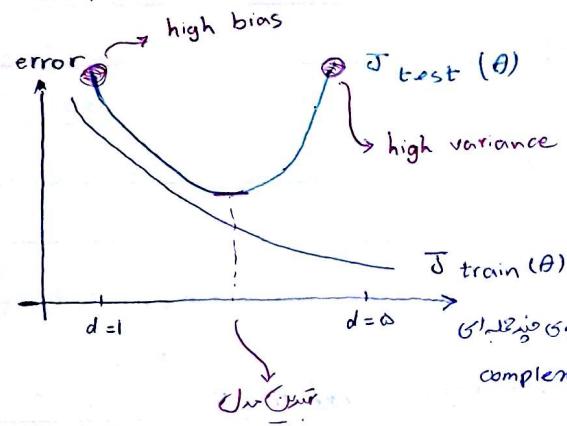
این مسابقه است

دروی

سل

$\rightarrow$  Netflix  $\rightarrow$  You are limited to one submission per day & three total

از داده های تست غیربرانش - خلاصه ناچار استفاده ننماید.



علیه تقدیرت تضمین (هم) پاسی دار نه خطای تست آن را را است.

برای سایر داده های یادداشت  
می شون

\* high bias  $\rightarrow$   $J_{train}(\theta) \uparrow$   
 $J_{validation}(\theta) \uparrow$   
(test)

\* high var.  $\rightarrow$   $J_{train}(\theta) \downarrow$   
 $J_{val.}(\theta) \uparrow$   
test

Bias تعاریف

① تعریف  
bias: likelihood a learner changes its mind  
doesn't

Preference غایل

regularization تل

② تعریف: how to push an algorithm in one direction, which is imposed from outside.

آن مدل را کوئی سوچیده از داده های پیشینه اختبار نماید.

bias نمی تواند چنین بگیرد

مثال: اگر بخواهد تبعی الدرست ها عمل نماید

margin نماین (SVM)

با میان فاصله در بین داده های training مدل انتقال را



غیر معمولی آنلود شده است. تحریل: دو حفته بعد

ایجاد تابع در مدل مسحیت نمودیم.

size(n)	Price(Y)
	↓ یکی

$x_1, x_2, \dots, x_n$	Y
------------------------	---

features  
 (x<sub>1</sub>, ..., x<sub>n</sub>)  
 ابعاد (تعداد ویژگی ها) = n

$$h(x) = \theta_0 + \theta_1 x$$

$$h: x \rightarrow y$$

$$\mathbb{R}^n \rightarrow \mathbb{R}$$

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

$$h(\bar{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

Normal Equation برای multivariate linear regression

(x<sup>n</sup> در اینجا مفهوم داده شده است)

$x_t \leftarrow x^t$  : observation  
 polynomial feature

ویژگی های خوب

(1) داده ها را خوب بایان کند

(2) توسط تخصصی حرفه ای ساخته شوند. (برای داده خوب را معرفی کنید)

(3) deep learning یا automated feature selection

Android --- (edges) توصیه: شاهد اثاثات  
 self learning

Computers teach themselves to Recognize cats, faces (Peter Norvig) : موسسه

feature engineering

اقراین ویژگی ها در زمان پیش بینی تیت خانه:

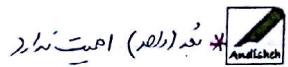
تعداد	تعداد صفات	تعداد خوب	...	Y
$x_1, x_2, \dots, x_n$	$x_1, x_2, \dots, x_n$	$x_1, x_2, \dots, x_n$	...	Y

تبدیل مقدار نیز نیز انجام می شود (برای افزایش سرعت محاسبه)

feature scaling

$$\frac{x_i - m_i}{\max(x_i) - \min(x_i)} \rightarrow x_i'$$

$$\frac{x_i - \bar{x}_i}{s_i} \rightarrow x_i'$$

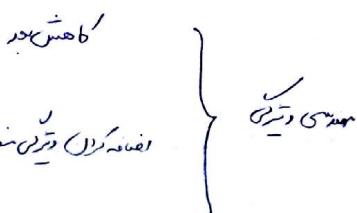


حالت

اندیلی (اندیلی) ایستاد

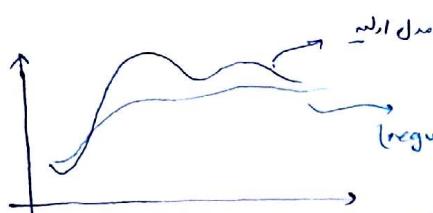
\* وقتی خطای بسیار سنی شود احمد اشغال است.

- Get more training examples
- Try smaller set of Features (PCA جهش خود)
- Try feature scaling / Normalization
- Try getting additional features (descibe the data better) داده را درست تر بگویند.
- Try adding polynomial Features (با درجه بیشتر)
- Increase or decrease of  $\lambda$



### Regularization

لکچر ۱۰: اشنایی با overfitting (برای لغزش نیست، که مخصوصاً در آزمون برای نیست) (high variance in training data)



خطای را طبقنی کردن (shrink) سلسله مطابق (regularized)

- حفظیه تغییر خوار - روابط بین داده ها زیرا و

خطای تنفسی را در پیدا کرده است این طبقنی کردن  $\rightarrow$  overfitting

تاج خوبی

$$J(\theta) = \frac{1}{m} \left[ \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=0}^n \theta_j^2 \right]$$

regularization

هدف  $\min_{\theta}$  در  $J(\theta)$  تاج خوبی است. اگر  $\lambda$  بزرگ باشد باید  $\theta$  را تحریک نمایند تا از  $\lambda$  خود کنترل کنند.  
اگر  $\lambda$  کوچک باشد ضیال کنترل نمایند و  $\theta$  را تحریک نمایند در حالتی که  $\lambda$  کوچک باشد.

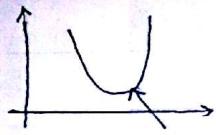
$$\boxed{\lambda \leq 0 \Rightarrow \text{استفاده رایج}}$$

$\lambda$	small $\lambda$	large $\lambda$
	variance $\uparrow$	bias $\uparrow$
	overfitting $\uparrow$	underfitting $\uparrow$

اصنایع در  $J(\theta)$  حدایی ایجاد نمایند این اشتاینی از مدل های بسیاره.

ردیس های تیر و smoothing هم بصردند. مثلاً  $\lambda \sum n \theta_j^2$  (این مبتدا های محبت باشند به معنای است)  
مقدیت تلاشان  $n$ : در تیر هایی من احتمال ندارم  $\theta_j$  بزرگ شود

(Normal Eq)  $\theta$  گیرنده جوابی Normal Eq جوابی



(Multivariate Linear Regression) MLR جوابی

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

goal: minimize  $J(\theta_0, \dots, \theta_n) = \frac{1}{m} \sum_{i=1}^m (h(x^i) - y^i)^2$

$$\bar{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} x_0 = 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \rightarrow h_{\theta}(x) = \bar{\theta}^T \bar{x}$$

$$\begin{array}{c|c} x_0 x_1 x_2 \dots x_n & Y \\ \hline \vdots & \\ \text{X} & \\ \vdots & \\ m(n+1) & \end{array} \rightarrow X = \begin{bmatrix} \bar{x}^1 & \bar{x}^2 & \dots & \bar{x}^n \end{bmatrix}^T$$

$$J(\theta) = \frac{1}{m} (\underbrace{(x\bar{\theta} - \bar{y})^T}_{\text{خطای ادغاما}} (x\bar{\theta} - \bar{y})) \quad (\text{خطای ادغاما})$$

$$\approx J(\theta) = (x\bar{\theta} - \bar{y})^T (x\bar{\theta} - \bar{y}) = ((x\bar{\theta})^T - \bar{y}^T)(x\bar{\theta} - \bar{y})$$

$$= (x\bar{\theta})^T (x\bar{\theta}) - \underbrace{(x\bar{\theta})^T \bar{y} - \bar{y}^T x\bar{\theta}}_{\text{از حم مساله نستاد}} + \bar{y}^T \bar{y} = \bar{\theta}^T x^T x \bar{\theta} - \gamma (x\bar{\theta})^T \bar{y} + \bar{y}^T \bar{y}$$

$$\frac{\partial J}{\partial \theta} = \gamma x^T x \bar{\theta} - \gamma x^T \bar{y} = 0 \Rightarrow x^T x \bar{\theta} = x^T \bar{y} \Rightarrow \bar{\theta} = (x^T x)^{-1} x^T \bar{y}$$

(inv)  $\neq$  معکوس نداشته باشد

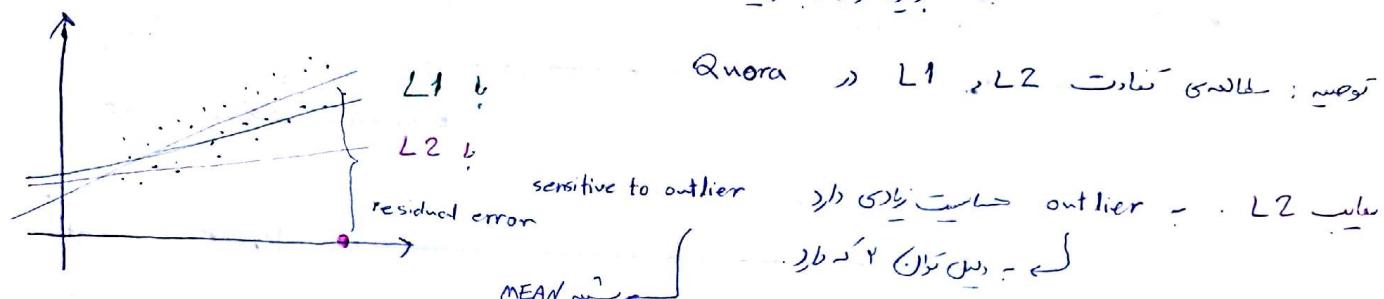
\* استناد از  $P_{\text{inv}} \neq 0$  ، اگر  $x^T x = 0$  بود که میتواند دیراست  
 از حم مساله نستاد باشد redundant features .  
 (تعداد دیرها خیلی زیاد هستند)  $n \gg m$  : علت دیر



٩٧، ٨، ١٠ . . . . .

أنواع خطأها في المنهج:

- o LSE (Least Square Error) =  $\sum_{i=1}^m (h(x^i) - y^i)^2$
  - o MSE (Mean Square Error) =  $\frac{1}{m}$
- } norm  
L2 Error  
(regularization)  
لإثبات انتشار الخطأ



3b) Ridge Regularization  $\rightarrow$  L2  $\rightarrow$  regularization

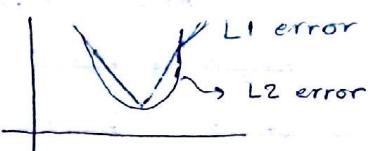
- o LAD (Least Absolute Deviation) =  $\sum_{i=1}^m |h(x^i) - y^i|$
  - o MAD (Mean Absolute Deviation) =  $\frac{1}{m}$
- } L1 Error  
لـ L1 خطأ

3b) Lasso Regularization  $\rightarrow$  L1  $\rightarrow$  regularization

MEDIAN  $\rightarrow$  Robust to outlier : L1 خطأ

o RMSE =  $\sqrt{\frac{1}{m} \sum (h(x^i) - y^i)^2}$

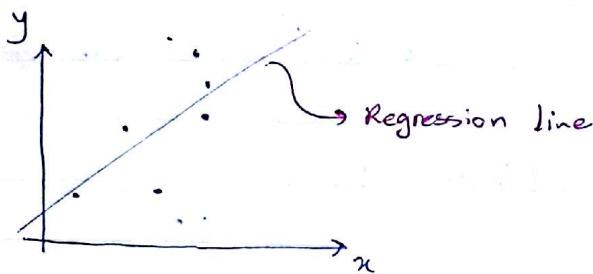
لـ L2 خطأ



لـ L2 خطأ

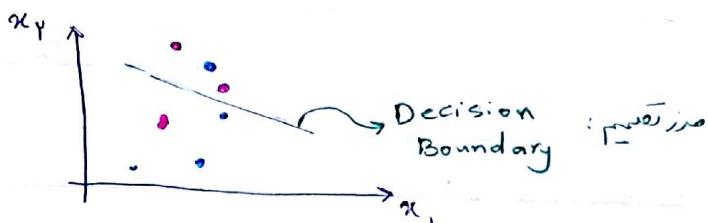


گروہیت: category / نامہ جزوی supervised



regression line  $\hat{Y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_j$  :-

وَصَدَّ مُهَاجِرَ مُتَقَلِّدَ دَالِبَةَ



نئی طبقہ نامیں (ھر دو ویرس میں)  $h: x \rightarrow \{+, -\}$

$h: x \rightarrow \{+, -\}$       binary classification

الgoritم الـ KNN (K-nearest neighbor) ينبع من الـ basic الـ k-\_nearest neighbor

$$d(\bar{u}, \bar{u}') = \sqrt{\sum_i (u_i - u'_i)^2}$$

Manhattan	v
Minkowski	v

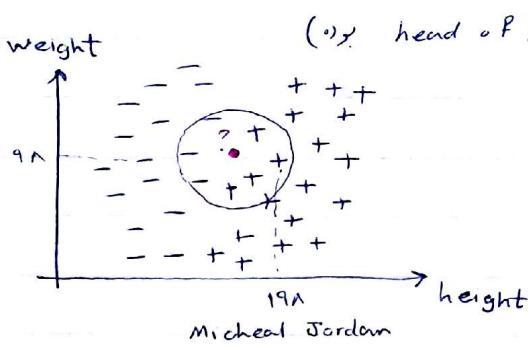
نامه ایشان را در حالت کلید (key) نوشته اند.

بیانار مدرسہ اسٹ (از حدود سال ۱۹۲۰)

pattern recognition

K عوّتّ خر انتخاب می شود و در مانند داده مسادیر بالا رنگ نقره ای می شود. طبق می تبست با K همچنان اطوان خود سمجده ای می شود که با این خارج از میان noise outlier برود

Weighted KNN ← KNN با وزن



(o) head oF Linked In

Ron Bekerman

$\vee - \sim N$   $\rightarrow$

? : +

## وثری های KNN

## Regularity based -

• 3b job training schools باختصار ← Instant based -

## (زیست‌حیاتی) Non-Parametric

پردازشی : عنوان تایپ داده ها که داشته باشند  $\rightarrow$  در KNN کاری تایپ ندارند  
نمایش کجاها داره  $\rightarrow$  dense است و...



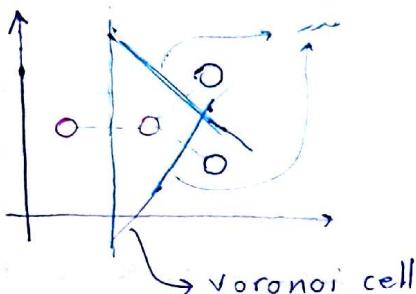
- الگوریتم Lazy : در سطحی training همچو کاری انجام نمی‌زند (از training طریق داده‌ها در سطحی test) .

$$\begin{cases} \text{training} \approx O(1) \\ \text{test} = O(m) \end{cases}$$

وقتی های الگوریتم سبل:

در دنیای big data این الگوریتم حافظه‌مند (non-scalable) بود. برای هر داده‌ی تست نیاز به داده‌ها را بررسی کند.

KNN : صادر IBK ← IB : weka ,  
↓  
Instant Based



بررسی آرزوی صدر رفعی:

خود منصف را بنیادی کاری ایندی رسماً نیست.

برای حالت 1-NN

تعداد پایه خط‌ها از تعداد داده‌ها است. معادله‌ی منتهی نیست  
و به تعداد پایه خط‌ها نایاب داده‌ی تست را با هر خط تفاسی کند (AND)  
هم‌چنین نایاب OR نایاب با جزئیات بیان نمود تفاسی کند.  $\Rightarrow$  تاثیری در order در نداشتن ندارد.

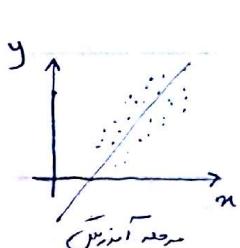
برای حالت 1-NN

(حرداً این فرم از آن ممکن نیست)

overfitting / high variance : درجه k

(کثر  $k=m$  باشد تنه همانه می‌شوند خواهد بود)

underrfitting / high bias : درجه k

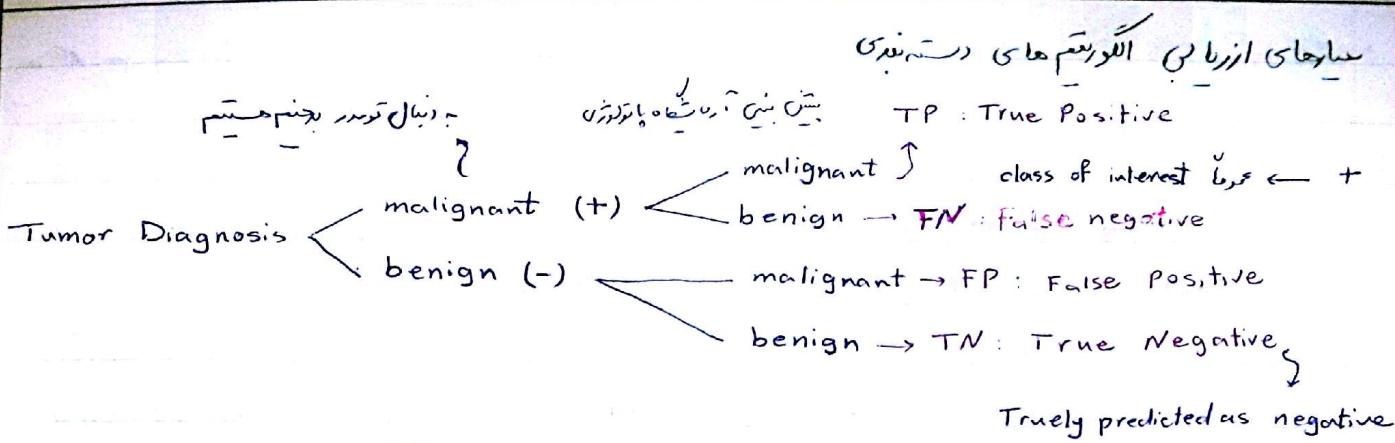


لطفی کاری عین درجه k :

در مدلی نیست به طوری که آمنی نیاز نداشتم که بین از دیگران طایر الگوریتم‌های

non-parametric KNN را در مدلی نیست - طوری که باید  $k$  را نیاز داریم .

parametric لجیستیک \*



$$\left\{ \begin{array}{l} TPR = \frac{\# TP}{\# P} = \frac{TP}{TP + FN} \\ TNR = \frac{\# TN}{\# N} = \frac{TN}{TN + FP} \end{array} \right.$$

= Sensitivity = Recall

= Specificity =

PR, IR میانی  
Information Retrieved  
Pattern Recognition

$$\left\{ \text{Accuracy} = \frac{TP + TN}{P + N} \right.$$

$$\left\{ \text{Error Rate} = 1 - \text{Accuracy} \right.$$

$$\text{precision} = \frac{TP}{TP + FP}$$

FNR, FPR \*  
accuracy, TNR \*

نمایش اطلاعات نظریه ای و الگوریتم های طبقه بندی شده با توجه به این اشاره کنید.

recall, precision \*

$$\left\{ \uparrow TPR + \downarrow FNR = 1 \right.$$

$$\left\{ \downarrow FPR + \uparrow TNR = 1 \right.$$

\* خطا های خطای نوع I و II هستند استراتیجی های در سامانه های این اهداف را می خواهند. انتباخته سیمیر این اهداف را می خواهد. accuracy = این معنی در درست های spam یا ایمیل هم را spambot می نامند. در حالیکه در این معنی خطاها فقط گیتی هایی هستند.

$$G\text{-mean} = (\text{precision} \times \text{sensitivity})^{\frac{1}{2}}$$

\* افوار سیمای این اسناد را در این میانه ای اسناد برآورد می کنند.

$$F1\text{-score} = \gamma \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

(F-measure)



		pos	neg
True Label	pos	TP	FN
	neg	FP	TN

Confusion Matrix (ماتریس خطا)

که اگر از تابع خود مخفون بگذاریم ماتریس ایجاد شود، جمله های خوب را بروز نماید.

Imbalanced class distribution  
 " learning

از داده های کم و بیش میان مخفون و مثبت است.  
 راه حل های آن:  
 ۱- اضافه کردن تعداد اعضا های طبقه مثبت و  
 ۲- SMOTE برای ایجاد توزیع متساوی  
 ۳- مخفون تعداد اعضا های مثبت و مخفون آن را برابر نمایم

حلبی ۹۷, ۷, ۲۲

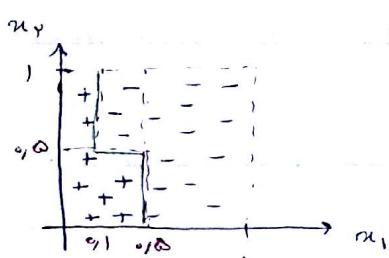
### (Decision Tree) درخت تقسیم

www.20q.net → با شنبه عجیب، با ۲۰ سوال - مخفون همراه

en.akinator.com → با DT شعیانه - در زنگ تراست پیرامون

hypothesis

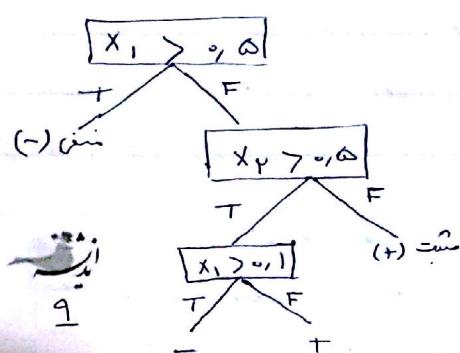
مثابل تقدیر است. (که عجیب دیده مائین) جزو مثابل تقدیر  
 مثبت است



مراسی ساده در نزدیکی محدوده مخفون است در نظر داشته شده

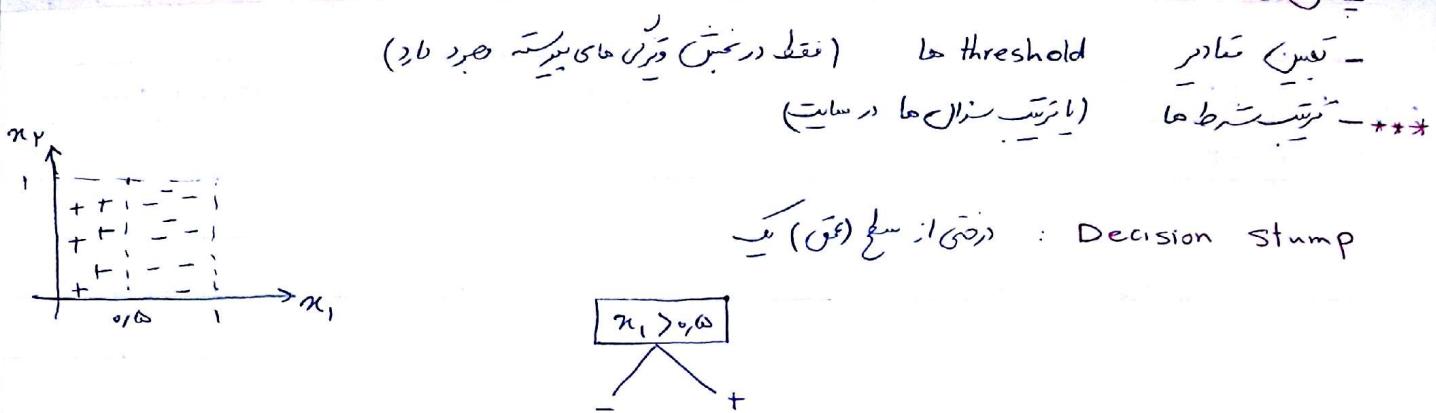
خطای درجه ۱: مثابل تقدیر

منطقی DT بسیار ساده از KNN است، smooth و مطابقت نسبتی دارد.



آن درخت است - ازین فضای حسب درجی درخت های مخفون بود

چالش‌ها



برآورد  
underfit  
عویض رفت  
(complexity)

برآورد  
overfit

\* بحدیث رخت - عویض رخت شده باشد

روخت باعنه! : اخطار نیزه‌ای آن که است دیالیس آن زیاد است.  
روختی نیزی اضافه شود - خاطر آن عرض نمایند

در حقیقت → ساز داده‌های مختلف برای داده ← دارایان بالای ۶۰

مردی به طوری درست توزیع ماره‌ها (کام) دهد ← باز استرس

non-parametric ← DT \*

منز تعمیم خوب ای ۲۶

DT الگوریتم قدیمی است . الگوریتم CART نیز همین‌هاست و به حجم بینه هستند .

نقطه دسته‌نیزی DT

CART (Classification & Regression Tree)

نمودار دسته‌بندی هم نیزه‌ی (CART) ← دیگر نیزه‌ی هست

هر دو الگوریتم از نظر آنتروپی استادی می‌شوند

نمودار آنتروپی (Entropy)

دیگر نیزه‌ی بولن می‌خریزی ماشین کی دهد . / عدم تغییر

حیث آنتروپی کسر (جایه) ← اطمینان نابل (نیمال سیستم)  
در عیش‌های رندیکاری صفر نمایند

که ندارید ، نعم نقطه داده است بروز نمودن ! منز تعمیم

منز دفعات منز تعمیم است



خریز این پذیره نه تعمیم است

+ more bits just in case  
2 bit

اهمال بولن جایزه کی نیز

→ دو تعدادی باینری ← تغییرات بولن

$H \rightarrow E$

$E: \text{ent}$

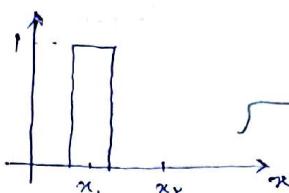
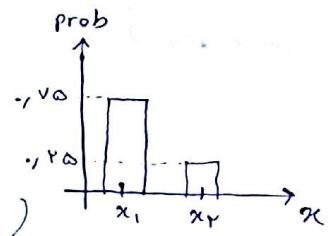
در قدم:

مادن آنتروپی

ردود بازیجی هسته: میان میان بین ناهم میان انتقال اطلاعات  
 آنتروپی کمتر تعدادی:

$$E(x) = \sum_i p(x_i) \log_2 \frac{1}{p(x_i)} = -\sum_i p(x_i) \log_2 p(x_i)$$

مقدار آن در میانی است. در حالت جنبه کم تغییر دارد از این نظر



$$\rightarrow E(x) = -(0.15 \log_2 0.15 + 0.25 \log_2 0.25) = 0.813 \text{ bit}$$

inductive learning پارامتریک اعلایی

آنتروپی در رسمتینی / پارامتریک مادن:

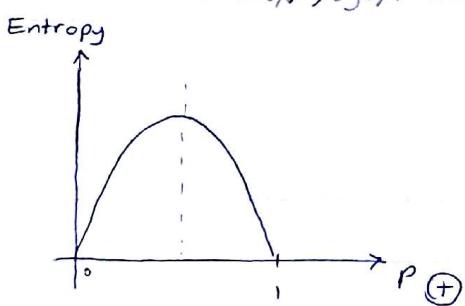
دیجیتال ناممکن بردن این حالت آنتروپی / سمعیه ای از نسل

+ تکمیل

$$E(s) = -P_{\oplus} \log_2 P_{\oplus} - P_{\ominus} \log_2 P_{\ominus}$$



$$= -0.1 \log_2 0.1 - 0.1 \log_2 0.1 = 0.88$$



نمایش انتروپی از سمت چهل و سه سمت لانه آنتروپی (سیاست 26)

\* همه آنتروپی ها "پارامتریک مادن" هستند. پس ما در حقیقت آنتروپی حرارت کوئی

جست آنتروپی فقط در  $\omega$ ، Tom Mitchell

[+9+, -10-]

[F1 = ?]

T F

[+1+, -0-] [A+, -0-]

①

②



در مجموع - نظریه ای  
 در تئوری شرمن است

[+9+, -10-]

[F2 = ?]

T F

[A+, -0-]

④

Feature by attribute

[11+, -2-]

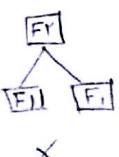
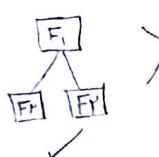
ساده

لطفاً

- اندیل

- اندیل

- اندیل



کام در درست راست؟

نهاده اند که F خوب نیست و مساختهای بحثی خوب نمی‌شوند. در دامنه با درنظر بردن مقطع آندریو اینجا

۲۶ به local opt. انتشار نمود.

حلقه ۱۵ ۹۷، ۷، ۲۴

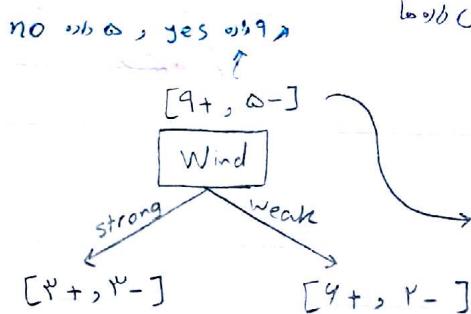
برای اطلاعات  $\leftarrow$  می‌بینیم سر (information gain)

برای اطلاعات می‌بینیم feature / می‌دانیم از متغیر طبق آندریو به دارایی جذابیت می‌باشد از خود این متغیر حاصل گشتند.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

آنقدر کل داده ها

جزوی شناسنامه  
مجموعه شناسنامه  
attribute/feature



PlayTennis/Wind : JL

$$E(S) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10} = 0,914$$

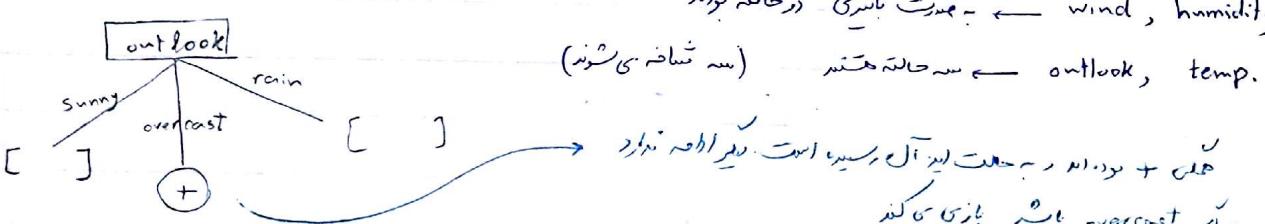
$$\begin{cases} E(\text{wind} = \text{strong}) = 1 \\ E(\text{wind} = \text{weak}) = -\frac{9}{10} \log \frac{9}{10} - \frac{1}{10} \log \frac{1}{10} = 0,181 \end{cases}$$

$$\text{Gain}(S, \text{wind}) = 0,914 - \frac{4}{10} \times 1 - \frac{6}{10} \times 0,181 = 0,048 \approx 0,05$$

دارایی تک ت ویرانهای ایجاد می‌شوند gain

بهمت بانیی در حالت باد و wind, humidity

(سه شناسه می‌شوند) سه حالت مختلف



پس از بروز این رخدادات بیرون می‌شوند است. در اینجا نیاز داریم بازی کنند.

بعد از داشتنی دیرینه gain می‌باشد.

Iterative Dichotomiser

hypothesis space search  
ID3 جستجو در فضای فرضی (Hypothesis Space) - الگوریتم



روضت ~ Iterative ID3 ← ساخته شود تا در مرشاده - توقف شود.

الgoritم ID3 - حدیث علی می نماید

local opt. سے اقبال نہیں کرے

فقط ماتسیس صافی کستہ طاری کرے (کوڈ میں نہیں)

عنوان است ویری فیکٹور دیجی دیجی خوب عمل نہیں کرے ID3 ایک دوسرے نہیں

rulequest.com C5.0 → نسبی عبارت آن C4.5 اصلاح نہیں ID3

ب وظیفہ آنہا نہیں!

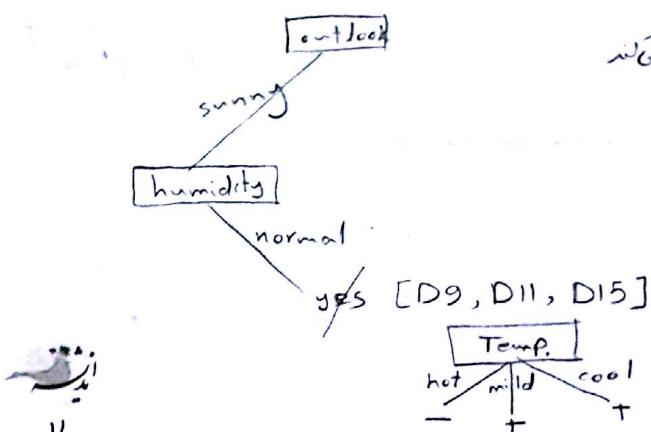
ب نوع اطمینان بعد دارو

\* دریں دیا اھست جیلوں باتتہ د ظاہر شدہ یا باسیز دریں کا کچھ بیوں دارتہ باشے => غیبت دریں  
 \* سرطخانہ؟  
 غیبت آئیں ایں است کہ جیلوی شاپتھا ہر بیک ریکنڈ دیں بیک دیں سانیل دیس جس آنکھیں غیر ملائیں  
 اور سرطخانہ ہے جو بیک کیلئے بلایم مکمل است مگن بیک سترد overfitting  
 استناد ارجمند التعریف

روضت است بیکارا، قل عیت در training (ایکم کی نظر) ← DT

آخر دادی صدی - train اھنامہ شود ← دس آن است کہ سراحت نہ اول نہ اول کسی دلیلی خواہی ہے روضت ملکی آن کا اعلانہ دیں

D15 = ( <out=sunny, Temp=hot, Hum=normal, wind=strong>, - )

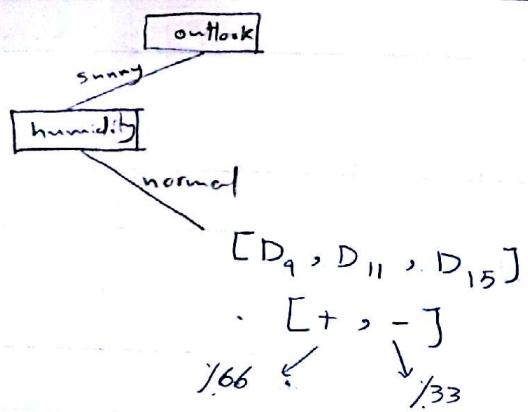


آخر دادی صدی نہیں علی کوئی

اھنامہ دلیل دادی نہیں \*

آخر دادی صدی ایں علی کوئی دلیل دادی نہیں درد

overfitting نہیں



همن طریق اضافی سیم، اعمالی درب را سفید کنند.  
- همین دلیل است که هر کسی (خجال) بگزیند.  
در حصار یا اضمال معتبر را عالش می دهند.

not balanced,  $\Delta n = 0$

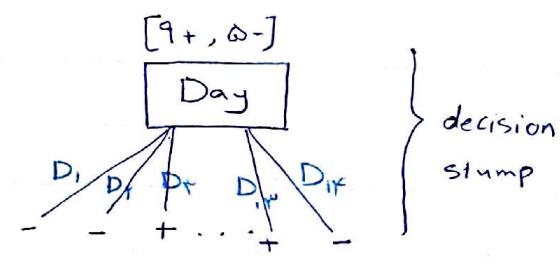
در C4.5 bias است به علت اینکه ریختان دوچیزه (باعنده متر) (اگر دو دست حمل کنیم راسته باشد دوچیزه دیناه تر را اینکه باید بکند)

→ سریع ریل های تلف ناچر کی ماسن را آن باره سازی نمود.

• C4.5 C4.5 J48 ←

٩٧، ٧، ٢٩ : ١١ مل

ـ نتائج ملحوظة في متغيرات Time Series Analysis (Temporal Feature) في data mining



نقطة مرجعية (Oasis) ID Code (Identifier)

مایل است نیز نباشد. (نمودار) در ۱۲۰ تا ۱۶۰ سانتیمتر

و شریعت های متعارف را در سازه لوحه

$$\text{Gain}(S, \text{Day}) = 0,94 - \frac{1}{14} \times (1 \times \log_{10} + 0 \times \log_{10}) \times 14 = 0,94$$

میکروسیستم  
حیاتیان

احمد دلیری زنگنه

کے ہم سائیسے کے نتایج میں دیگر حاصل ہوں گے  $\leftarrow$  bias ناچالب رہتے

Mitchel میچل

میز جدیدی از خریل Gain باینی مایس:

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Entropy}(S, A)}$$

انشارهای تعدادی باینی مایس  
برابرستد با Gain

$$\text{Split Entropy}(S, A) = - \sum_{i=1}^{|ValesA|} \frac{|S_i|}{|S|} \log \frac{|S_i|}{|S|}$$

برخاف قبل میان target

Gain Gain Ratio

$\text{Gain}(S, \text{Hum}) = 0.151$	0.152
$\text{Gain}(S, \text{outlook}) = 0.246$	0.156
$\text{Gain}(S, \text{wind}) = 0.048$	0.049
$\text{Gain}(S, \text{Temp}) = 0.029$	0.019

انشارهای با ترتیب جدید این تعداد مشترک شدند  
جیل ایکس ایکس نتیجه outlook بر ایندtemp در حالت بود.

$$\text{Split Entropy}(S, \text{outlook}) = -\left( \frac{1}{4} \log \frac{1}{4} + \frac{3}{4} \log \frac{3}{4} \right) = 1.84$$

$$\text{Gain Ratio}(S, \text{outlook}) = \frac{0.246}{1.84} = 0.1344$$

$$\text{Split Entropy}(S, \text{day}) = \left( \frac{1}{4} \log \frac{1}{4} \right) \times 14 = 1.84$$

$$\text{Gain Ratio}(S, \text{day}) = \frac{0.194}{1.84} = 0.1044$$

نتیجه از هم حل کرد.

در ابتدا نایر خیلی دریخهای متاسف شدند و ارجمند preprocessing خویش شدند - داده‌نایاب شدند ایشان  
... ID code

C 4.5

ID3

GainRatio

(numerical) کسری، پیوسته

Gain (Basic Info Gain)

nominal, count

~ ~ missing values!

pruning

X (back track)

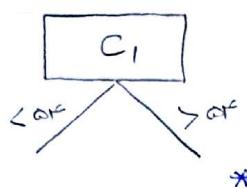
۱۲

دریج ممکن است از این نظر overfitting و generalization می‌شود و نتیجه ایست. (دسته‌بندی نیز)

### (C4.5) numeric درخت داده‌های عددی

40	48	60	72	80	90
-	-	+	+	+	-
$C_1 = 54$ یعنی				$C_2 = 85$	

sort اولیه در اینجا است



$$\left\{ \begin{array}{l} E(C_1 < \alpha_f) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 0.9V \\ E(C_1 > \alpha_f) = 0 \\ \text{Gain}(S, C_1) = 1 - \frac{1}{4} \times 0.9V - \frac{1}{4} \times 0 = 0.19 \end{array} \right.$$

$$\left\{ \begin{array}{l} E(C_1 < \alpha_f) = 0 \\ E(C_1 > \alpha_f) = -\frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} = 0.11 \\ \text{Gain}(S, C_1) = 1 - \frac{1}{4} \times 0 - \frac{1}{4} \times 0.11 = 0.144 \\ (-0.144 + 0.144) E(S) = 1 \end{array} \right.$$

اولیه درخت  $C_1$

جزوی تبدیل داده‌ها برای استفاده کنیم.  $\leftarrow$  C4.5  $\leftarrow$  Gain Ratio

missing value  
در هری داده‌ای ترسیه آن، متاداری را حذف نمی‌کنیم.  
هر کدامی که می‌تواند با دررسی بیان شوند

non-parametric  $\leftarrow$  آنالیزی آن) استفاده کنیم (فرض در صور توزیع داده‌ها غیردستگیر

درخت عاملی  $\leftarrow$  طبقه‌بندی الگوریتم

جلسه ۱۹ : ۹۷/۸/۱

scalable machine learning

اداری دیتای نیزین

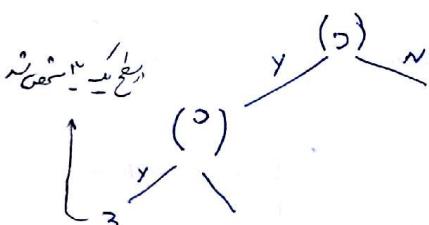
(۴۶۰) میان روش دیگر  $\leftarrow$  تحقیق اداری  $\rightarrow$  ایجاد نویز سیگنال  $\leftarrow$  روش دیگر  $\leftarrow$  image processing

۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
۰	۸	۲	۳	۴	۵	۶	۷	۸	۹

OCO

۴

Feature vector



multi class confusion matrix

پرسش پرسانی (وسم چندی) برای حالت خوبی:

	A	B	C	D	E	→ predicted
A	TP <sub>A</sub>	E <sub>AB</sub>	E <sub>AC</sub>	E <sub>AD</sub>	E <sub>AE</sub>	
B	E <sub>BA</sub>	TP <sub>B</sub>	-	-	-	
C	E <sub>CA</sub>	-	TP <sub>C</sub>	-	-	
D	E <sub>DA</sub>	-	-	TP <sub>D</sub>	-	
E	E <sub>EA</sub>	-	-	-	TP <sub>E</sub>	

actived

A  $\downarrow$   $\leftarrow$  TN

class of interest

یک طبقه سمت باقی ماند

برای طبقه مورد نظر: TP<sub>x</sub>

نهازی Y برابر با TP<sub>xy</sub>

چه خواصی مطابق با این را برای الگوریتم معتبر به علاوه:

مجمع تغطیه اصلی (برای طبقه مورد نظر)  $\leftarrow$  در حقیقت: TP

neg  $\leftarrow$   $\neg$  pos A  $\leftarrow$  A میتواند  $\neg$  طبقه ای داشته باشد  $\leftarrow$  TP<sub>A</sub> (A میتواند  $\neg$  طبقه باشد): FP

TP<sub>A</sub> میتواند A میتواند: FN

(برای طبقه A):  $\neg$  میتواند A میتواند: TN

و حال میتواند  $\neg$  میتواند

دباسیت های استاندارد:

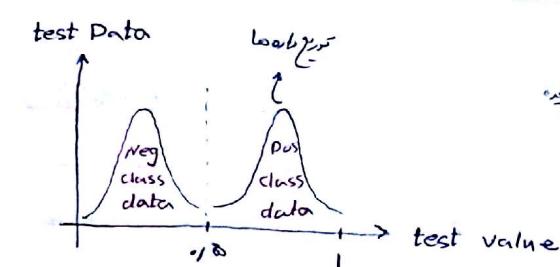
دباسیت استاندارد از دسته طبقه کالیفیریا

همه طبقه ای از دسته شده دباسیت استاندارد نمایند

UCI ML Repository

دباسیت های استاندارد برای تهییزی غیر اولیه دباسیت های مختلف استاندارد می شوند.

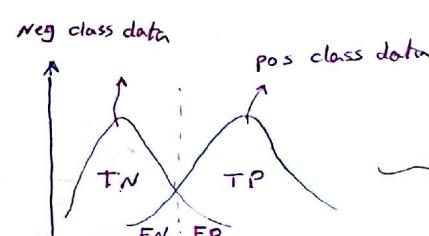




ROC : سینی حایی  $\rightarrow$  از مقدار چنانچه در نتیجه گیری کارکرد آن

if test value > Treshold ;  
positive cut off  
else حسی خوب

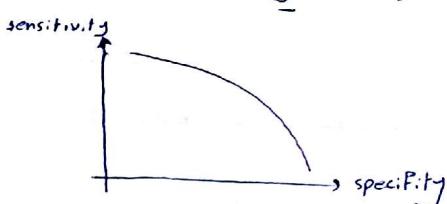
negative



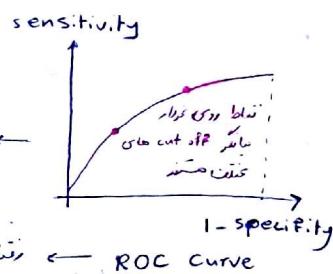
پلت ایندیکاتور است  $\rightarrow$  این سیستم پس از این کار  $\rightarrow$  neg

sensitivity افرادی از این خانواده  
specifity افرادی از این خانواده cut off حایی جایی

sensitivity افرادی از این خانواده  
specifity افرادی از این خانواده cut off حایی جایی



این خوارد را به مدل‌های مناسب نیز می‌دانند.



بهمت بایع این دو دلایل می‌باشد که این مدل‌ها می‌توانند

میزان اطمینان

یا نظر الگوریتم دستیابی نمایند ایست

الگوریتم شناختی باشند

نیز حسی ممکن است

آیا الگوریتم داریم که از برگریل است؟

- نه، آنچه الگوریتم داشت هیچ نوبه نداشت

الگوریتم‌ها در نظر گیرانده تعدادی عمل می‌نمایند

هر کدامیکی از آنها در الگوریتم می‌ساخت نیز نهاده می‌گیرند

AUC = Area Under Curve

AUC	نیفست
0.9 - 1	Excellent
0.8 - 0.9	Good
0.7 - 0.8	Fair
0.4 - 0.7	Poor
0.0 - 0.4	Fair



## جلسه ۱۳: معرفی شبکه عصبی - یادگیری عمیق

دلخواهی دارند → درود

عمل: اسی → درودی را → خروج تبدیل کنند function

linear Regression : UFLDL : پیشوازه بجزی سعی  
Prediction : Nando de Freitas

classification: Softmax, cross Entropy ۱, ۲

UFLDL : Softmax Regression

unsupervised Feature learning

tutorialspoint سایت

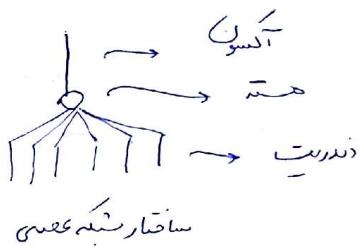
خرچه دیگرها بینه شد بازنمایی بینه شد و بینه شد

دونه → دصل کردن نایک خط

دونه رخط نهیسته ← بیان زب (اقوام) که به (یادگیری) نایک از خط جدا کرده آغاز افتاد

MNIST ← عکس از اینام دست نویی ← سیم عکس ۲۸x۲۸ بینه ← کیه بردار ۴x۷۸۴ دیگر

بریزهای یادگیری ماین هست یادگیری عمیق میتواند با طریق های مختلف گلودر خوبی را نهاده



از عکس به ساختار هسته ای برسی

فناوت از هسته های ایست

اطماعات عجیب شود دلکر - تنداری بسیار

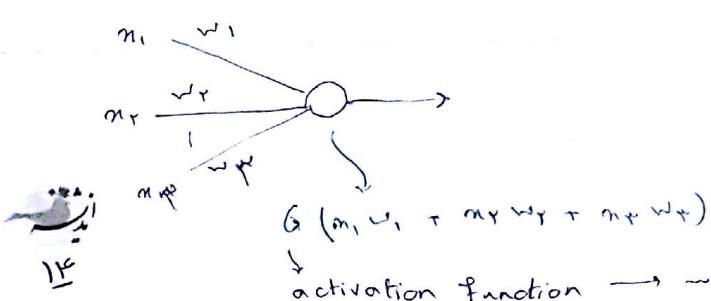
کی شد trigger

لے اور در نه سر شد دلکر برای هسته عصبی (neuron)

feature هایی در نظر نماید deep learning

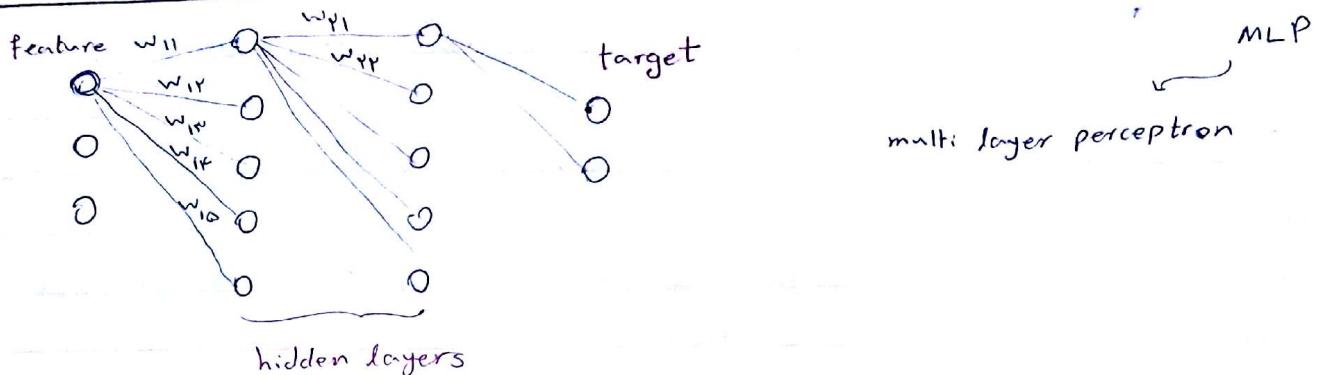
خرچه شبکه عصبی n+1 ← نهاده میکس

perceptron : one



$G(m_1w_1 + m_2w_2 + m_3w_3)$

activation function → w, sigmoid



courseera

جنس ۱۳ : ۱۰ جمادی ۹۷، آ

دستور افعال:

فرضیه های چهار تغیر تعدادی (WetGrass, Rain, Sprinkler, cloudy) می باشد  
متغیر T = خودکشیده

C	S	R	W	Pr
F	F	F	F	0.9
F	F	F	T	0.1
.	.	.	.	.
.	.	.	.	.

اگر دریچه آنرا احتمالات را از نظر باشیم حساب می کنیم که ای تراویم به صورت ۱۴٪ است  
- فرضیه دلخواه آن احتمال برابر باشد. - دنبال نهادن متغیری  
اگر آن را محاسبه کنیم:

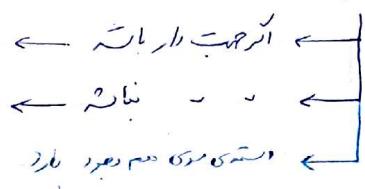
probability

(PGM) Graphical Models

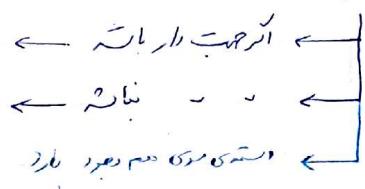
عکس فشرده تر

برای مجموعه ای از متغیرها

(belief net) Bayesian network



Markov Random field



کاربرد این مدلها استغفار، اطلاعات زیستی

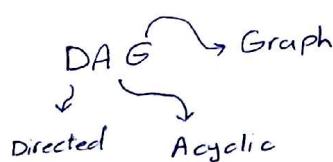
خانم

هم حسب

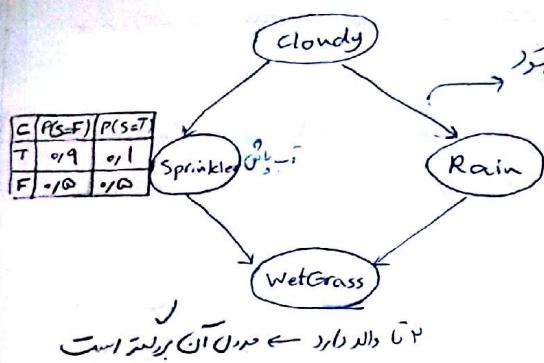
courseera

\* مدل های هفت طبقه causality

\* درایف سیکلیک دور (cycle) می باشد (این مدل های خودکشیده ای از فرضیه ایست)



Subject: \_\_\_\_\_  
Year: \_\_\_\_\_ Month: \_\_\_\_\_ Date: \_\_\_\_\_



C	P(S=F)	P(S=T)
T	0.9	0.1
F	0.1	0.9

C	P(R=F)	P(R=T)
T	0.4	0.6
F	0.6	0.4

P(C=F)	P(C=T)
0.1	0.9

prior probability

اول احتمالی داده شده باشند. اول احتمالی داده شده باشند. اول احتمالی داده شده باشند. اول احتمالی داده شده باشند.

خوب است

(0.1, 0.9, 0.1, 0.9)

CPTs

محظوظ از خود را افکار می‌شود

S	R	P(W=F)	P(W=T)
T	T	0	1
F	T	0.1	0.9
T	F	0.1	0.9
F	F	1	0

divergence arrow

PGM چیزی

W, S, C  $\perp\!\!\!\perp$  W, R, C : serial connection

S, R, C  $\perp\!\!\!\perp$  : head to head connection

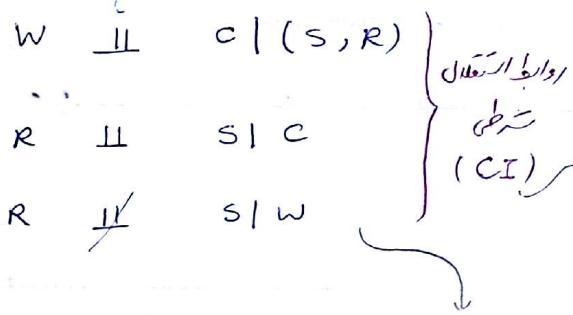
convergence arrow

W, S, R  $\perp\!\!\!\perp$  : tail to tail n  
(V structure)

استثنای سیم

قانون اول آنکه اگر لذت می‌شوند اینکه در حالت حال باشند

Markov



از S از R مستقل است بشرط که C باشد.

با شرط که R باشد از R مستقل است بشرط که C باشد.

با شرط که C باشد از R مستقل است بشرط که R باشد.

با شرط که W باشد از R مستقل است بشرط که R باشد.

با شرط که W باشد از R مستقل است بشرط که R باشد.

marginalize: مخفی  $\rightarrow$  محض نویزی  $\rightarrow$  مخفی نویزی  $\rightarrow$  مخفی نویزی  $\rightarrow$  مخفی نویزی

سطو: یعنی محض نویزی دلیل برای حیثیت بردن چیزی داشته باشد

$$P(S=1 | w=1) = \frac{P(S=1, w=1)}{P(w=1)} = \frac{\sum_{C} P(C=c, S=1, R=r, w=1)}{P(w=1)} = \frac{0.48}{0.48} = 0.48$$

$\Rightarrow$  Rain: طوفان

محض نویزی

$$P(R=1 | w=1) = \frac{P(R=1, w=1)}{P(w=1)} = \frac{0.44}{0.48} = 0.916$$

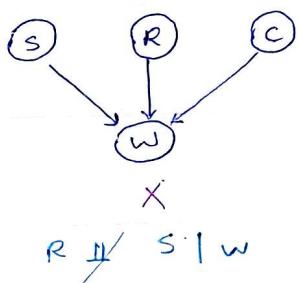
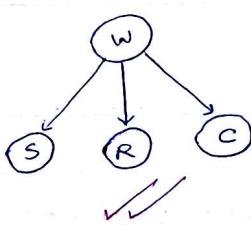
درین سند جوی نخواهد بیسی داشت این از آن صفت نخواهد داشت اگر غرایص عبور اعیان های تند

normalization Factor

\* بُرْجی نیشن  $\hat{w}$  برای  $\hat{w}$  کاپسی  $\arg\max$  سنت.

\* خوب نشسته درون مدل:  $\hat{w}$  از  $w$  های آمریکی  $\hat{w}$  با این مدل مبالغه است.  
 که سُفْرِ کران ساختار درون سیار دیگر دستور است.

\* بُرْجی نیشن ساختار ایجادی و ترتیبها در ترتیب مدل ساختار است.  
 آنچه کنم درون  $w$  از  $w$  های دیگر متصل است:  $\leftarrow$   
 Naive Bayesian



و این  $w$  دیر  
متصل نیست

برای  $w$  روابط علت مدل داشتم

\* در جلسه درستهای بین درج در:

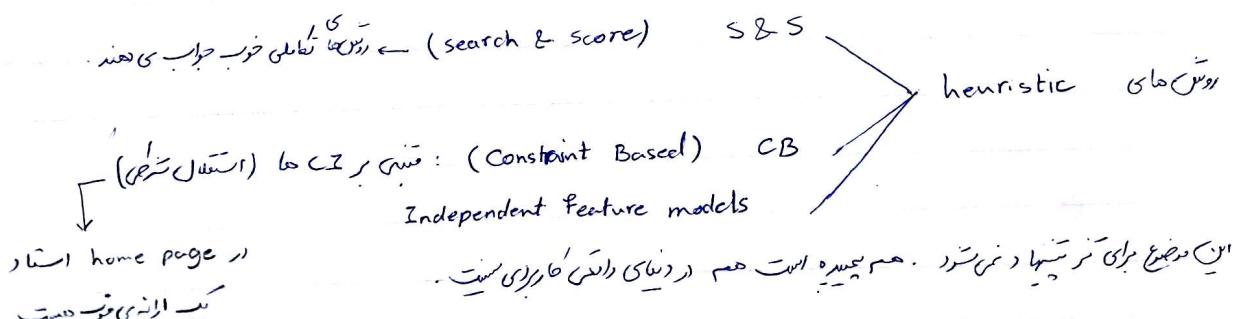
$$P(w|R, S, C) = P(w|R, S)$$

ساده می شود ب

1) بُرْجی نیشن DAG  
 2) بُرْجی نیشن CPT  $\rightarrow$  استعمال ترتیب باعث ساده بُرْجی نیشن CPT های پر کرد

Chickering et al. 2004

بُرْجی نیشن ساختار BN، مدل NP-hard



9V, A, 10 : 10 نمایش

### Naïve Bayes Classifier (NBC) (Independent feature model)

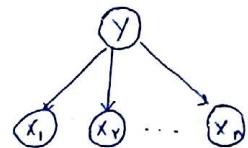
$$h: x \rightarrow \{+, -\}$$

}

$$\langle x_1, x_2, \dots, x_n \rangle$$

متغیر تعیین کننده

با احتمال این درست شدن این مجموعه از متغیرها



$$P(Y|X) = \frac{P(X|Y) P(Y)}{P(X)} \propto P(X|Y) P(Y)$$

maximizing argument

$$\begin{aligned} y_{MLE} &= \underset{y \in Y}{\operatorname{argmax}} P(Y=y|X) \propto \underset{y \in Y}{\operatorname{argmax}} P(X|Y=y) P(Y=y) \\ &= P(x_1|y) P(x_2|y) \dots P(x_n|y) \underset{\text{constant}}{\cancel{P(y)}} \end{aligned}$$

مثال: شمعون در طبقه اول است. و در های تبریز (تهران).  
(children ۱۲ > adults ۴) از عمر ۱۶ تا ۲۰

$$\left. \begin{array}{l} P(a) = \frac{4}{14} = \frac{1}{4} \\ P(c) = \frac{12}{14} = \frac{6}{7} \end{array} \right\} \text{prior prob.}$$

تصویب داریم شعبه با Gaussian NB  
برای معرفت

فرض کنیم توزیع داده است (دروازه های اثبات حس سی خواهد بود)

$$\begin{aligned} \bar{x}_{h,a} &= \frac{1}{4} \sum_{i:y_i=a} h^i \\ s_{h,a}^2 &= \frac{1}{4} \sum_{i:y_i=a} (h^i - \mu_{h,a})^2 \end{aligned}$$

unbiased estimator of  $n-1$  است و  $n=4$  است

$(\mu_{w,a}, s_{w,a}^2) \neq (\mu_{h,a}, s_{h,a}^2)$  برای adults

دلیل است

$(\mu_{w,c}, s_{w,c}^2) \neq (\mu_{h,c}, s_{h,c}^2)$  برای children

دلیل است

چون با غیره هایی داشتیم نه حاجی داشتیم نه نامناظری ها صدق نمود



محضی تمرین: داده های از دست آمده را در تابعی کنید

$$\{ P(h_x | c) = \frac{1}{\sqrt{2\pi} \delta_{h,c}} e^{-\frac{(h_x - \mu_{h,c})^2}{2\delta_{h,c}^2}} \sim N(\mu_{h,c}, \delta_{h,c})$$

$$P(w_x | c) = \sim w_c$$

$$\{ P(h_a | a) = \frac{1}{\sqrt{2\pi} \delta_{h,a}} e^{-\frac{(h_a - \mu_{h,a})^2}{2\delta_{h,a}^2}}$$

$$P(w_a | a) = \sim w_a$$

$$\{ P(y=a | x) \propto P(h_x | a) P(w_x | a) P(a)$$

$$\{ P(y=c | x) \propto P(h_x | c) P(w_x | c) P(c)$$

$$y_{MLE} = \operatorname{argmax}_y P(x | y) P(y)$$

مثال ارائه شدهای لمسه: NBC نیوزیم PlayTennis را دنیاگیری کرد. آنها دسته ای از باری داده های تمرینی داشتند که برای آنها آنرا متنهای ایجاد کردند.

( $\langle \text{Rainy}, \text{cool}, \text{high}, \text{True} \rangle$ , ?)

این الگوریتم درست نیست و در مساحت training ایام سی خود را برای تنبیه ای ایجاد کرد. این ایجاد شد.

$$P(y=+) = \frac{9}{11}$$

$$P(y=-) = \frac{2}{11}$$

$$P(\text{outlook}=\text{Rainy} | +) = \frac{3}{9}$$

$$P(\text{Rainy} | -) = \frac{1}{9}$$

$$P(\text{cool} | +) = \frac{4}{9}$$

$$P(\text{cool} | -) = \frac{1}{9}$$

$$P(\text{high} | +) = \frac{3}{9}$$

$$P(\text{high} | -) = \frac{4}{9}$$

$$P(\text{True} | +) = \frac{3}{9}$$

$$P(\text{True} | -) = \frac{1}{9}$$

$$\sum_{i=1}^4 P(x_i | +) P(+) = 0.10059$$

$$\sum_{i=1}^4 P(x_i | -) P(-) = 0.01347$$



AAAI, KDD, NIPS, ICML + UAI

نایوی بیجس نامه با استنبت است Naive Bayes \*

کنفرانس های لفچ پیشتر:

بزرگی scalability ~ داشتن سادگی آن Naive Bayes \*

جلسه ۱۴ : ۹v, A, Y.

مطالعه عمری: عیث NBC  $\rightarrow$  smoothing مطالعه شود. تل لایاس

: BC, NBC مفاسی

پارامترهای این است انت

اگر قصد طبقه بانیم که BC را با تحسین  $P(x|y)$ ,  $P(y)$  آوریم دیگر

چه تعداد امثال آن برای تحسین کتاب اعماق BC و NBC نیاز داریم؟

$$h: x \rightarrow \{0, 1\}$$

مثال: برای تحسین  $P(x|y)$  چه تعداد پارامتر نیاز است؟

پارامترهای سُنی

$$\cdot \theta_{ij} = P(x_i = x_i | y = y_j)$$

$$\langle x_1, x_2, \dots, x_n \rangle \quad \text{دروز فون دریم} \Rightarrow \begin{cases} i = 1 - n \\ j = 1 - 2 \end{cases}$$

$$\text{تفصیل: } BC \text{ حدوداً } 2 \times 2^n \text{ پارامتر نیاز دارد.} \\ \text{sample complexity: } O(2^n) \approx 2 \times (2^n - 1) \quad \text{ذین خبر}$$

برای تحسین  $P(x|y)$  چه پارامتر نیاز دارد؟ NBC

$$\theta_{ij} = P(\langle x_1, x_2, \dots, x_n \rangle | y_j)$$

$$= P(x_{i1} | y_j) P(x_{i2} | y_j) \dots P(x_{in} | y_j)$$

و تقریباً متساوی خواهد بود

$$\Rightarrow \text{sample complexity: } 2 \times n = O(n)$$

بعینی دلخواه است! از  $2^n$  به ضمیر جمعی طی می‌باشد

Subject: .....  
Year ..... Month ..... Date .....

$$\text{windy} = \{\text{True}, \text{False}\}$$

$$\text{class} = \{+, -\}$$

$$\left\{ \begin{array}{l} p(\text{windy} = \text{True} | +) = k_1 \Rightarrow p(\text{windy} = \text{False} | +) = 1 - k_1 \\ p(\text{windy} = \text{True} | -) = k_2 \Rightarrow p(\text{windy} = \text{False} | -) = 1 - k_2 \end{array} \right.$$

$O(n)$

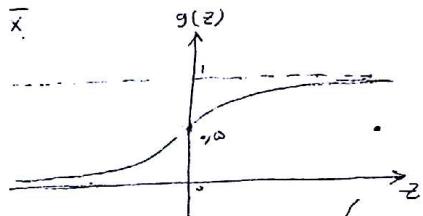
soft classifier  $\xrightarrow{(LR)}$  logistic Regression  $\xrightarrow[\text{classification}]{} \text{hard}$

لطفاً سه سالنگ طبقه بندی کریم و در فایل های خواهد بود  
نمای LR چنان نسبتاً آغاز شو جراحت سهی تریم میشود.

soft sigmoid یعنی logistic

$$\text{soft}: h(x) = w_0 + w_1 x_1 + \dots + w_n x_n = \bar{w}^T \bar{x}$$

$$\bar{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} x_0 = 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$



ستینه های عمل کرده (کسر از ۰، برای مبتدیان)

$$g(z) = \frac{1}{1 + e^{-z}}$$

تابع sigmoid رسم کرده عبارت خواهد بود زیرا  $(1 - e^{-z})$  درست ۰ تا ۱ بین است (خطی  $\times$  خطی)

$$g(z) > 0.5 \Rightarrow \text{when } z > 0$$

$$\left\{ \begin{array}{ll} g(\bar{w}^T \bar{x}) > 0.5 & \text{when } \bar{w}^T \bar{x} > 0 \\ g(\bar{w}^T \bar{x}) < 0.5 & \text{when } \bar{w}^T \bar{x} < 0 \end{array} \right.$$

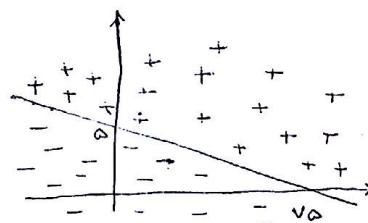
class +

class -

$$\bar{w}^T \bar{x} = w_0 + w_1 x_1 + w_2 x_2$$

$$\rightarrow -v_0 + x_1 + v_2 x_2 > 0$$

soft classifier نسبتی باید باشد



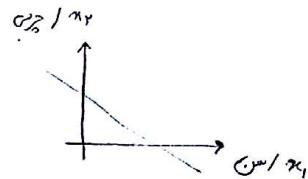
Temporal BN

برای این سیستم مولودی شد است، نامناسبان  
Dynamic Bayesian Network (DBN)

توالی (TBN)

تالی از: actuary / احتمال نزد کار اخراج را بررسی کند.

$$P(\text{death} | \bar{x}) = g(\bar{w}^T \bar{x}) = g(x_1, x_2) = g(\text{age} / \text{سن}, \text{cholesterol} / \text{کلسترول})$$



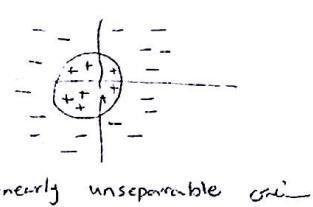
$$P(\text{death} | \bar{x}) = g(\bar{w}^T \bar{x}) \quad (\text{I})$$

$$g(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + \exp(-z)} \quad (\text{II})$$

$$\text{I} \rightarrow \text{II} \quad h(x) = \frac{1}{1 + e^{-\bar{w}^T \bar{x}}} \quad \text{parametrized by } \bar{w}, \bar{x}$$

$$\left\{ \begin{array}{l} P(\text{death} | \bar{w}, \bar{x}) = \frac{1}{1 + \exp(-\bar{w}^T \bar{x})} = P(Y=1 | \bar{x}, \bar{w}) \\ P(\text{alive} | \bar{x}, \bar{w}) = 1 - \frac{1}{1 + \exp(-\bar{w}^T \bar{x})} = \frac{\exp(-\bar{w}^T \bar{x})}{1 + \exp(-\bar{w}^T \bar{x})} = P(Y=0 | \bar{x}, \bar{w}) \end{array} \right.$$

$$\frac{P(Y=0 | \bar{x})}{P(Y=1 | \bar{x})} = \exp(-\bar{w}^T \bar{x}) \leq 1 \quad \ln \bar{w}^T \bar{x} \geq 0 \Rightarrow \text{class: +}$$



Linearly inseparable case

توالی کلاسیفایر LR \*

توالی سهندی سهندی LR \*

توالی پرپترن سهندی LR \*

LR  $\approx$  Perceptron

\* LR یک کمین تردی ناچ است -  $P(Y=1 | \bar{x}) / P(Y=0 | \bar{x})$  مستقیماً کمین نزد است  $\leftarrow$  آن درین  
درینی BN، حالا راعرض کیلدر  $P(\bar{w}|y)$  را تعیین نزدیکی دادیم  $P(y|\bar{x}) / P(y|\bar{x})$  - طور خوب سهندی کمین نزد است. (ABC)

(P(y),

(NBC)

- کمین اتفاقی داشت که در هر دو محدودیت مذکور نزد است

- کمین احتمالی هایی، آندریت های generative لخته میگشند

1) Generative & Discriminative classifiers: Naive Bayes

کمین در مدل خود را میگشند  $\leftarrow$  طبق محدودیت

\* Bias (NBC)  $\rightarrow$  Bias (LR)



بسته سوچ دارد پریزهای خود را از حجم مستقل باشند

$\Rightarrow$  میان ریل درست NBC از LR بسته است

False}



\* variance (NBC)  $<$  variance (LR)



آزادی عل لR بسته است.

\* در صورت محدود شدن داده های خود (خواهی)، LR عکس های NBC است -  $\rightarrow$  NBC

الدویجه نیز نیازی ندارد از حضور داده های خود که باید داشت تا variance آن نیازی نداشد باز هم مانند NBC داشت آنرا طبقاً کمتر NBC عکس های NBC است -

در صورت داده های طیفی، NBC کمتر LR است

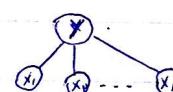
خواهد بود.

حلبیه IV: ۹۷, ۸, ۲۲

$h: X \rightarrow Y$

$\mathbb{R}^n \rightarrow \{0, 1\}$   $\xrightarrow{\text{پارامتر}} P(Y|X) \propto P(X|Y)P(Y)$

$$X = \langle x_1, x_2, \dots, x_n \rangle$$



NBC شرط

TAN  $\leftarrow$  Tree Augmented Naive Bayes

جنایل شناخته شده است ولی اینست شده عکس های NBC است -

در این الگوریتم هر دویزه کوتاه نه علاوه بر Y، ترسیم می کنند تا هم کوتاه شوند (جایگزین Parent) (استناد به NBC در صورتی که در NBC هر دویزه فقط یکی کوتاه باشد).

$$= \frac{1}{1 + e^{-z}}$$

> ۰, ۰

$$(\bar{w}^T \bar{x}) \geq 0$$

$$(\bar{w}^T \bar{x}) < 0$$

\* جنایل دستیابی و استنباطی مدل generative (joint generative) لغته های می شوند. نویج

Joint  $P(x,y)$  را مارکم می توانیم به  $P(x)P(y)$  تبدیل کنیم

synthetic

\*  $P(x,y)$  را مستقیماً می توانیم به  $P(x|y)P(y)$  تبدیل کنیم لغته های می شوند.

generative و discriminative (شان) این متد (اسن) نیز  $\rightarrow$  Ng & Jordan (2001) عرباً بتر از مثل های

discriminative می شوند.

$$\bar{x} = w_0 + w_1$$

$$\rightarrow -w_0 + w_1$$

دیگر دستیابی عل می شوند.



:<sup>نحو</sup> Statistical learning theory  $\rightarrow$  مطالعه معروف  $\rightarrow$  Vapnic گفت.

One should solve a classification problem directly & never solve a more general problem as an intermediate step (such as modeling  $P(x|y)$ ).

این انتسابی دسته‌های فرآیند: از کم تنبایه کرد زبان آن آنچه بازنشوی را دریاف است  
 classifi- نسبی-  $\left\{ \begin{array}{l} \text{در دو دسته از نسبی دسته} \\ \text{ترکیبی-} \end{array} \right.$  تنبایه کرد زبان آن آنچه بازنشوی است  
 دو دسته هر زبان)  $\rightarrow$  بارگذاری شده داده زبان بازنشوی کند داده زبان بازنشوی است  $\leftarrow$  (نمایشی داده داده داده است

ایرانی عکس های فرآیند

Latent Dirichlet allocation



{ Generative: BN, NBC, HMM, DBN, LDA, RBM

Deep belief network

Restricted Boltzmann machine

{ Discriminative: LR, DT, SVM, CRF, NN

منزه (متوجه) GNB

متوجه شویم  $x = (x_1, x_2, \dots, x_n)$ ,  $y$  داده شویم  $y = 0$  یا  $y = 1$

$$P(y=1|x) = \frac{P(x|y=1) P(y=1)}{P(x)} = \frac{P(x|y=1) P(y=1)}{P(x|y=1) P(y=1) + P(x|y=0) P(y=0)}$$

↓  
marginalize  $y$

$$= \frac{1}{1 + \frac{P(x|y=0) P(y=0)}{P(x|y=1) P(y=1)}} = \frac{1}{1 + \exp \left[ \ln \frac{P(x|y=0) P(y=0)}{P(x|y=1) P(y=1)} \right]}$$

(جذب)

$$= \frac{1}{1 + \exp \left( \ln \frac{P(x|y=0)}{P(x|y=1)} + \ln \frac{P(y=0)}{P(y=1)} \right)} \stackrel{CI}{=} \frac{1}{1 + \exp \left( \sum_{i=1}^n \ln \frac{P(x_i|y=0)}{P(x_i|y=1)} + \ln \frac{P(y=0)}{P(y=1)} \right)}$$

\*\*\*

(I)

ترجمی سرشناسی میراث ما نظریه مدل برای برداشت میانی ها مطابق شود.

Subject: ... Year: ... Month: ... Date: ...  
با نظریه برای برداشت میانی طبقه بندی ممکن است (الگوریتم رایسون ایمپلیکت)

$$\sum_{i=1}^n \ln \frac{P(X_i | Y=0)}{P(X_i | Y=1)} = \sum_i \ln \frac{\frac{1}{\sqrt{2\pi\delta_i}} \exp\left(-\frac{(X_i - \mu_{i0})^2}{2\delta_i^2}\right)}{\frac{1}{\sqrt{2\pi\delta_i}} \exp\left(-\frac{(X_i - \mu_{i1})^2}{2\delta_i^2}\right)} = \sum_i \frac{(X_i - \mu_{i1})^2 - (X_i - \mu_{i0})^2}{2\delta_i^2}$$

$$= \sum_i \frac{X_i (\mu_{i0} - \mu_{i1}) + \mu_{i1}^2 - \mu_{i0}^2}{2\delta_i^2} = \sum_i \left( \frac{\mu_{i0} - \mu_{i1}}{\delta_i^2} X_i + \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\delta_i^2} \right) \quad (II)$$

برآورد نزدیکی از ای تسانی دارای میانی

$$I, II \Rightarrow P(Y=1 | X) = \frac{1}{1 + \exp \left[ \sum_{i=1}^n \left( \frac{\mu_{i0} - \mu_{i1}}{\delta_i^2} X_i + \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\delta_i^2} \right) + \ln \frac{P(Y=0)}{P(Y=1)} \right]}$$

$$P(Y=1) = \pi \\ P(Y=0) = 1 - \pi \quad \left\{ \begin{array}{l} \Rightarrow \\ = \frac{1}{1 + \exp \left( \sum_{i=1}^n w_i X_i + w_0 \right)} \end{array} \right.$$

where  $w_i = \frac{\mu_{i0} - \mu_{i1}}{\delta_i^2}, \quad w_0 = \ln \frac{(1-\pi)}{\pi} + \sum_{i=1}^n \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\delta_i^2}$

Deci

$$P(Y=0 | X) = 1 - P(Y=1 | X)$$

میں ~~بایو~~ درخت تسانی دارای میانی نزدیکی GNB میں نظری است (یہ کم دستributed حل است)

میں چیز خوب خنای تھیم سمجھ کر است، دراون حفظ میں صرف است.

9V, 8, 9 : 19 جلسہ

### Support Vector Machine(s)

ترجمی Vladimír Vapnik طبقہ

ویکی عربی زبان خاندانwidest street approach

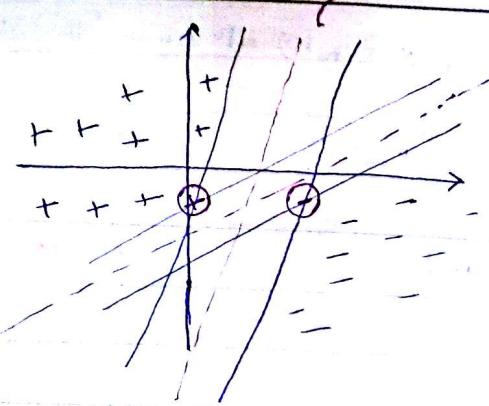
$$h : x \rightarrow \{-1, +1\}$$

mathematical convenience

نتیجہ درجہ سادسازی برابر



$$1 + \text{margin} - 1$$



هدف فرضی برای خیال می‌باشد  
هدف پیش‌گزین خط حالت‌های ساده‌تر را به صورت ایجاد کند.

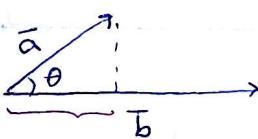
ساختاری بروی نیم‌دایریست در دامنه آنکاره است.

Decision Rule:  $\bar{w} \cdot \bar{x}_{\text{test}} + b$

نکته: مادری همچویی داشتیم. نقطه  $\bar{x}$  را با معنی  $w \cdot \bar{x} - b$  خیال نمودیم.

برای  $\bar{x}$  بردار  $w$  تصوری نیم  $\leftarrow$  صفر (اطی)  $\rightarrow$  test

محول است



$$a \cdot b = |a||b| \cos \theta$$

Decision Rule:  $\bar{w} \cdot \bar{x}_{\text{test}} + b \geq 0 \Rightarrow +$   
else -

$$b = -c$$

داریست

$$\textcircled{1} \quad \bar{w} \cdot \bar{x}_{\text{test}} + b \geq 0 \rightarrow \oplus \quad \text{else } \ominus$$

نماینده دستگاه

جهت سمت می‌باشد  $y=1$  یعنی  $y=1$  می‌باشد

\* \*

$$\begin{cases} \bar{w} \cdot \bar{x}_+ + b \geq 1 & (y=1) \\ \bar{w} \cdot \bar{x}_- + b \leq -1 & (y=-1) \end{cases}$$

$$\begin{cases} \frac{\bar{w} \cdot \bar{x}_+ + b}{xy} - 1 \geq 0 \\ \frac{\bar{w} \cdot \bar{x}_- + b}{xy} - 1 \geq 0 \end{cases}$$

(-1, +)

ساختاری آندر

$$\bar{w}' \cdot \bar{x} + b' = k$$

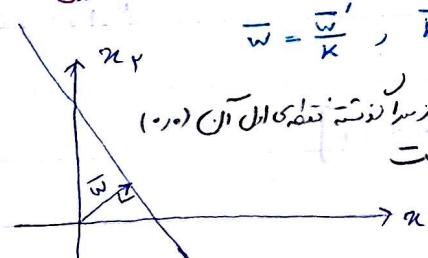
$$\bar{w} = \frac{\bar{w}'}{k}, \quad b = \frac{b'}{k}$$

ساختاری آندر  $\bar{w} \cdot \bar{x} + b = 0$  - خواهد بود

اگر خواهیست از خط ایجاد شد

$$x_2 = \frac{w_2}{w_1} x_1$$

$$\text{ضدیلیت ها: } \frac{w_2}{w_1} \times \frac{-w_1}{w_2} = -1$$



$$\bar{w} \cdot \bar{x} + b = 0 \rightarrow w_1 x_1 + w_2 x_2 + b = 0 \Rightarrow x_2 = \frac{-w_1}{w_2} x_1 - \frac{b}{w_2} *$$

ساختار خط

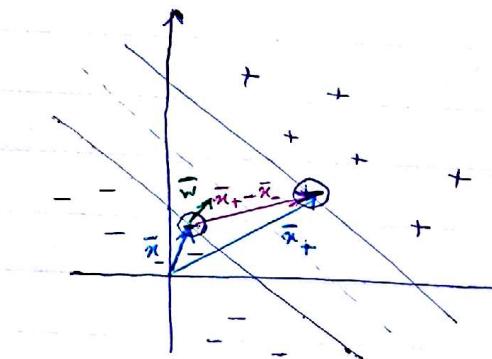
$$\tan(\alpha - \beta) = \frac{\tan \alpha - \tan \beta}{1 + (\tan \alpha \tan \beta)} = \infty = \tan 90^\circ$$



(\*)  $y_i(\bar{w} \cdot \bar{x}_{train} + b) - 1 \geq 0 \quad \xrightarrow{\text{نحو}} \quad y_i(\bar{w} \cdot \bar{x} + b) - 1 \geq 0$

(اندیشه ای اینزیر) بردارهای شناسان

①  $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0 \quad (\text{حدودت})$



(از سری به کم تر نکه از SV می‌شود . از جمله جای تغییر SV مخفی  
صلح کالیلم)

بردار از محدود شرطی خواهد شد . در جبر خالی می‌بینیم که بردار را می‌توان بخطی مدارل آن (معاری دهم  
(نوازه با بردار اولی) جایز نیست . بنابراین  $\bar{w}$  را از تغییرات شفته شده سه  
می‌کنیم .

width =  $(\bar{x}_+ - \bar{x}_-) \cdot \frac{\bar{w}}{\|\bar{w}\|} = \frac{\bar{x}_+ \cdot \bar{w} - \bar{x}_- \cdot \bar{w}}{1-b} \times \frac{1}{\|\bar{w}\|} = \frac{1}{\|\bar{w}\|}$

برای این بدل ایندریک بردار به نرم  $\bar{w}$  تقسیم دویم

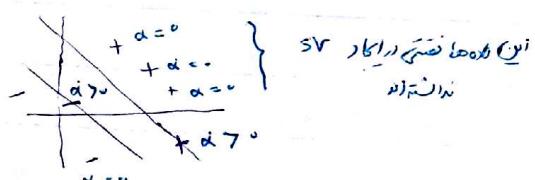
حل معقول می‌باشد که  $\bar{w}$  را  $\max$  کند خود  $\leftarrow$  عرض خالی

minimize  $\|\bar{w}\| \sim \min \frac{1}{\gamma} \|\bar{w}\|^2$

② : عرض خالی  $\frac{1}{\|\bar{w}\|} \xrightarrow{\text{goal}} \min \frac{1}{\gamma} \|\bar{w}\|^2$  نتیجه :

تابع فرازه  $L(\bar{w}, b) = \frac{1}{\gamma} \|\bar{w}\|^2 - \sum_i \alpha_i [y_i(\bar{w} \cdot \bar{x}_i + b) - 1]$

نتیجه داده های خالی از این ایده SV یعنی کلمه



تابع اول و دوم :



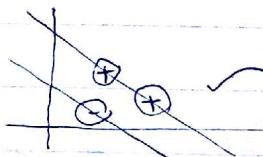
$$\|\bar{w}\|^2 = \bar{w} \cdot \bar{w} = \bar{w}^T \bar{w}$$

می‌اسطراحته فیزیک

دیره، است

$$\frac{\partial L(\bar{w}, b)}{\partial \bar{w}} = \boxed{\bar{w} - \sum_i \alpha_i y_i \bar{x}_i = 0}$$

$$\frac{\partial L(\bar{w}, b)}{\partial b} = \boxed{-\sum_i \alpha_i y_i = 0} \rightarrow \text{حده وزن دار. مایع تجسس را شل} \rightarrow \text{است} \rightarrow \text{کی دهد}$$



میان درها + مایع تجسس

بار است باره

بالا بزرگتر شود

$$L(\bar{\alpha}) = \frac{1}{2} \left( \sum_i \alpha_i y_i \bar{x}_i \right) \left( \sum_j \alpha_j y_j \bar{x}_j \right) - \left( \sum_i \alpha_i y_i \bar{x}_i \right) \left( \sum_j \alpha_j y_j \bar{x}_j \right)$$

$$- \sum_i \alpha_i y_i b + \sum_i \alpha_i$$

$$= \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{\bar{x}_i \cdot \bar{x}_j}$$

نقطه اینست بردار

تجسس اسفل

ط اینجا جزویت برای محاسبه training

$$\text{test نجاش: } \sum_i \alpha_i y_i \bar{x}_i \cdot \bar{x}_{\text{test}} + b > 0 \Rightarrow \text{class: +}$$

$$\text{sgn} \left( \sum_i \alpha_i y_i \bar{x}_i \cdot \bar{x}_{\text{test}} + b \right)$$

در معادلات فیزی می‌بینیم که برآوردهای بینش است دو صفت زیر را دارد که تابع نیزیم.

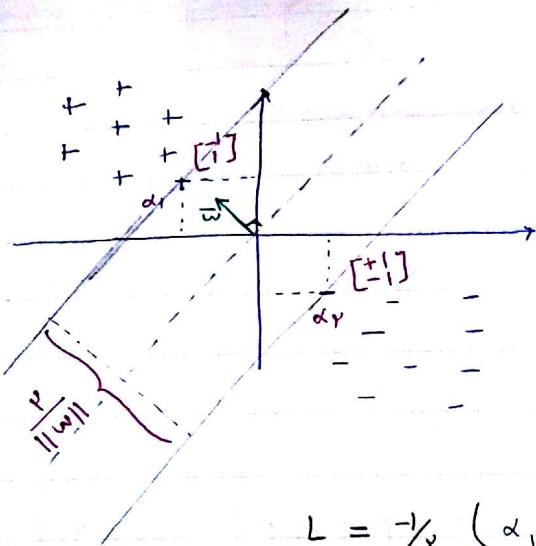
صفت اول: مقدارهای عرضی و عرضی خیان / دهای

$$\begin{aligned} \bar{w} \bar{x} + b &= 1 \\ \bar{w} \bar{x} + b &= 0 \end{aligned}$$



$$L = -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j + \sum_{i=1}^m \alpha_i \quad (\alpha_i \geq 0)$$

سؤال: مطابق است به چهار و بیان



ارحل دستی - همچنین پیشنهادهای سی ای سازند  
تئیفی دیگر نیست. ارحل با ماشین با استاد بر بود آنها برای آنها تئیفی خواهد

جیت  $\bar{w}$  را هم در بحث مثبت ها در نظر نمی دیسیم. اگر به بحث نزدیک  
در نظر بگیریم در معادلات که نشان از اضافه آید

$$L = -\frac{1}{2} (\alpha_1 \times \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \alpha_2 \times \begin{bmatrix} +1 \\ -1 \end{bmatrix}) \cdot (\alpha_1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \alpha_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}) + \alpha_1 + \alpha_2$$

$$= -\frac{1}{2} \begin{bmatrix} -\alpha_1 - \alpha_2 \\ \alpha_1 + \alpha_2 \end{bmatrix} \cdot \begin{bmatrix} -\alpha_1 - \alpha_2 \\ \alpha_1 + \alpha_2 \end{bmatrix} + \alpha_1 + \alpha_2 = -\frac{1}{2} \times \times \times (\alpha_1 + \alpha_2)^2 + \alpha_1 + \alpha_2$$

$$= -\alpha_1^2 - 2\alpha_1 \alpha_2 - \alpha_2^2 + \alpha_1 + \alpha_2 \quad \text{رسانی اولی سی ای دستی}$$

که متفاوت بینی سازی است. به مرور زمان می توانند حل نشون

رسانی اولی سی ای دستی

$$\frac{\partial L}{\partial \alpha_1} = -2\alpha_1 - 2\alpha_2 + 1 = 0$$

$$\frac{\partial L}{\partial \alpha_2} = -2\alpha_1 - 2\alpha_2 + 1 = 0$$

$$\sum_i \alpha_i y_i = 0 \Rightarrow \sum \alpha_+ = \sum \alpha_- \Rightarrow \alpha_1 = \alpha_2$$

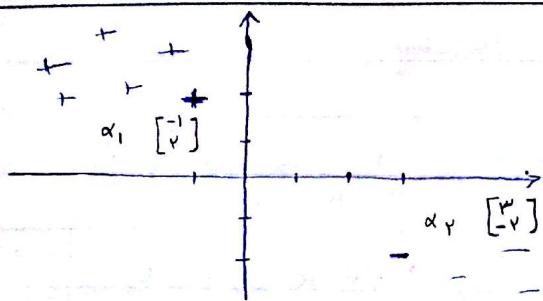
امروزه معادلات از توابع محض معتبره این ماده را حل نمی کنند.

$$\bar{w} = \sum_i \alpha_i y_i \bar{x}_i = \alpha_1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \alpha_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$$y_i (\bar{w} \cdot \bar{x} + b) - 1 = 0 \xrightarrow{\substack{y_i = 1 \text{ یا } y_i = -1 \\ \text{نشانی}} \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} + b - 1 = 0 \Rightarrow b = 0}$$

نشانی خوب است





$$L = \frac{1}{4} (\alpha_1 \begin{bmatrix} -1 \\ Y \end{bmatrix} - \alpha_y \begin{bmatrix} Y \\ -1 \end{bmatrix}) \cdot (\alpha_1 \begin{bmatrix} -1 \\ Y \end{bmatrix} - \alpha_y \begin{bmatrix} Y \\ -1 \end{bmatrix}) + \alpha_1 + \alpha_y$$

$$\Rightarrow L = \frac{1}{4} \begin{bmatrix} -\alpha_1 - Y\alpha_y \\ Y\alpha_1 + Y\alpha_y \end{bmatrix} \cdot \begin{bmatrix} -\alpha_1 - Y\alpha_y \\ Y\alpha_1 + Y\alpha_y \end{bmatrix} + \alpha_1 + \alpha_y$$

$$= -\frac{1}{4} \left[ (-\alpha_1 - Y\alpha_y)(-\alpha_1 - Y\alpha_y) + (Y\alpha_1 + Y\alpha_y)^2 \right] + \alpha_1 + \alpha_y = -Y\alpha_1^2 - Y\alpha_y^2 - V\alpha_1\alpha_y + \alpha_1 + \alpha_y$$

$$\alpha_1^2 + 4\alpha_1\alpha_y + 9\alpha_y^2 \quad 4\alpha_1^2 + 4\alpha_y^2 + 1\alpha_1\alpha_y$$

$$2\alpha_1^2 + 12\alpha_1\alpha_y + 10\alpha_y^2$$

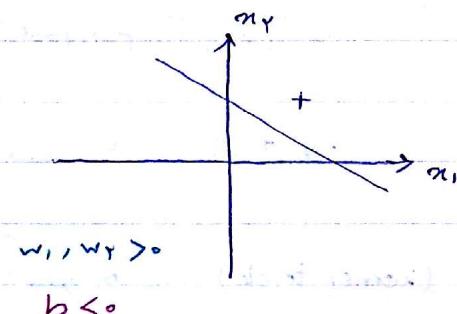
$$\frac{\partial L}{\partial \alpha_1} = -2\alpha_1 - V\alpha_y + 1 = 0 \quad \left. \begin{array}{l} + \\ \Rightarrow 4\alpha_1 + 10\alpha_y = 1 \end{array} \right\}$$

$$\frac{\partial L}{\partial \alpha_y} = -12\alpha_y - V\alpha_1 + 1 = 0 \quad \left. \begin{array}{l} \downarrow \\ \text{جواب ممکن نیست} \\ \text{دعاوی اینکه } \alpha_1 = \alpha_y \text{ اثبات نداریم} \\ \text{سازلی اول را با هم بخواهیم} \end{array} \right\}$$

$$\Rightarrow \alpha_1 = \alpha_y = \frac{1}{14}$$

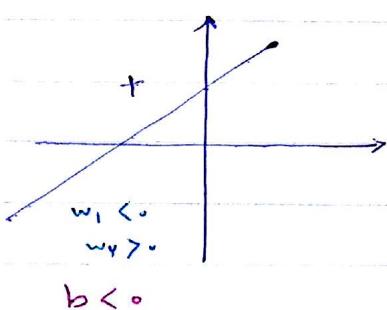
$$\alpha_1 = \alpha_y$$

حالات زیر:  $\Rightarrow$  حلول



$$w_1, w_y > 0$$

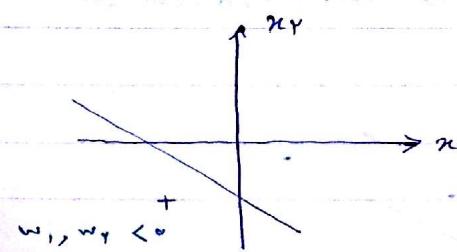
حالت اول:  $w_1, w_y > 0$



$$w_1 < 0$$

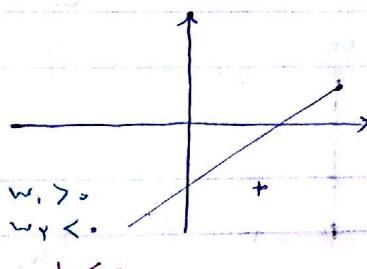
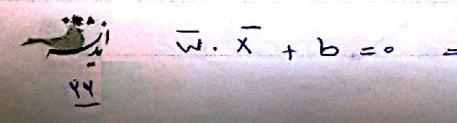
$$w_y > 0$$

$$b < 0$$



$$w_1, w_y < 0$$

حالت دوم:  $w_1, w_y < 0$



$$w_1 > 0$$

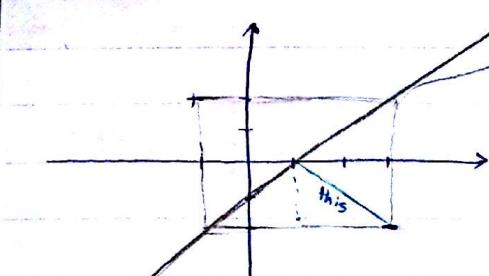
$$w_y < 0$$

$$b < 0$$



$$w_1 \bar{x}_1 + w_y \bar{x}_y + b = 0 \Rightarrow w_1 x_1 + w_y x_y + b = 0 \Rightarrow x_y = -\frac{w_1}{w_y} x - \frac{b}{w_y}$$

کلیت b ناچاری است تا میانگین در میان دو میانگین داشت  $w_1 + w_y$  بخوبی شود



نصف عکس خیان

$$this = \sqrt{w_1^2 + w_2^2} = \frac{1}{\|w\|} \Rightarrow \|w\| = \frac{1}{\sqrt{w_1^2 + w_2^2}}$$

$$= \sqrt{w_1^2 + w_2^2} = \sqrt{1} \Rightarrow k = \frac{1}{\sqrt{1}} \Rightarrow \begin{cases} w_1 = -\frac{1}{\sqrt{1}} \\ w_2 = \frac{1}{\sqrt{1}} \\ b = \frac{1}{\sqrt{1}} \end{cases}$$

جهت همای خیان (آن خطا را

شیدم (اتجاه مربوطة)

سادهی  $w$

خط دیگر سیم  
از  $(2, -2)$  و  $(-1, 2)$  عبور کرد

$$\Rightarrow x_2 = x_1 - 1$$

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{b}{w_2}$$

چون  $w_1 > 0$  و  $w_2 < 0$

$$\frac{w_1}{w_2} = -1 \quad , \quad \frac{b}{w_2} = 1 \quad \Rightarrow \quad \begin{cases} w_1 = -k \\ w_2 = k \end{cases} *$$

یعنی به محاسبه داده است چون در ماده حل ماتریس ابتدا داده ای داشت بی اوریم  $w$ ,  $b$  را محاسبه سین دی از

محبایی داده را ب داشت اوریم:

$$\begin{aligned} \alpha_1 = \alpha_2 \\ \begin{bmatrix} -\frac{1}{\sqrt{1}} \\ \frac{1}{\sqrt{1}} \end{bmatrix} = \alpha_+ \begin{bmatrix} -1 \\ 2 \end{bmatrix} - \alpha_- \begin{bmatrix} 2 \\ -2 \end{bmatrix} \quad \left\{ \begin{array}{l} \alpha_1 = \alpha_2 \\ \alpha_1 + \alpha_2 = \frac{1}{\sqrt{1}} \\ \alpha_1 - \alpha_2 = \frac{1}{\sqrt{1}} \end{array} \right. \quad \Rightarrow \alpha_1 + \alpha_2 = \frac{1}{\sqrt{1}} \quad \left\{ \begin{array}{l} \alpha_1 = \alpha_2 \\ \alpha_1 + \alpha_2 = \frac{1}{\sqrt{1}} \end{array} \right. \quad \Rightarrow \alpha_1 = \alpha_2 = \frac{1}{\sqrt{1}} \end{aligned}$$

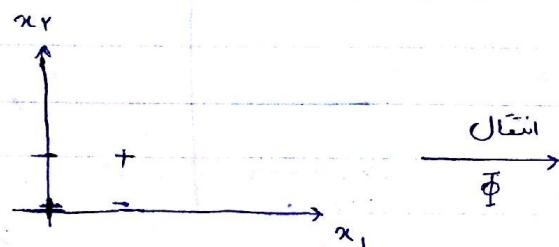
اولین دفعه از آخر مقاله آمد

جلسه ۲۱ : ۹/۹/۱۱

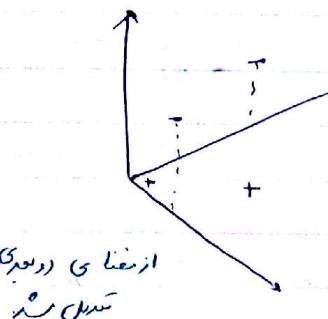
سرد هست (kernel trick)

linearly inseparable

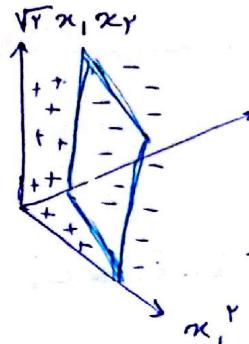
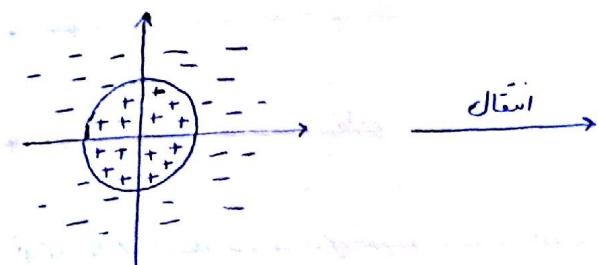
که باعث SVM برای حل مسئله جوانابی خواهد شد



انتقال  
 $\Phi$



از اینجا رویکری انتقال می‌داریم که مسئله ساده‌تر و به سهایی جای پذیر خواهد شد



متغیری ممکن است پیدا نرود این  
نه متغیری ناممکن است ایجاد نرود

آن نهایا نمایی است ( $x_1^*, x_2^*, x_1^* \cdot x_2^*, x_1^* \cdot x_2^* + x_2^* \cdot x_1^*$ )

استقل ممکن است با ابعاد مستقل محدود داشت به عبارت ایجاد ابعاد ایجاد حتم ممکن است.

$$\Phi: \mathbb{R}^n \rightarrow \mathbb{Z}$$

$$\bar{x} \longrightarrow \Phi(\bar{x})$$

$$(x_1, x_2) \longrightarrow (x_1^*, x_2^*, x_1^* \cdot x_2^*, x_1^* \cdot x_2^* + x_2^* \cdot x_1^*)$$

برآوردهای پیشانی

$$\left\{ L = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j [\bar{x}_i \cdot \bar{x}_j] + \sum_i \alpha_i \right\} \text{train سطحی مارکوار}$$

$$h(\bar{x}_{\text{test}}) = \text{sgn} \left( \sum_i \alpha_i y_i [\bar{x}_i \cdot \bar{x}_{\text{test}}] + b \right)$$

پیش از برآوردهای پیشانی

①

$$\underbrace{\Phi(\bar{x})}_{\Phi(\bar{x})} \rightarrow L = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j [\Phi(\bar{x}_i) \cdot \Phi(\bar{x}_j)] + \sum_i \alpha_i \quad ②$$

نیازی نیست فضای جدید را بسازیم  $\leftarrow$  داشت فضای داخلی آنها برای این کار طبقه است.

رنیازی نیست  $\Phi$  را باستخوان  
تابع

$$L = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j [K(\bar{x}_i, \bar{x}_j)] + \sum_i \alpha_i \quad ③$$

در واقع داده را ایجاد کرد

از محاسبی ① مستقیماً بمحاسبی ③ کوچک

و توابع این عبارت را kernel نامند

\* می‌توان یک فضای معمولی (valid kernel) خوب باشی در بردار را در فضای (some space) معتبر کرد.

که برای همه تابع خاصی بگویی می‌توان خاصیتی داشت (ترز دهنده)

۲۳

\* بُعْدِ حَسَنَةِ خَاصَّتِ جَابِجَانِيَّةِ

\* از لَغْيَةِ تَلْكِيجَةِ مَدْرَسَةِ دِيرْطَارِ:

آنچه تکذیب استناد شده از نوع خود برداشته باشند لفظ در این مسئله بوده است.

Linear kernel:  $(\bar{x}_i \cdot \bar{x}_j + c)$

این نظر مبتداً در کتاب های عصیانی

polynomial kernel:  $(\bar{x}_i \cdot \bar{x}_j + c)^d$

RBF Kernel:  $\exp\left(-\frac{\|\bar{x}_i - \bar{x}_j\|^2}{\gamma^2}\right) \equiv \exp\left(-\gamma \|\bar{x}_i - \bar{x}_j\|^2\right)$

Gaussian Radial Basis Function

$$\gamma = \frac{1}{\sqrt{2}}$$

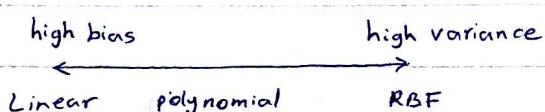
در این مدل فرض داشتیم متساوی می شود. به عنوان مبنای انتشاری صور خوب داشتیم RBF را انتخاب کرد (با خواسته Latex)

کلیه محتوای

Inverse multiquadratic:  $K(\bar{x}_i, \bar{x}_j) = \frac{1}{(\|\bar{x}_i - \bar{x}_j\|^2 + c)^k}$

Sigmoid kernel:  $K(\bar{x}_i, \bar{x}_j) = \tanh(\bar{x}_i \cdot \bar{x}_j + c)$

RBF موجزتر و ساده تر است. با سطح میانی دارد. ... میان آنست که بغضنهای بزرگ است بدینها می بینیم. (دیگر برخی متنها نیزی انتشار از خصیصه ای دارند)



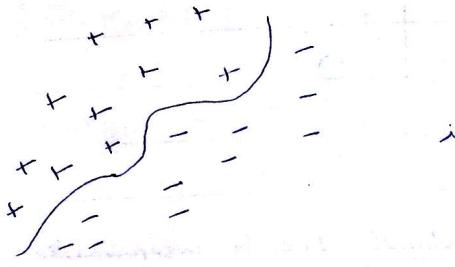
time of learning: linear < poly < RBF

ability to fit any data: " "

risk of overfitting: " "



\* فن رانی جی کرنی باری کے مدد سے برداشت و باعث تراویح مختلف طرز سادہ ب پیغمبر امّاک دستیم با پیغمبر نلام میں عکس  
برداشت ہے۔ (از طرف تفت، خال و ...)



\* هر دو فن است با حالات margin نداره حالا از حرم جدا ہیں  
کے درجاتی مختصر نظر SVM

حلیہ ۹۷، ۹، ۱۳ : ۴۴

$$h: X \rightarrow \{-1, 1\}$$

$$\Phi: X \rightarrow Z$$

$$\bar{x}, \bar{x}' \in X \rightarrow \underbrace{\Phi(\bar{x}) \cdot \Phi(\bar{x}')}_{K(\bar{x}, \bar{x}')}$$

کے نتیجے میں  $\Phi$  میں  $\bar{x}_x = \begin{cases} x_1 \\ x_2 \end{cases}$  میں مثال

$$\Phi(\bar{x}) = [1, x_1, x_2, x_1^2, x_2^2, x_1 x_2]^T$$

$$K(\bar{x}, \bar{x}') = \Phi(\bar{x}) \cdot \Phi(\bar{x}') = 1 + x_1 x'_1 + x_2 x'_2 + x_1^2 x'^2_1 + x_2^2 x'^2_2 + x_1 x_2 x'_1 x'_2$$

وہی مثال تابع  $\Phi$  را میں طبقہ میں نہیں کر سکتے ہیں اس کے بعد (لائسنس پڑھنے پر) نہیں کر سکتے ہیں تابع  $\Phi$  کی تابعیت کیا ہے؟

کے نتیجے میں  $\Phi$  کو معرفہ کیا گیں: polynomial kernel (ڈسکریپٹو)

$$K(\bar{x}, \bar{x}') = (1 + \bar{x} \bar{x}')^p$$

$$= (1 + x_1 x'_1 + x_2 x'_2)^p = 1 + x_1^p x'_1^p + x_2^p x'_2^p + p x_1 x'_1 + p x_2 x'_2 + p x_1 x'_1 x_2 x'_2$$

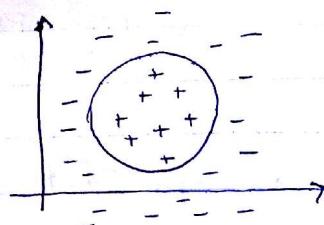
اگر خواهیم از دوی خرمن تابع کریں غیرہی خرمن تابع کریں ما پیرا کیسیم: زریں خوب (اصلی) دردراست، عکس عکسی کیسیم



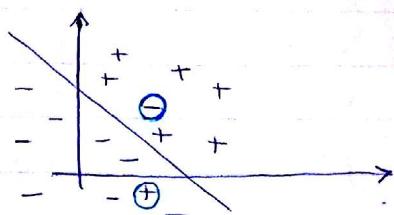
$$\Phi(\bar{x}) = [1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2]^T$$

کے سب سے درجہ عالی ترین

### Soft-margin SVMs



seriously linearly inseparable



slightly linearly inseparable

→ حلی است برای این سیستم kernel

پرانتز ۸ (برای هندسه  
گرانشی مفهوم استفاده شود)

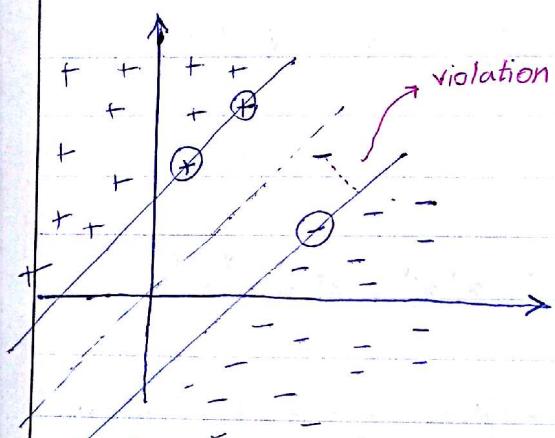
soft-margin SVMs ; ۱۶

در سال ۱۹۹۵ معرفی شد

C پارامتر

\* در اینجا مسأله از کار C مفهوم شده، بنابراین حالت را در نظر گرفته ایم

RBF گروه



$$y_i (\bar{w} \cdot \bar{x}_i + b) \geq 1 \quad \text{این رابطه دیر صفحه}$$

slack variable  $\xi_i$  برای هر دو violation می باشد

$$y_i (\bar{w} \cdot \bar{x}_i + b) \geq 1 - \xi_i$$

$$\text{total violation} = \sum_{i=1}^m \xi_i \geq 0$$

$$\min \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^m \xi_i$$

subject to :

$$y_i (\bar{w} \cdot \bar{x}_i + b) \geq 1 - \xi_i \quad \& \quad \xi_i \geq 0 \quad (i=1 \dots m)$$

$C \rightarrow \infty$  : over fitting می شود و چنانچه می تواند

$C \rightarrow 0$  : under fit

مسأله دشوار شد و نتیجه ندارد



نمایه درستی SVM برای hard-margin SVM

$$L(\bar{w}, b, \bar{x}, \bar{\xi}_i, \bar{\beta}) = \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^m \xi_i - \sum_{i=1}^m \alpha_i [y_i(\bar{w}\bar{x} + b) - 1 + \xi_i] - \sum_{i=1}^m \beta_i \xi_i$$

subject to:  $\alpha_i, \beta_i \geq 0$

$$\begin{aligned} \frac{\partial L}{\partial \bar{w}} &= \bar{w} - \sum_{i=1}^m \alpha_i y_i \bar{x}_i = 0 \\ \frac{\partial L}{\partial b} &= - \sum_{i=1}^m \alpha_i y_i \\ \frac{\partial L}{\partial \xi_i} &= C - \alpha_i - \beta_i = 0 \end{aligned}$$

برای  $\bar{w}$  و  $b$  از مقدار مینیموم است  
برای  $\xi_i$  از مقدار مکنیموم است

$$\left. \begin{array}{l} C - \alpha_i = \beta_i \\ \beta_i \geq 0 \end{array} \right\} \rightarrow \boxed{0 \leq \alpha_i \leq C}$$

کردن باعث برای  $\alpha_i$  می شود

با خواص لامباد است:  $\min \bar{\xi}_i, b, \bar{w}$

$\max \bar{\beta}_i, \bar{\alpha}_i$

$$\alpha_i \geq 0 + \xi_i \times \sum_{i=1}^m (C - \alpha_i - \beta_i)$$

لطفاً می بینید

عنوان ماتریسی میان میان  $(C - \alpha_i) \beta_i$  که در soft-margin hard-margin میان میان را به نسبت

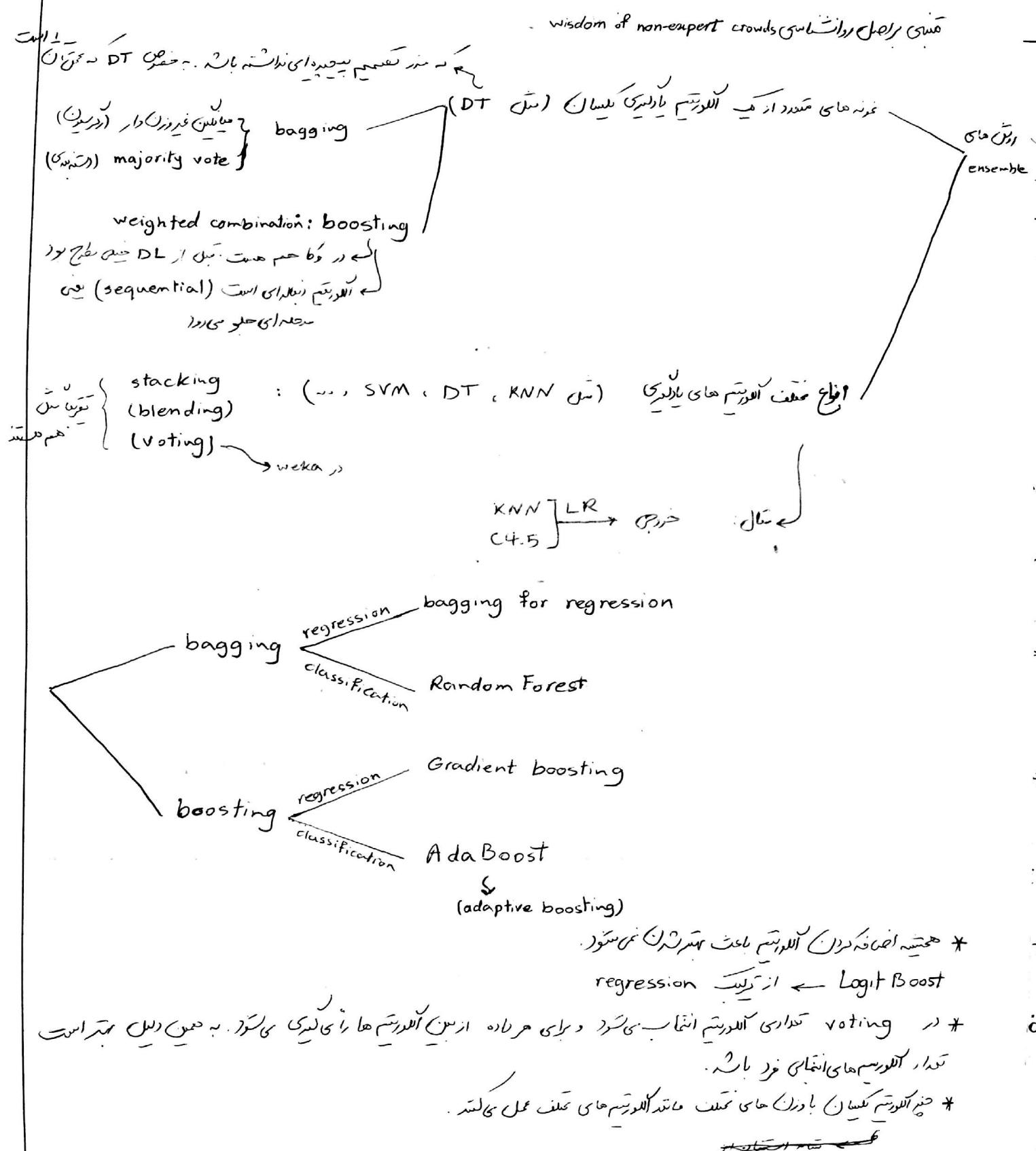
جلسه ۹، ۹، ۱۰

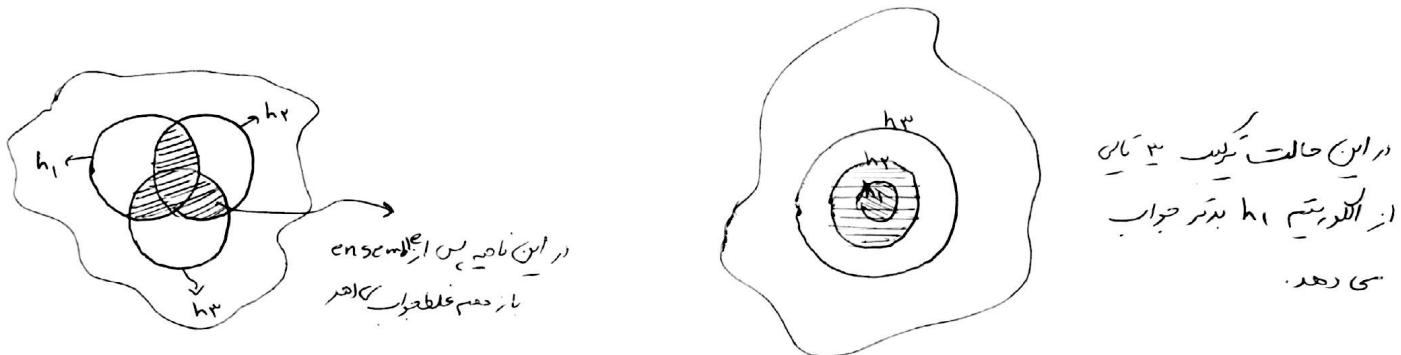
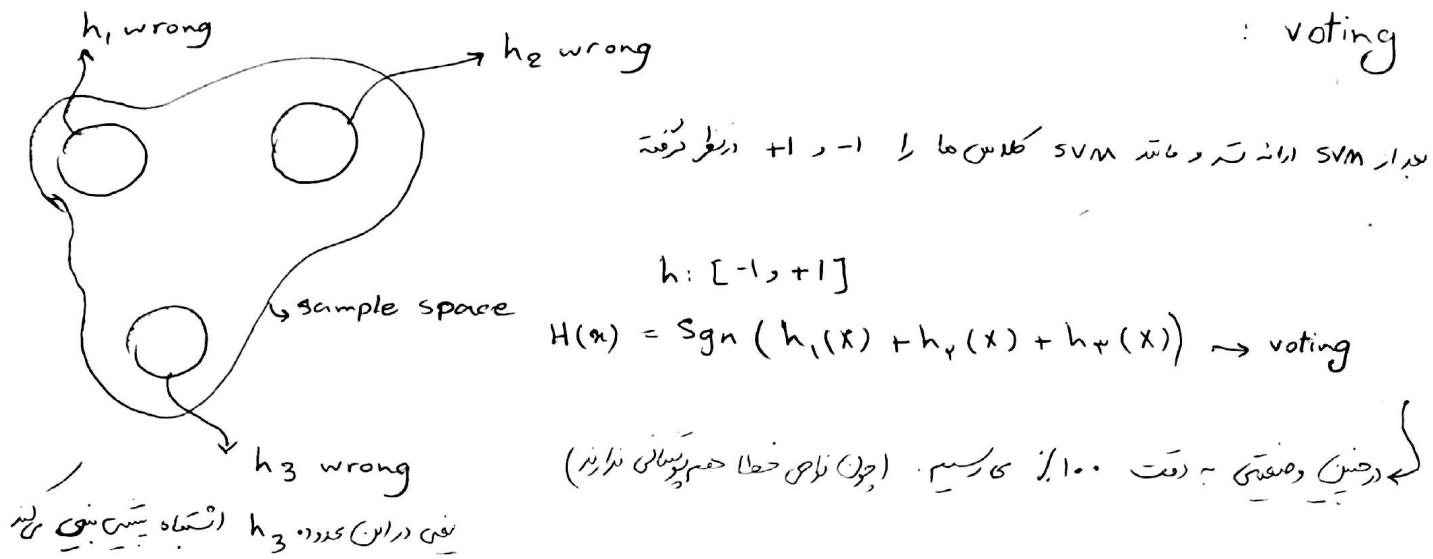
Francis Galton ← عثثهای دریلریک، نند داده است: آشنازی اثبات آنکه سرم توانی ادوسی جانعه را نداشته

wisdom of crowds ← عقل جمیع ← عقل جمیع ← Nature → Galton

نمایه برای نیکوین دل کاو نزد در مقابله ساینس نظر افراد بسیار به راقعیت نزدیکی بود

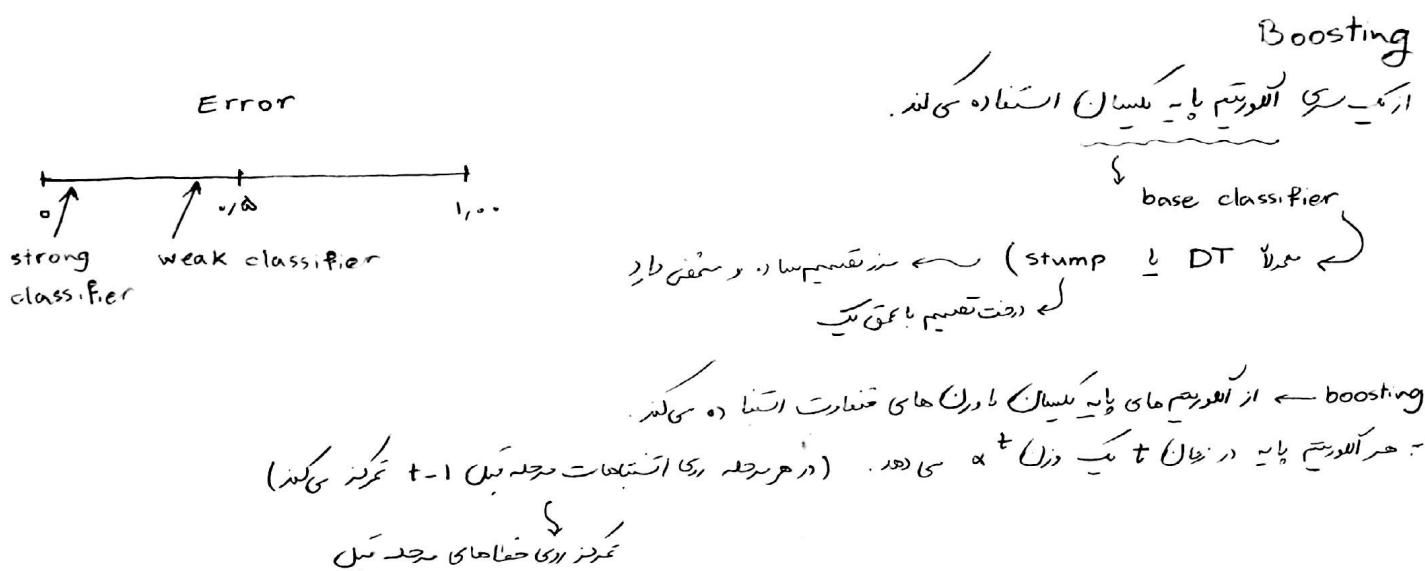
## Golkon ← Ensemble Methods





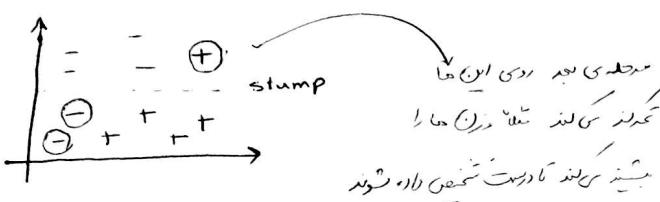
سؤال: چه ترسیم‌هایی بفرمایی ترکیب خوب الگوریتم از گویی اولین‌ها بدر عمل نمایند (اصل تاییدشونه) که سعی داشت در مورد خطایها

حلیمه: ۹۷, ۹, ۹٪



{ Yael Freund & Robert Schapire (97) ← مقاله بسیار خوب و معنادار ← Gödel Prize (2003)

$\leftarrow$  AdaBoost



$$\sum_i w_i^t = \frac{1}{m} \Rightarrow \sum_i w_i^t = 1$$

$$\in^t = \sum_i w_i^t \Rightarrow \in^t = \sum_i w_i^t$$

(Incorrect) (Incorrect)

SVM <

$$h: X \rightarrow \{-1, +1\} \quad \text{; classifier}(j),$$

$$H(\bar{x}) = \text{Sgn} (\alpha' h'(x) + \alpha'' h''(x) + \dots)$$

لیکن رای نیز در مدار  $\omega$  و  $\omega_0$  می باشد

$$w_i^{t+1} = \frac{1}{Z^t} \times w_i^t \times \exp(-\alpha^t h^t(x) y(x))$$

$$\alpha^t = \frac{1}{\gamma} \ln \frac{1 - e^t}{e^t}$$

تصویر تقسیم شوند. (تعادل اسلوب‌ها) classifier voting کا طریقہ در ریاضیات و آن دینامیکس با دریں های و بالیدیر لیب میں زیر دریاس سے

$$\boxed{w_i^{t+1} = \frac{w_i^t}{z^t} \times \begin{cases} \sqrt{\frac{e^t}{1-e^t}} & h^t(x_i) = y_i \text{ (correct)} \\ \sqrt{\frac{1-e^t}{e^t}} & h^t(x_i) \neq y_i \text{ (incorrect)} \end{cases}}$$



$$z^t = \left( \sum_i w_i^t \right) \times \sqrt{\frac{e^t}{1-e^t}} + \left( \sum_i w_i^t \right) \times \sqrt{\frac{1-e^t}{e^t}} = \gamma \sqrt{e^t(1-e^t)}$$

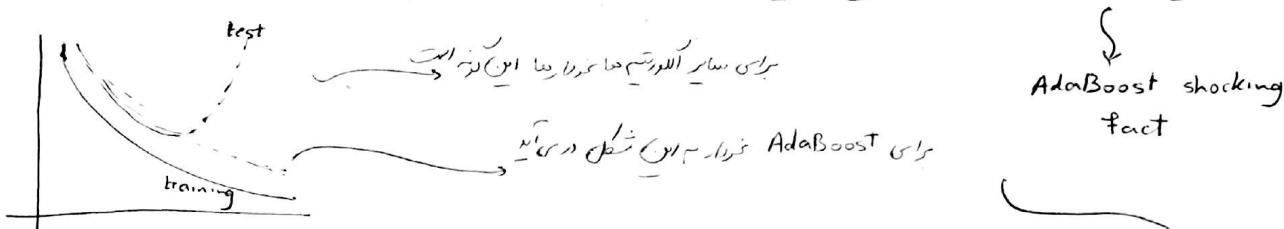
$\checkmark$  حالتی که  $z^t$  را رابطه

$$\begin{cases} w_i^{t+1} = \frac{w_i^t}{\gamma} \times \frac{1}{1-e^t} & \text{correct} \\ w_i^{t+1} = \frac{w_i^t}{\gamma} \times \frac{1}{e^t} & \text{incorrect} \end{cases}$$

نماینده از ترتیب انتخاب شده می‌شود  
که ممکن نیست. و مرتبط با آن داده  
آن معتبر نیست.

در هر سمعه داده‌ها توزیس مارکوچ فرآیند می‌گذرد

براساس نتایج عجیب داده‌ها این overfit نامیده داده‌ها باید از آن دور شوند.



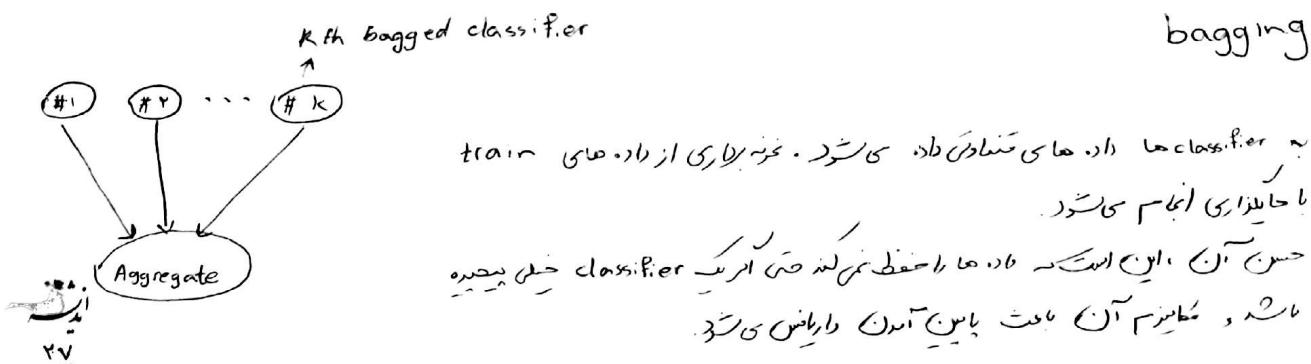
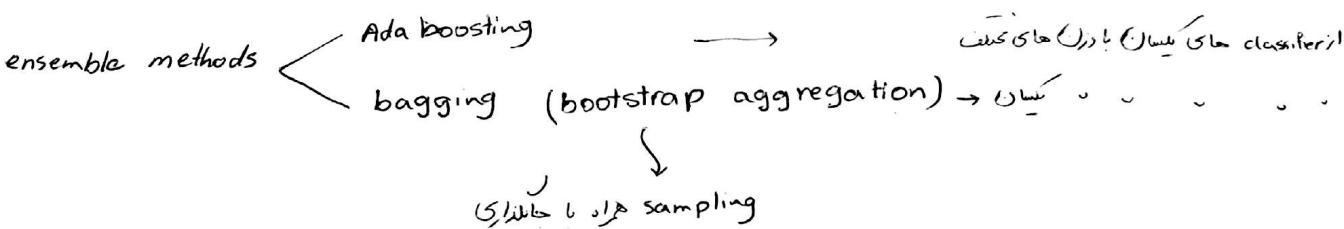
AdaBoost shocking fact

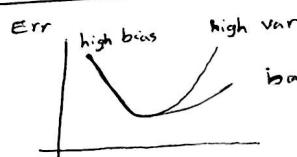
آن نه است که در آن داده‌ها نیز داده‌ی برتر باید نباشد. جزو آن داده‌ها می‌باشد.

بیشتر طبقه برای توجه overfitting: آن اگر داده‌ی پیش‌نیازه باشیم و بخطه تئیین (جست آن) آن داده‌ی پیش‌نیازه  $e^{-t}$  بیار می‌گذرد و باتوجه به این  $e^{-t}$ ، مقادیر  $\alpha^t$  کم می‌شوند. بنابراین هر داده‌ی برتر را از نظر قوتة در دنیا نیز آن اوضاع خاص باشند.

classifier  $t$

جلسه ۹، ۹، ۲۰ : ۲۰





bagging  $\rightarrow$  میانگین از همه

$$h: X \rightarrow \{-1, +1\}$$

$$H(\bar{x}_{test}) = \underbrace{\text{Sgn}(h_1(\bar{x}) + h_2(\bar{x}) + \dots)}_{\text{majority vote}} + h_k(x)$$

memorization  $\downarrow$

overfitting  $\downarrow$

variance  $\downarrow$

bagging  $\downarrow$

bootstrap:

training



- create a random subset of  $\frac{\#m}{\#m}$  data by sampling

- Draw  $m'$  of  $m$  samples with replacement ( $m' \leq m$ )

bagging:

- repeat  $k$  times / create  $k$  learners

- create training set  $\#i$  of the bootstrap examples.

- train classifier  $\#i$  on the training set  $\#i$ .

$i=1$  to  $k$

\* آر سند (رسان) بود جواب نهایی، میانگین غیر قابل دار جواب خارجی بود.

\* هر چه  $k$  بسته باشد از طریق مبتدا اختناب کنیم  
ایران

برای آنکه آر سند در تنبیه باشد و ار رسانی کرد  $\Rightarrow$  voting

random forest (ار رسانی)  $\Leftarrow$   $n$  bagged tree  $\Leftarrow$  bagged classifier

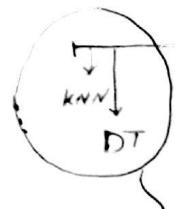
diversity brings prosperity  $\rightarrow$  برای ایجاد تنوع بسته، در هر سطح درختها یک دست  $n$ subset از درختها

برای bagging میانگین از همه classifier \*  
برای استabilitی خود خوب است. (هر یکی از هم خوب خواهد بود)



more to unstable  $\rightarrow$  bagging  $\rightarrow$  more stable  $\rightarrow$  جیل بازی +

stability



unstable  $\leftarrow$  جیل بازی  $\rightarrow$  more stable  $\rightarrow$  جیل بازی +

or SVM  $\rightarrow$  more stable  $\rightarrow$  جیل بازی +

more stable  $\rightarrow$  جیل بازی +

more stability  $\rightarrow$  جیل بازی +

big data

more stable  $\rightarrow$  جیل بازی +

more stable  $\rightarrow$  جیل بازی +

GD  $\xrightarrow{\text{scalable}}$

(stochastic Gradient Descent)

SGD

not scalable  $\rightarrow$

BN  $\rightarrow$

NBC

scalable  $\rightarrow$  KNN

clustering

discovery

کوچک کردن

دسته بندی

k-means  $\rightarrow$  جیل بازی +

K-Means clustering

نحوه انتخاب cluster  $\rightarrow$  K انتخاب شده باشد + اعداد نمونه انتخاب شده

پارامتر  $\rightarrow$  نیازمند (centroids)

: ۰.۱

کمترین خطا (minimum error)  $\rightarrow$  K (centroids)  $\rightarrow$  نیازمند (centroids)

نیازمند (centroids)  $\rightarrow$  نیازمند (centroids)

update centroid

۲) حال  $\rightarrow$  نیازمند (centroids)  $\rightarrow$  نیازمند (centroids)

centroids  $\rightarrow$  نیازمند (centroids)

matlab  $\rightarrow$  function

عویض مبتدا که overfitting را که k=m است (که هر داده را در یک گروه قرار دهد) می خواهیم جلوگیری کرد.

K-means (در درستی k-means plus)

intelligent k-means

k-means plus

جلسه ۲۸ : ۹۷، ۹۸

density estimation

PCA

خواص نسبتی

$\{\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_m\}$

unsupervised learning

یادگیری بدون نظارت / معلم

دسته آنلاین

descriptive

clustering

نمایشی کاربر یادگیری بدون نظارت

معتاد طبقه بندی کاربر یادگیری بدون نظارت و clustering

الوریشم های خواص نسبتی

آنالیز بنیان بر اساس داده ها

1) partitive algorithms e.g. k-means

معظم راهنمایی خواص نسبتی

2) hierarchical methods

3) density-based methods

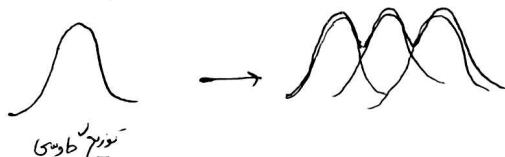
معتاد حیثیتی است، انتشار این اتفاق را در

4) probabilistic

: Expectation Maximization

کتابخانه مخصوص است

GMM (Gaussian Mixture --)



linearly inseparable

از چیزی که

Kernel k-means

که کاری که k-means

محضین یعنی آنها سعی کرد



الgoritم های سلسه مراتبی

اسلايدها

در شال اسلايدها درسترا overfitting و underfitting دارد هر دوی که نتیجه جذاب است و در آخر اینجا overfitting دارد که عرضی را در آن خوب نیست و نتیجه خوب نیست.

کے پایه cut-off بعد خوبها سفید نیست.

روش های density based سلسه مراتبی آن بین خوب برآورده است.

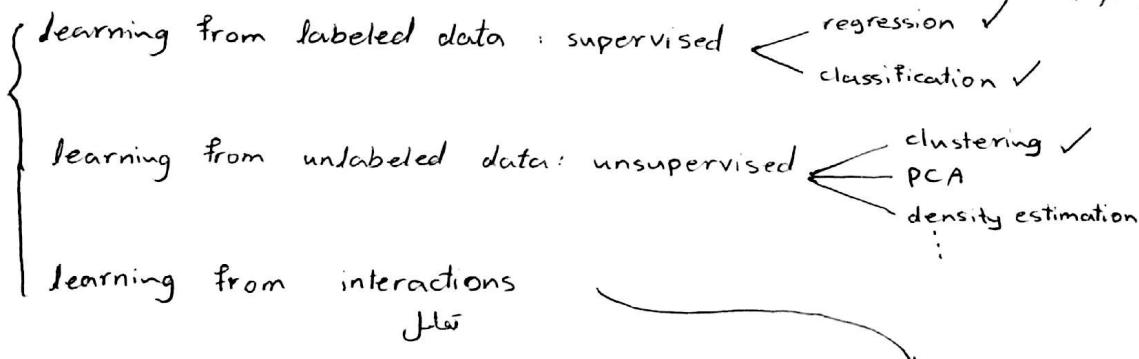
اسلايدها

اینها از جیسیل تکوات آنچه حس دیده در فصل نهاده

در فصل ۹ و ۱۰ مادر دایرها هستند (سطر آن به است)  
 پارامترها  
 { ع : عواملی که باعث (نفعها) ← نفع عیشه باید شوند  
 مدل آنها را تاکی خوب است حاصل آن .

در صورت زدن الgoritم جبری، نوشی سودانه در شاله توصیه می شود.

جلسه ۹۷، ۱۰، ۱۲



ادهی آشنی نداریم ہیں لاری جیسی مدرس نداریم. جیسے آشنی طبیعی agent، انجام کر دست یا ایش کی نیز در فصل ایون امریت جیسی مسند

شال: نظم یا لیکی حرکت مانش (عن جو در فصل)

(Markov Decision Process)

جیسے یعنی MDP مدرس مسند

ذهن مارن (memoryless)

$$P(x_{t+1} = x_1 | x_t = x_1, x_{t-1} = x_2, \dots) = P(x_{t+1} = x_1 | x_t = x_1)$$

# و' Reinforcement Learning $\rightarrow$ Richard S. Sutton

Subject: .....  
Year: .... Month: .... Date: ....

آلرنس مایسیم افغانه دیشم، پایه‌بردهی معرفت - MDP مدل کوچک  $\leftarrow$  باز حس و انت اولت  
نهی حالت‌های سلسله نامه مارک

بران) با حس و تواند نسبت فرضی نه حس ببران  
اون) حدف یادگیری یک سیاست (policy) اولت

رسانیدها . هدف  $\max_{\pi} \mathbb{E}[R_t]$  مع reward (در مداخل انتهاست) و  $\max_{\pi} \mathbb{E}[R_t]$   
کم کشید (ترتیب آنها را بین دو دیگران بفرمود)  
نابرابر - مداخل اولیه اهمیت سنتی خواهد کشید  
exploit  $\Rightarrow 0$   
explore  $\Rightarrow 1$

منابع پارسی  $\rightarrow$  باورش ماین سل حل دینم dynamic programming  
teacher/ human demonstration  
imitation learning  
همه‌یکی - بیان برترین ایله سفارخ طریق صحیح را انجام می‌دهند که از این طریق  
تکرار  $\rightarrow$  نسبت به پیش ترتیب ربات  
تکرار از یادگیری تقویتی

self-taught learning  $\leftarrow$  deep learning در داشته Andrew Ng  
خواهش نیلیم های  $\rightarrow$  رای یادگیری گین تقویتی کشید دیوستود

