生物统计学：
生物信息中的概率统计模型

2024年秋

Yann LeCun
Geoffrey Hinton
Yoshua Bengio
Andrew Ng

生物统计学2024

有关信息

- 授课教师：宁康
  - Email: ningkang@hust.edu.cn
  - Office: 华中科技大学东十一楼606室
  - Phone: 18627968927

- 课程网页
  - http://www.microbioinformatics.org/Biostatistics.html
  - QQ群：717914581

生物统计学2024

课程安排

- 生物背景和课程简介
- 传统生物统计学及其应用
- 生物统计学和生物大数据挖掘
  - Hidden Markov Model (HMM)及其应用
    - Markov Chain
    - HMM理论
    - HMM和基因识别 (Topic I)
    - HMM和序列比对 (Topic II)
  - 进化树的概率模型 (Topic III)
  - Motif finding中的概率模型 (Topic IV)
    - EM algorithm
    - Markov Chain Monte Carlo (MCMC)
  - 基因表达数据分析 (Topic V)
    - 聚类分析-Mixture model
    - Classification-Lasso Based variable selection
  - 基因网络推断 (Topic VI)
    - Bayesian网络
    - Gaussian Graphical Model
  - 基因网络分析 (Topic VII)
    - Network clustering
    - Network Motif
    - Markov random field (MRF)
  - Dimension reduction及其应用 (Topic VIII)
- 面向生物大数据挖掘的深度学习

研究对象:
生物序列,
进化树,
生物网络,
基因表达
...

方法:
生物计算与生物统计

生物统计学2024

第10章：面向生物大数据挖掘的深度学习

- 深度学习是什么
- 深度学习的基本方法（以及神经网络算法）
- 生物大数据的深度学习方法
- 生物大数据深度学习的应用案例

生物统计学2024

Part I

深度学习是什么

生物统计学2024

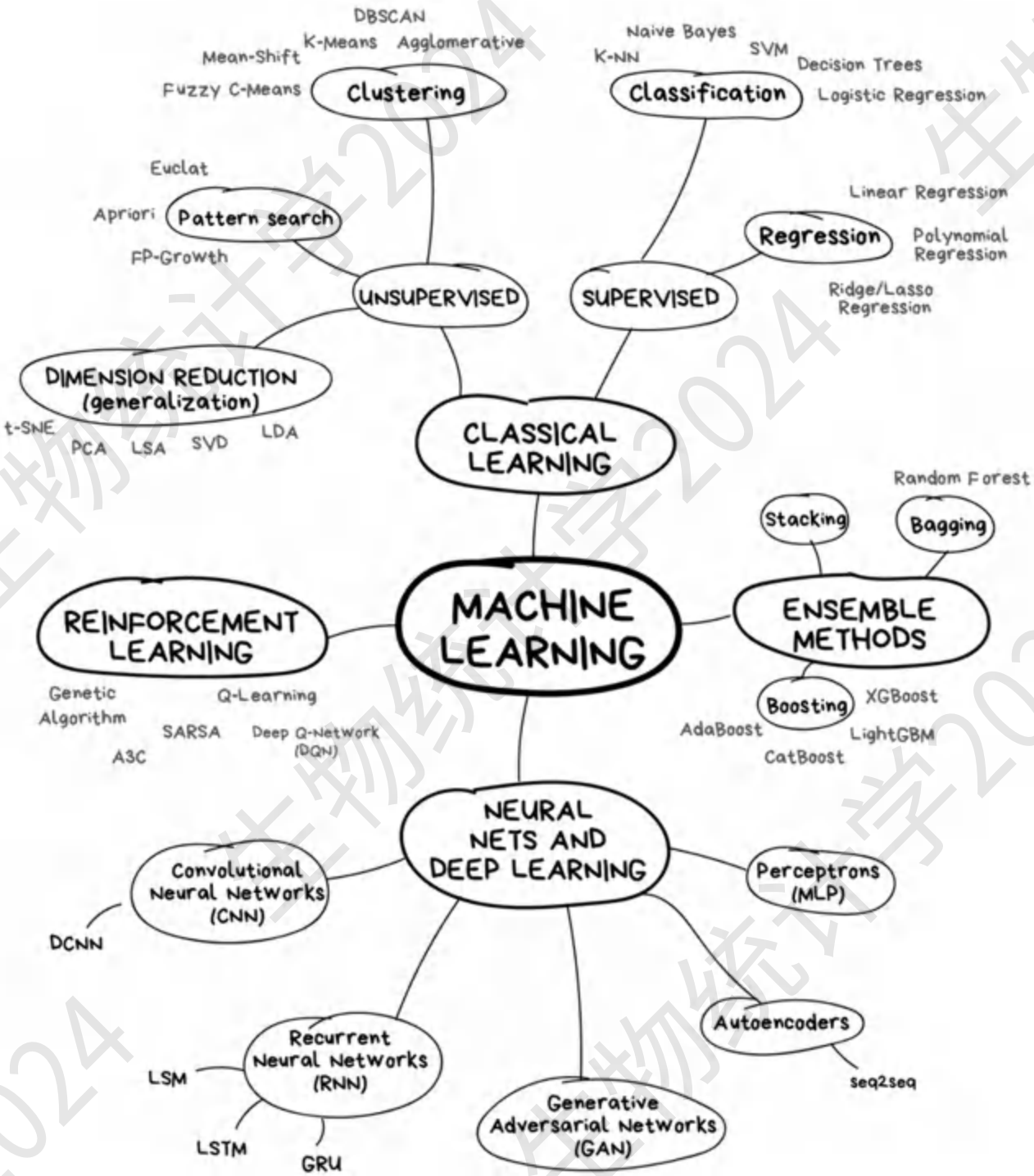
$$\frac{d\mathcal{L}}{dW} = \sum_i \frac{d\mathcal{L}}{da_i}\frac{da_i}{dz_i}\frac{dz_i}{dW}$$
$$= \frac{1}{N}\sum_i -\left(\frac{y_i}{a_i} - \frac{1-y_i}{1-a_i}\right)\cdot\frac{\exp(-z)}{(1+\exp(-z))^2}\cdot x_i$$
$$\frac{1}{N}\sum_i -\left(\frac{y_i - a_i}{a_i(1-a_i)}\right)\cdot a_i(1-a_i)\cdot x_i$$
$$= \frac{1}{N}\sum_i -(y_i - a_i)\cdot x_i$$

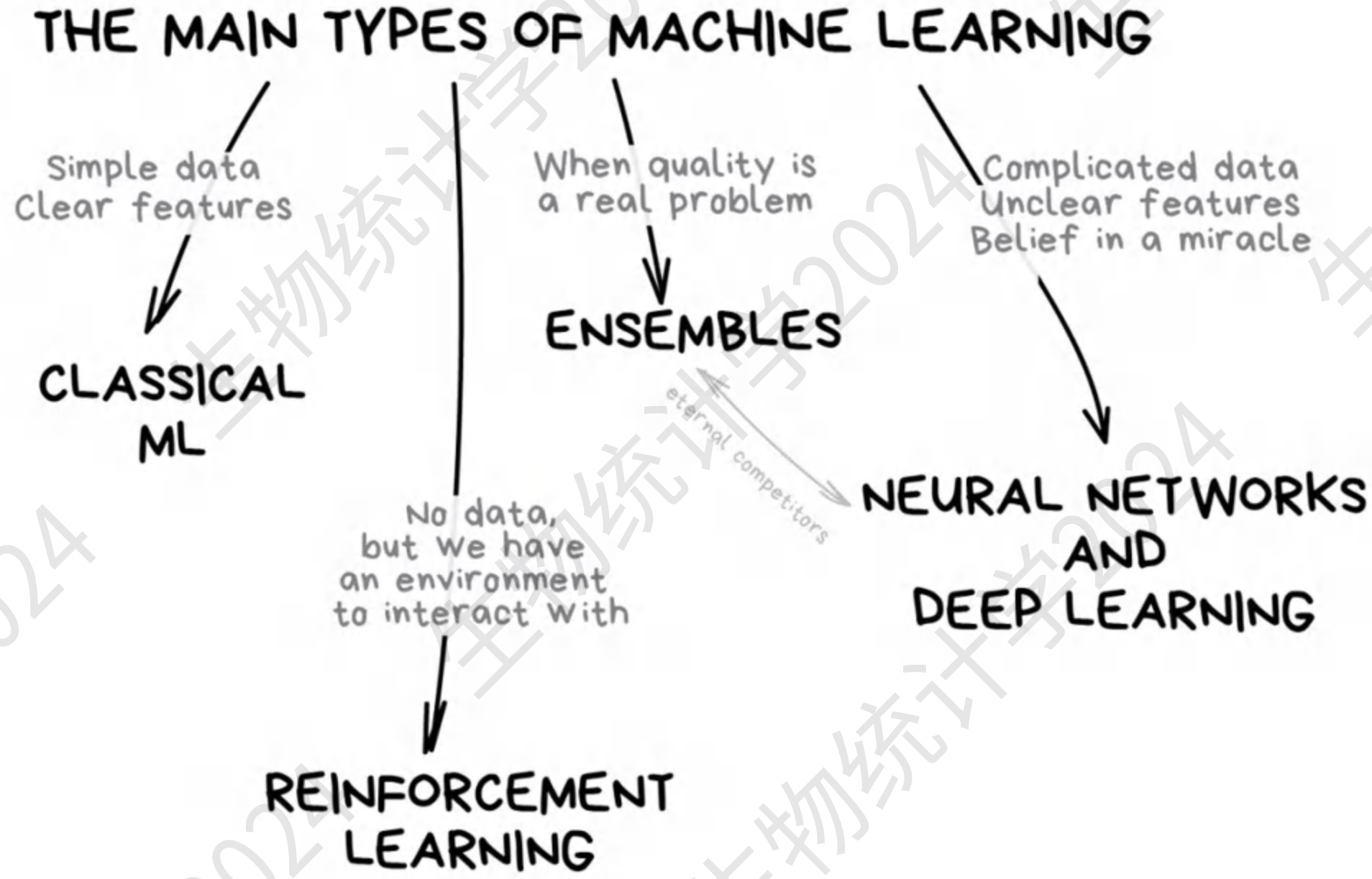
AS WE CAN SEE HERE,
THIS IS OBVIOUS!

PROGRAMMERS ARE PROGRAMMING!
DATASCIENCE!
PROFESSION OF FUTURE!
IN THE NEXT FIVE YEARS...
EXPONENTIAL GROWTH!!!
SMART MACHINES!
A-A-A-A-A-A-A-A-A-A-AAA!!!!!!!

TWO TYPES OF ARTICLES ABOUT MACHINE LEARNING

生物统计学2024


MACHINE LEARNING
CLASSICAL LEARNING
UNSUPERVISED
Clustering
DBSCAN
K-Means
Agglomerative
Mean-Shift
Fuzzy C-Means
Pattern search
Euclat
Apriori
FP-Growth
DIMENSION REDUCTION (generalization)
t-SNE
PCA
LSA
SVD
LDA
SUPERVISED
Classification
Naive Bayes
K-NN
SVM
Decision Trees
Logistic Regression
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
ENSEMBLE METHODS
Stacking
Bagging
Random Forest
Boosting
XGBoost
AdaBoost
LightGBM
CatBoost
REINFORCEMENT LEARNING
Genetic Algorithm
Q-Learning
SARSA
Deep Q-Network (DQN)
A3C
NEURAL NETS AND DEEP LEARNING
Convolutional Neural Networks (CNN)
DCNN
Perceptrons (MLP)
Autoencoders
seq2seq
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)


THE MAIN TYPES OF MACHINE LEARNING
Simple data
Clear features
CLASSICAL ML
No data,
but we have
an environment
to interact with
REINFORCEMENT LEARNING
When quality is
a real problem
ENSEMBLES
eternal competitors
Complicated data
Unclear features
Belief in a miracle
NEURAL NETWORKS AND DEEP LEARNING

生物统计学2024

CLASSICAL MACHINE LEARNING

Data is pre-categorized or numerical

SUPERVISED

Predict a category

CLASSIFICATION

«Divide the socks by color»

Predict a number

REGRESSION

«Divide the ties by length»

Data is not labeled in any way

UNSUPERVISED

Divide by similarity

CLUSTERING

«Split up similar clothing into stacks»

Find hidden dependencies

DIMENSION REDUCTION (generalization)

«Make the best outfits from the given clothes»

Identify sequences

ASSOCIATION

«Find what clothes I often wear together»

生物统计学2024

THE SIMPLEST SPAM-FILTER

(used until 2010)

hey	... 1829
I'm	...1710
no	... 1191
where	... 1012
you	...985
speak	... 873
learn	...747
one	... 739

good letters

672 times

«KITTY»

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

BAYES' THEOREM

NOT SPAM

13 times

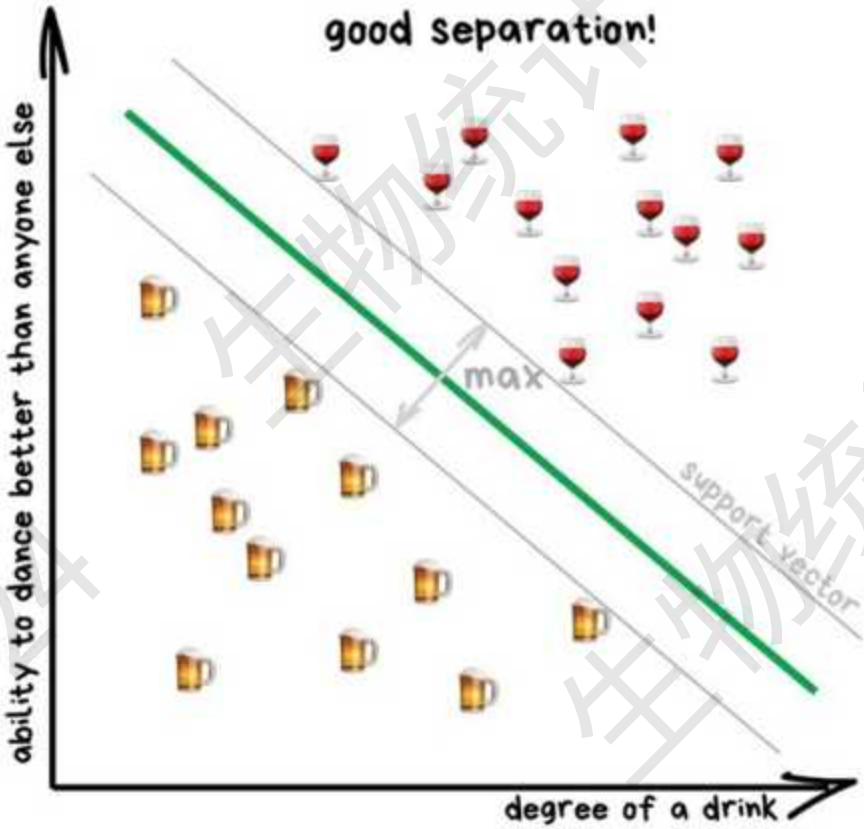
viagra	... 1552
casino	... 1492
100%	... 1320
credit	... 1184
sale	... 985
press	... 873
free	... 747
enlarge	... 739

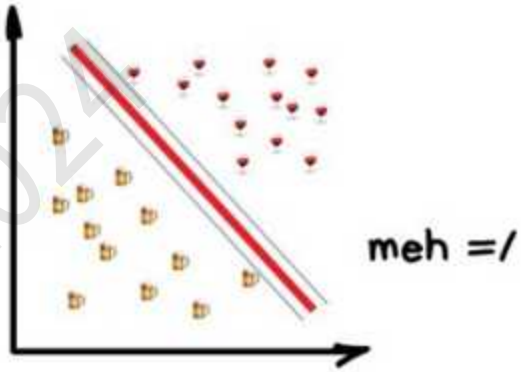
spam letters

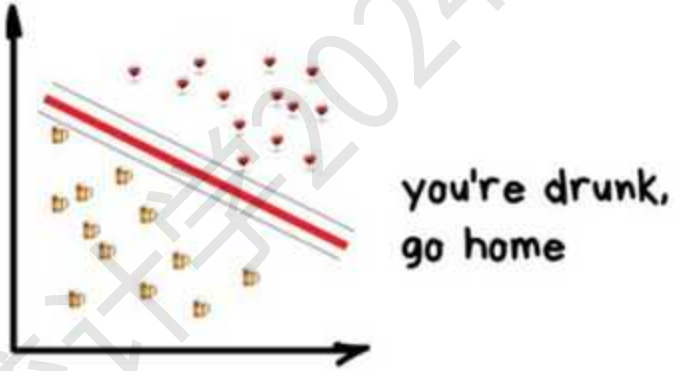
NAIVE BAYES

生物统计学2024

SEPARATE TYPES OF ALCOHOL



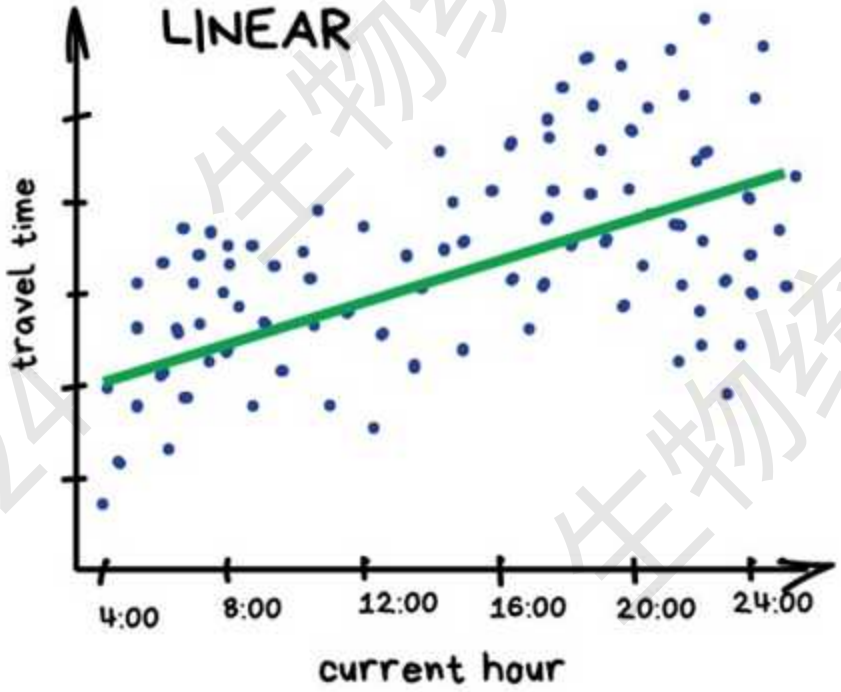


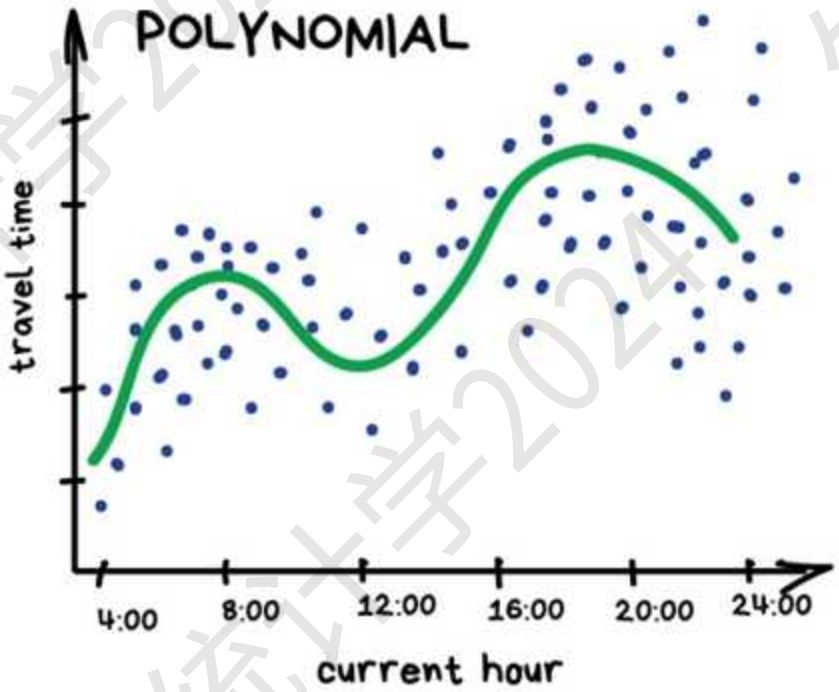


SUPPORT VECTOR MACHINE

生物统计学2024

PREDICT TRAFFIC JAMS



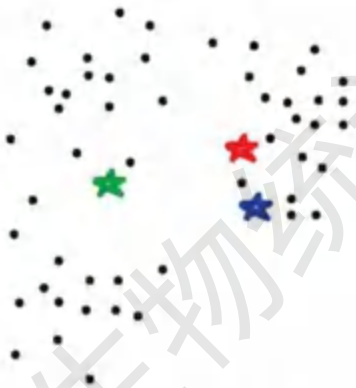


REGRESSION

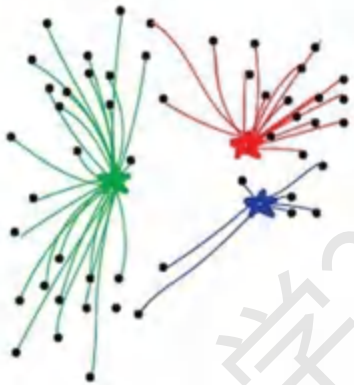
生物统计学2024

PUT KEBAB KIOSKS IN THE OPTIMAL WAY

(also illustrating the K-means method)

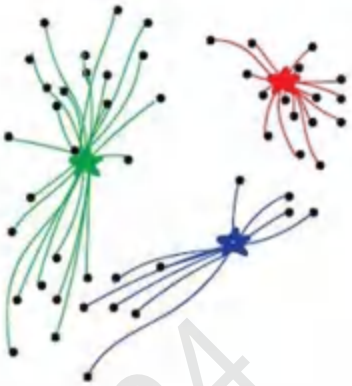


1. Put kebab kiosks in random places in city



2. Watch how buyers choose the nearest one

3. Move kiosks closer to the centers of their popularity



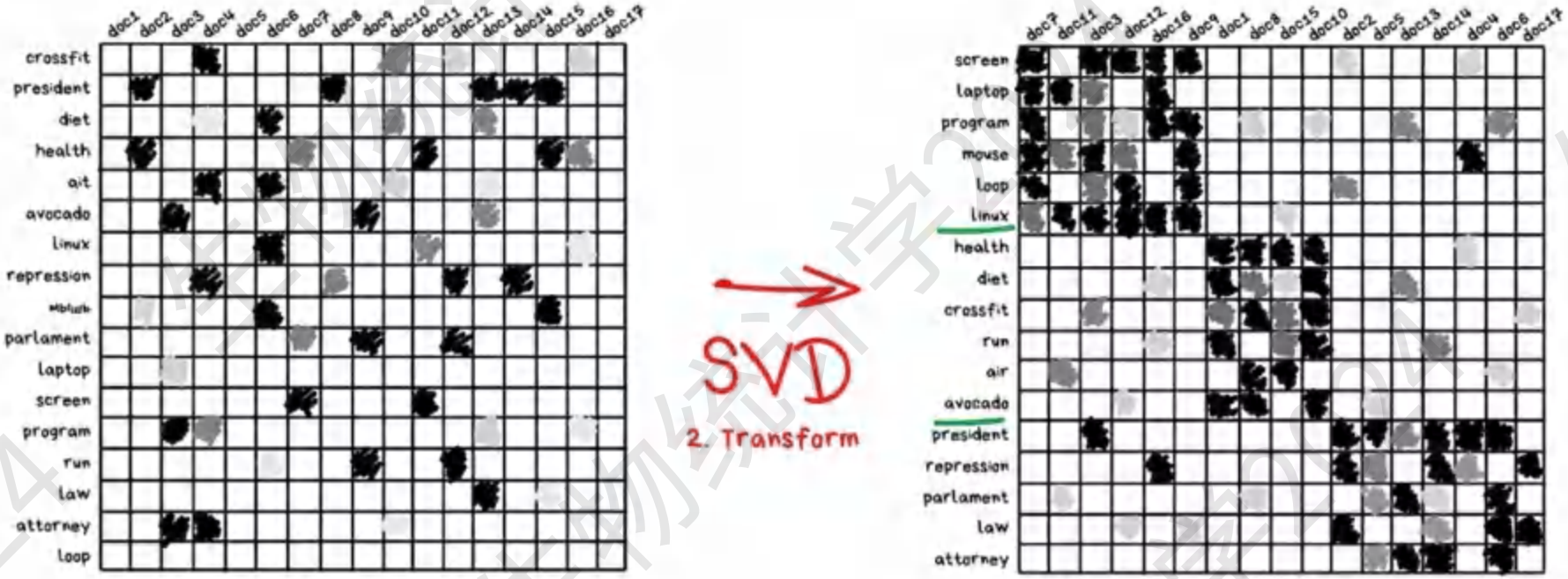
4. Watch and move again

5. Repeat a million times

6. Done!
You're god of kebabs!

生物统计学2024

SEPARATE DOCUMENTS BY TOPIC



1. Build a matrix of how often each word can be found in each document
(black - more often)

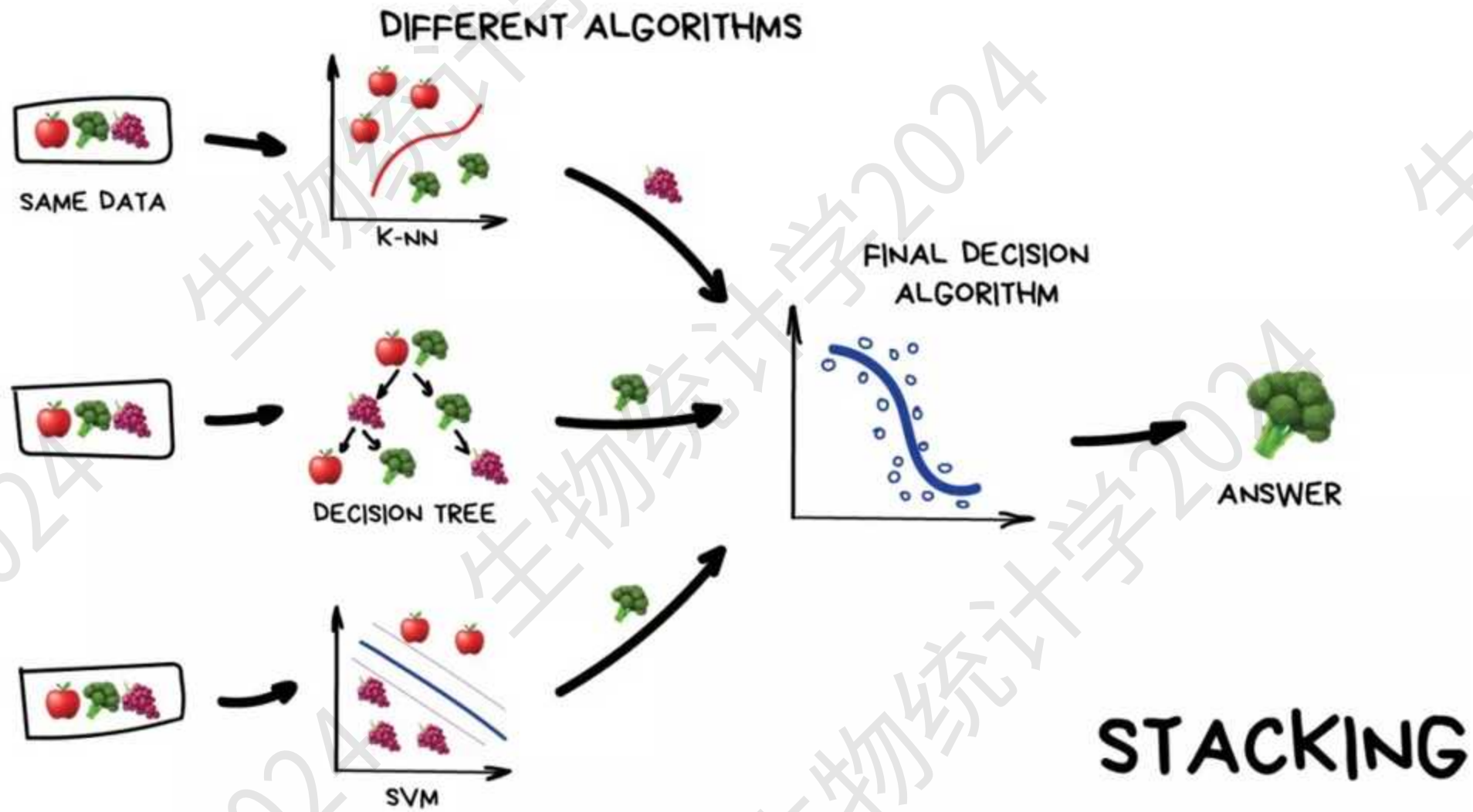
3. Get visual topic clusters.
Even if the words haven't met together

LATENT SEMANTIC ANALYSIS (LSA)

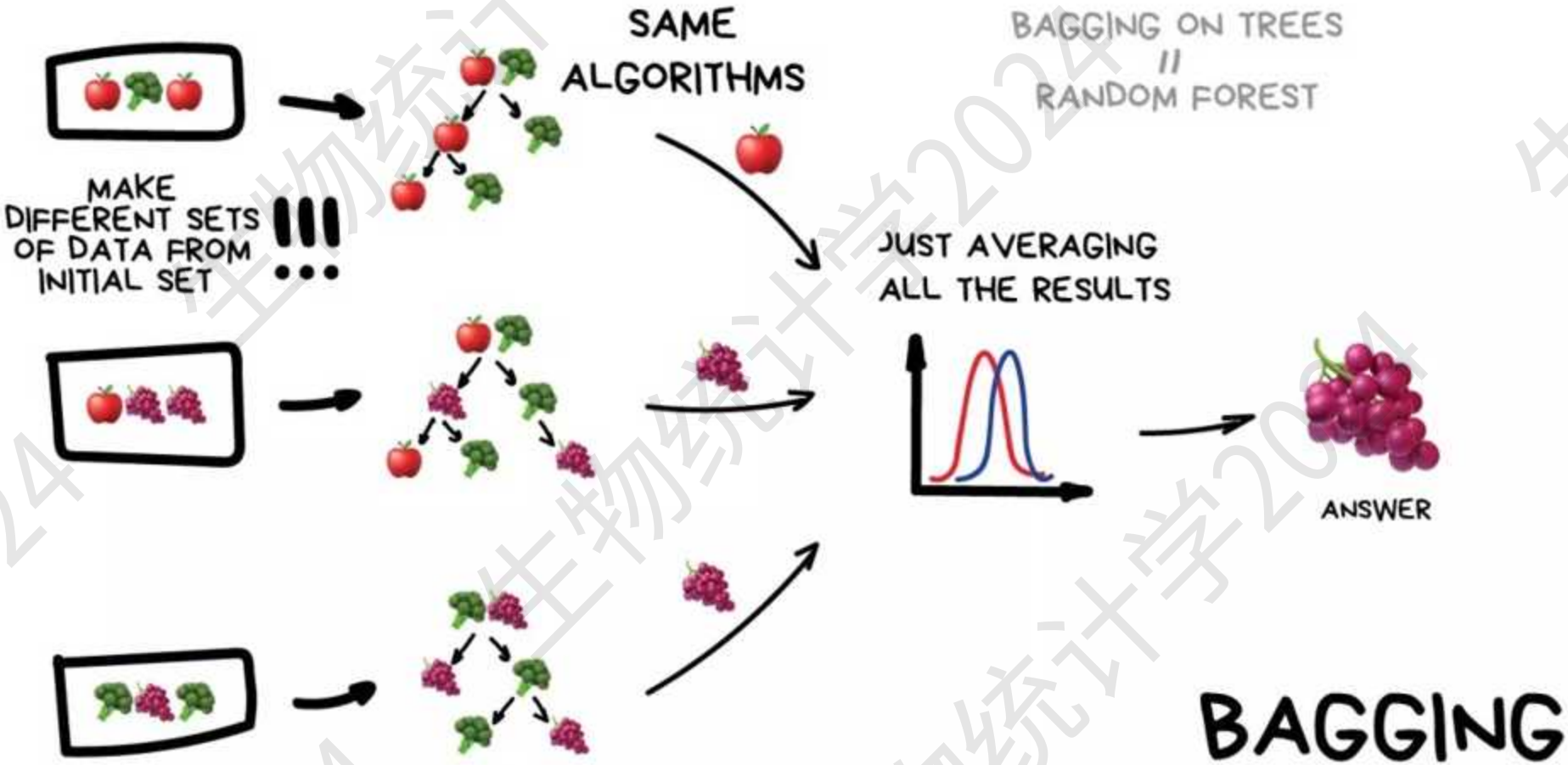
生物统计学2024

THAT MEATBAG BOUGHT A SOFA
PROBABLY, HE LOVE SOFAS!!!
RECOMMEND HIM 148 MORE SOFAS
ebay
AliExpress
amazon

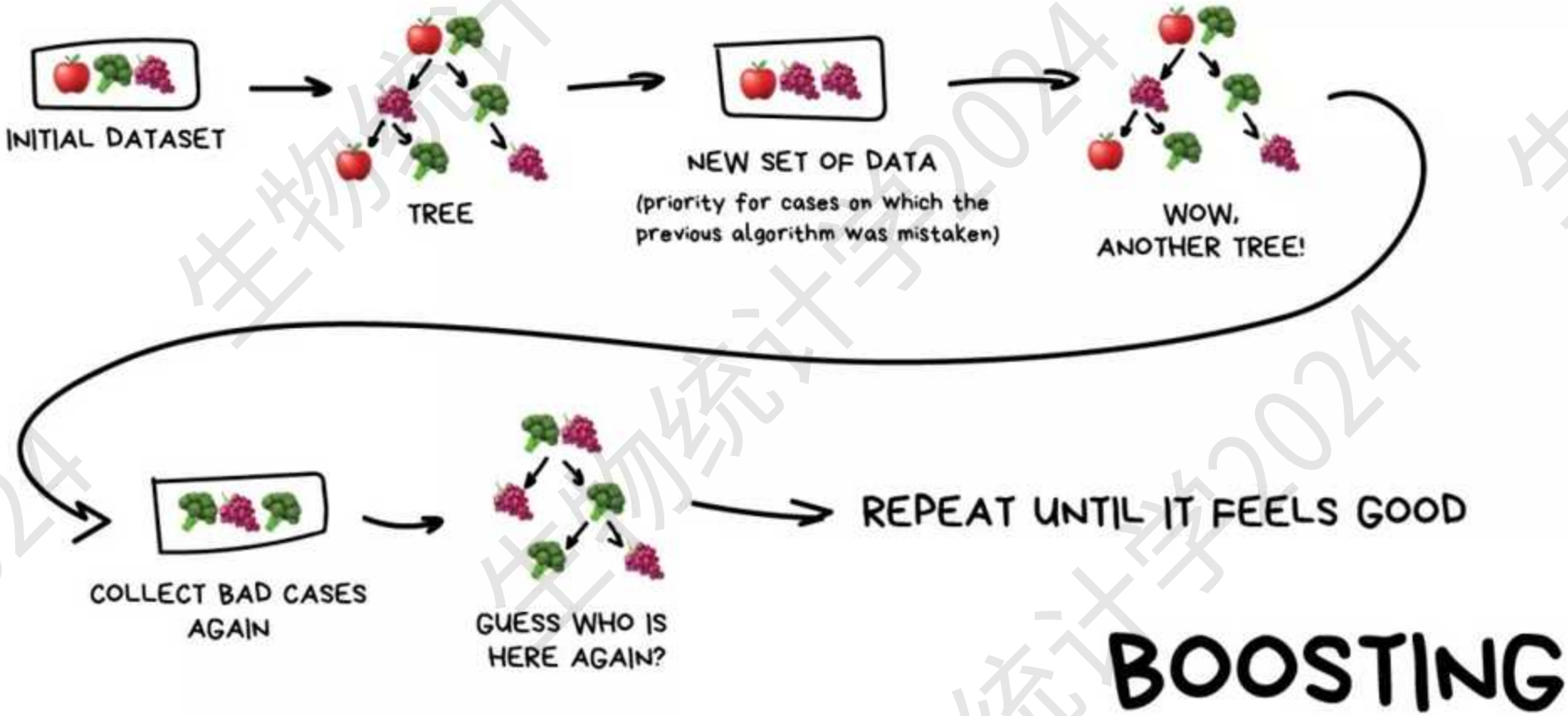
生物统计学2024


DIFFERENT ALGORITHMS
SAME DATA
K-NN
DECISION TREE
SVM
FINAL DECISION ALGORITHM
ANSWER
STACKING

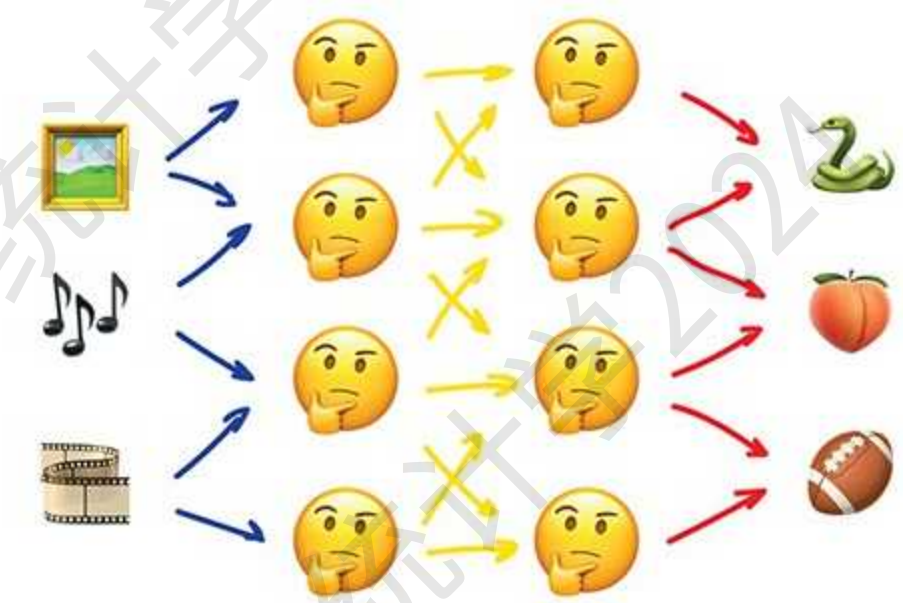
生物统计学2024


SAME ALGORITHMS
BAGGING ON TREES
//
RANDOM FOREST
MAKE DIFFERENT SETS OF DATA FROM INITIAL SET !!!
JUST AVERAGING ALL THE RESULTS
ANSWER
BAGGING

生物统计学2024

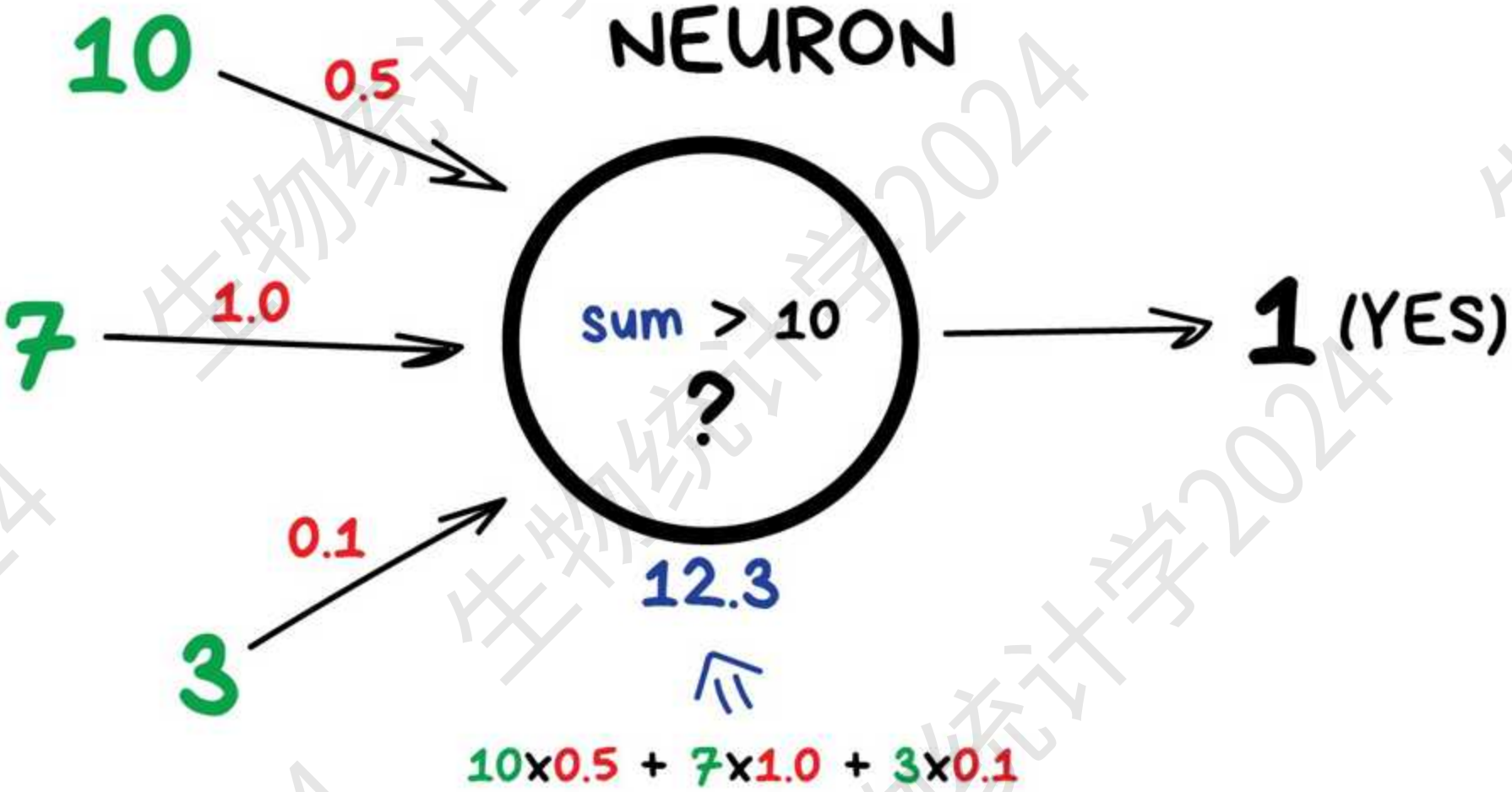

INITIAL DATASET
TREE
NEW SET OF DATA
(priority for cases on which the previous algorithm was mistaken)
WOW, ANOTHER TREE!
COLLECT BAD CASES AGAIN
GUESS WHO IS HERE AGAIN?
REPEAT UNTIL IT FEELS GOOD
BOOSTING

生物统计学2024

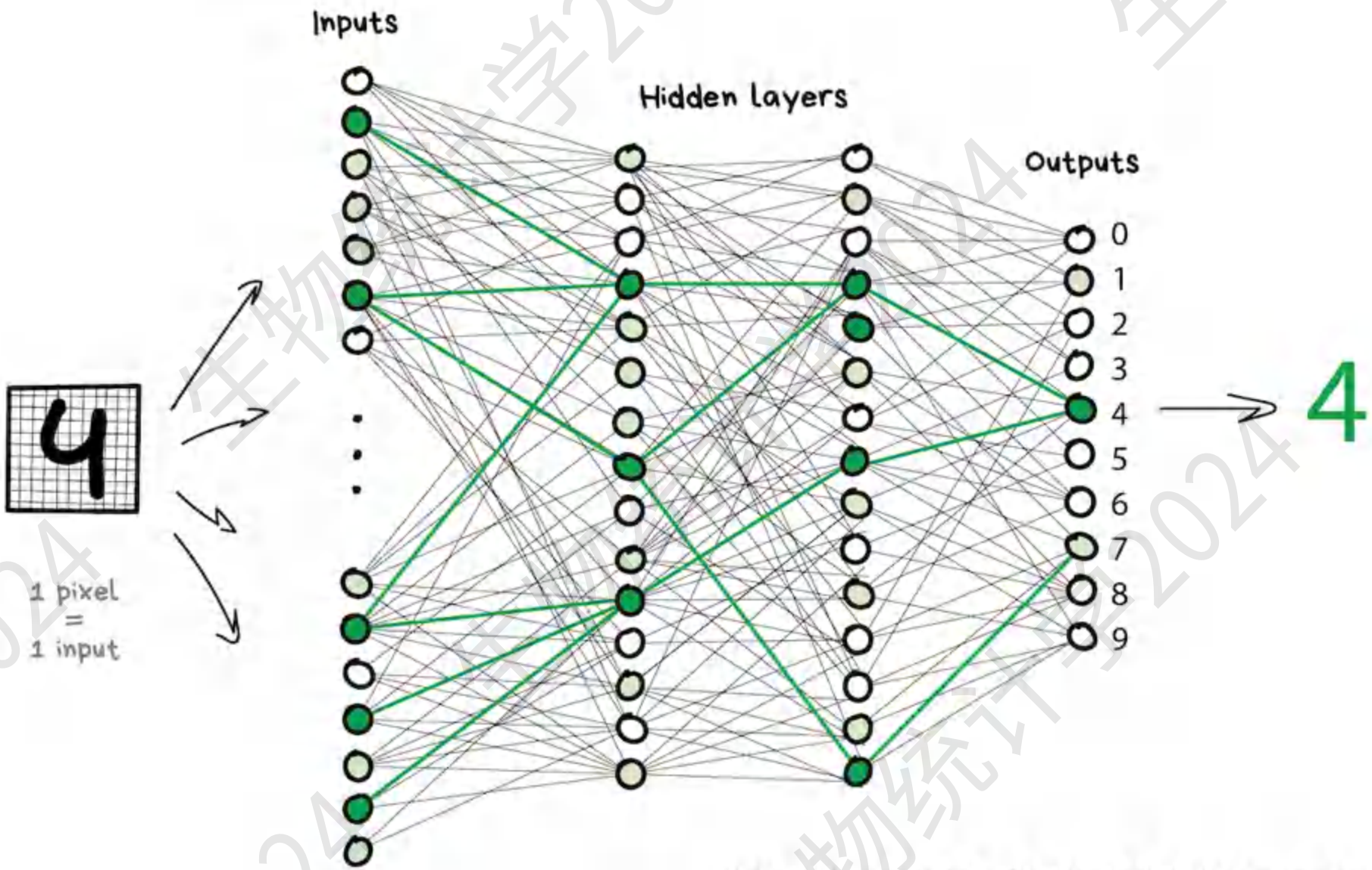


Neural Networks

生物统计学2024

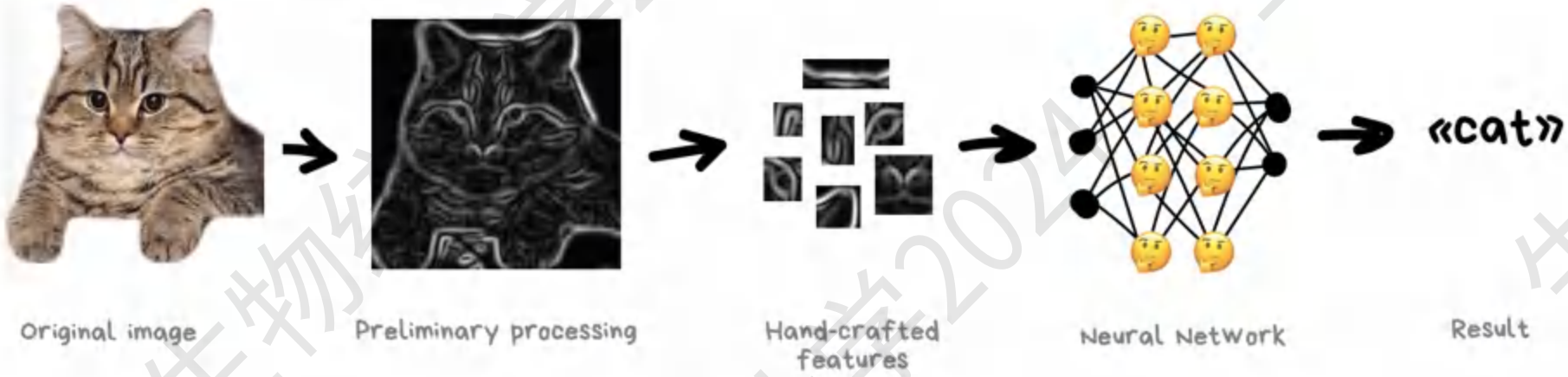


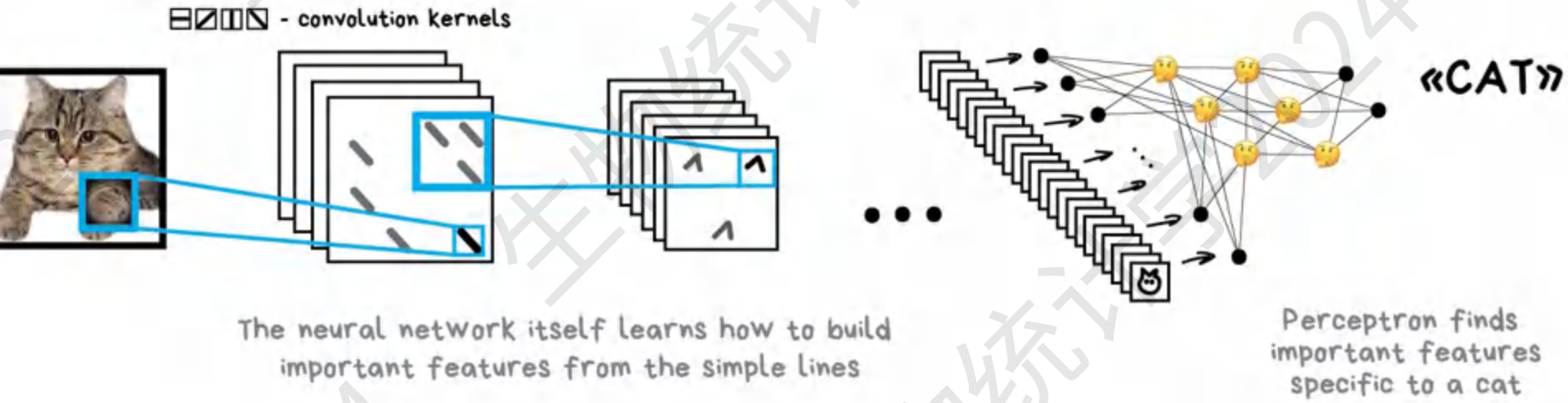
生物统计学2024



MULTILAYER PERCEPTRON (MLP)

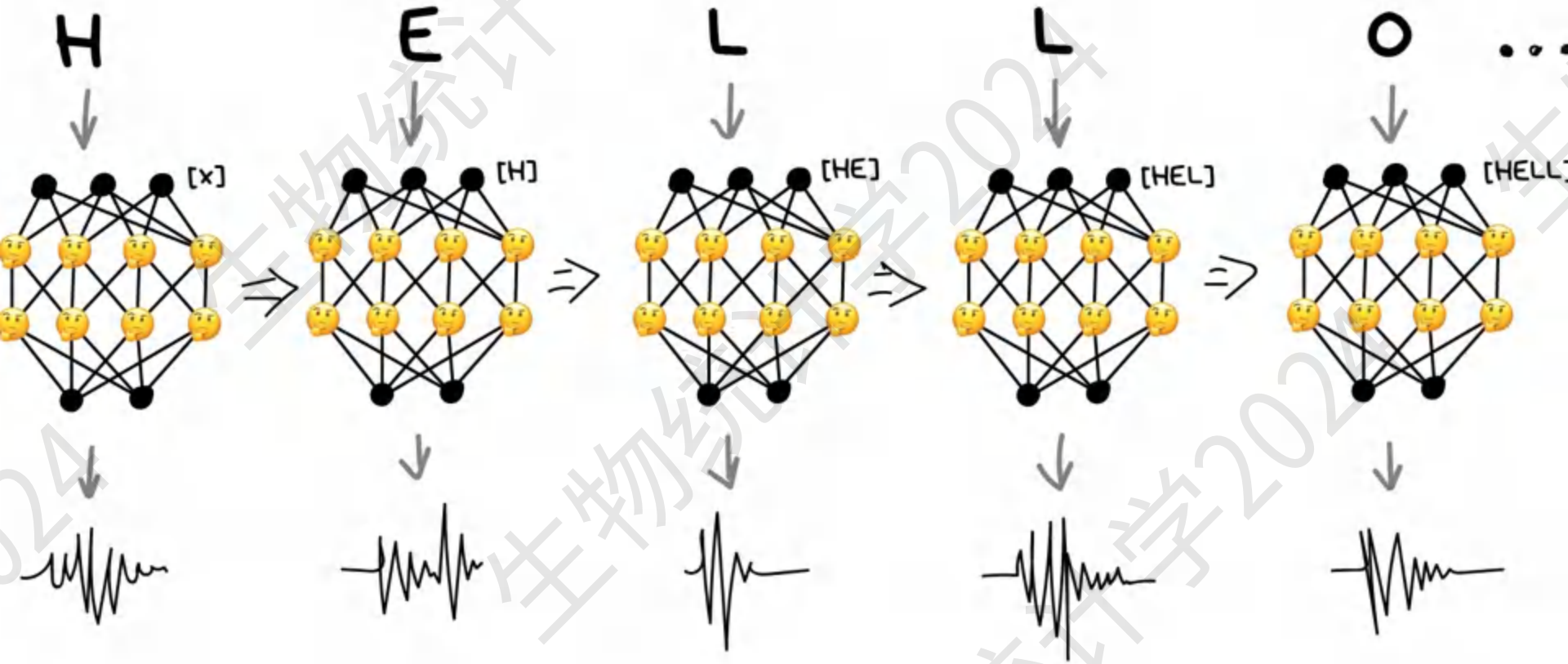
生物统计学2024





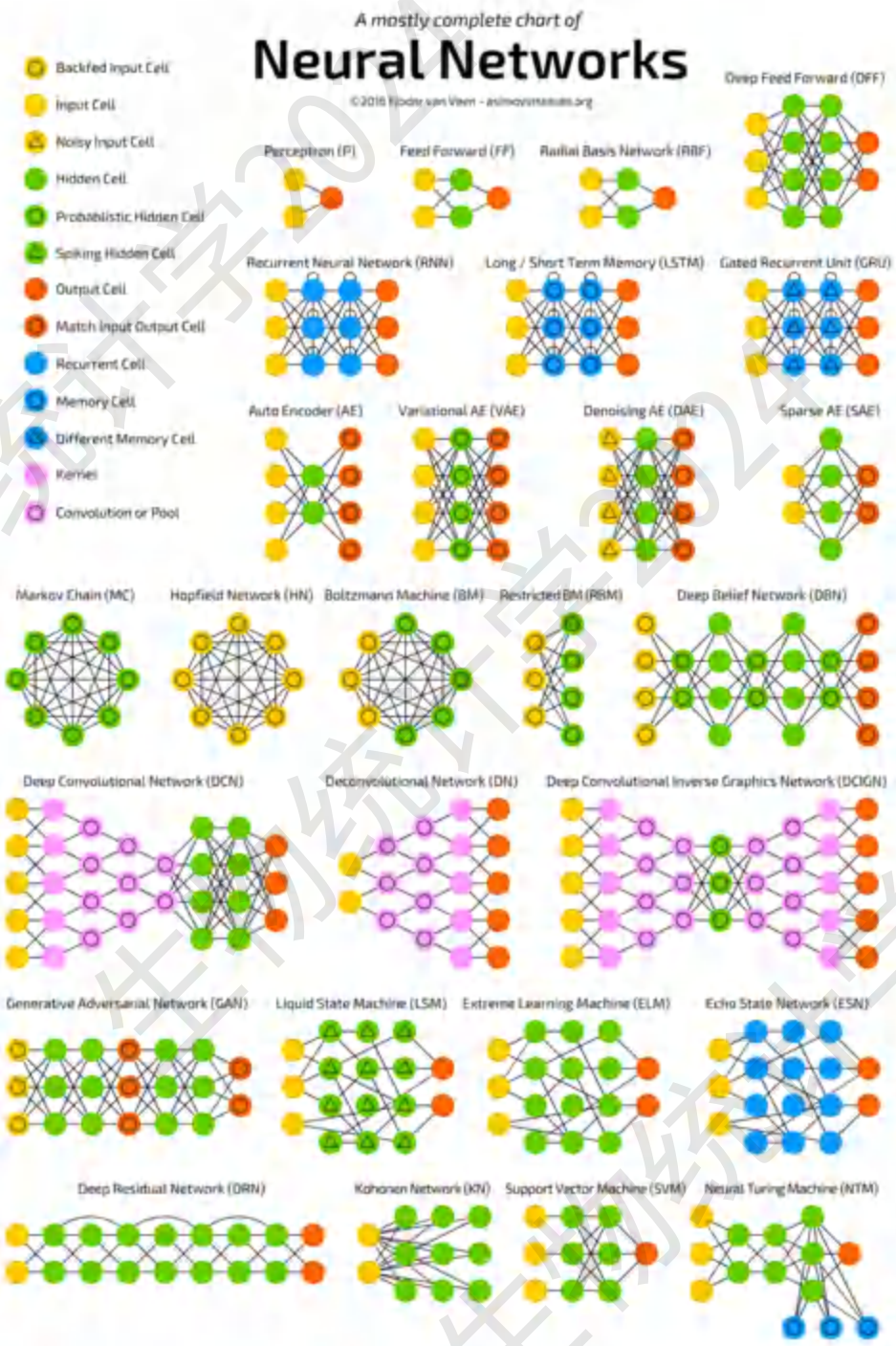
CONVOLUTIONAL NEURAL NETWORK (CNN

生物统计学2024

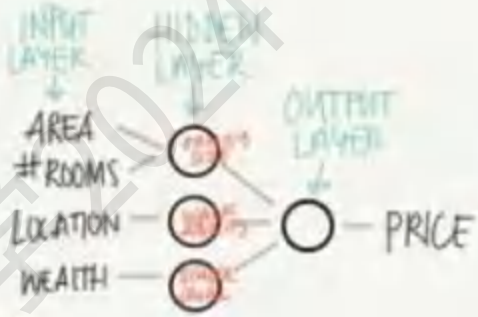


RECURRENT NEURAL NETWORK (RNN)

生物统计学2024


A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probabilistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Kohonen Network (KN)
Support Vector Machine (SVM)
Neural Turing Machine (NTM)

INTRO TO DEEP LEARNING



SUPERVISED LEARNING

INPUT X	OUTPUT Y	NN TYPE
HOME FEATURES	PRICE	STANDARD NN
AD + USER INFO	WILL CLICK ON AD (0/1)	
IMAGE	OBJECT (1...1000)	CONV NN (CNN)
AUDIO	TEXT TRANSCRIPT	RECURRENT NN (RNN)
ENGLISH	CHINESE	
IMAGE/RADAR	POS OF OTHER CARS	CUSTOM/HYBRID

NNs CAN DEAL WITH BOTH STRUCTURED & UNSTRUCTURED DATA

STRUCTURED

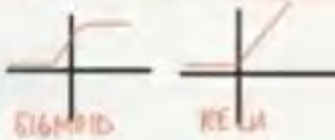
"THE QUICK BROWN FOX"

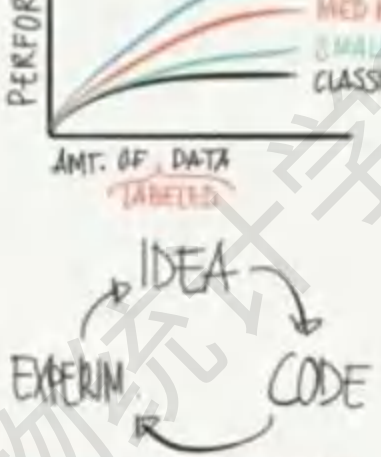
UNSTRUCTURED

HUMANS ARE GOOD AT THIS

WHY NOW?

ONE OF THE BIG BREAKTHROUGHS HAS BEEN MOVING FROM SIGMOID TO RELU FOR FASTER GRADIENT DESCENT

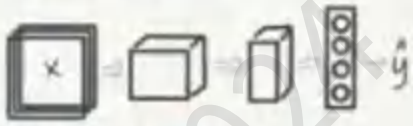




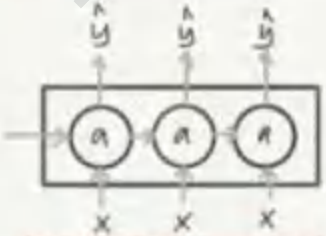
FASTER COMPUTATION IS IMPORTANT TO SPEED UP THE ITERATIVE PROCESS

NETWORK ARCHITECTURES

STANDARD NN



CONVOLUTIONAL NN



RECURRENT NN

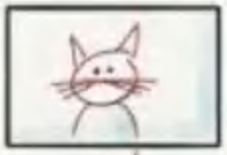
知乎 @Sophia
@lesstellendez

MEDICAL NETWORKS · DEEP LEARNING · COURSERA · ANDREW NG + MR. L

LOGISTIC REGRESSION

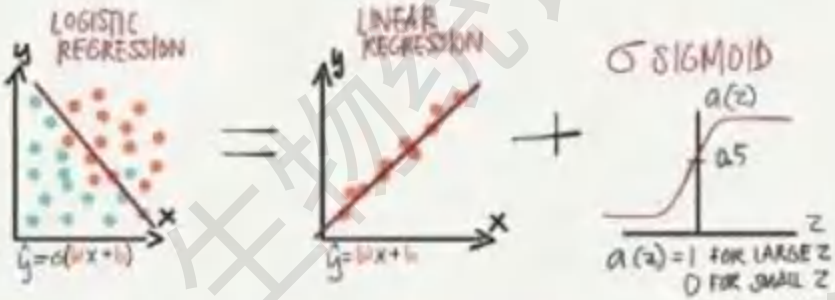
AS A NEURAL NET

BINARY CLASSIFICATION



→ 1 : CAT
0 : NOT CAT

X Y



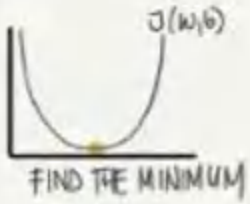
THE TASK IS TO LEARN w & b BUT HOW?

A OPTIMIZE HOW GOOD THE GUESS IS BY MINIMIZING THE DIFF BETWEEN GUESS ($\hat{y}$) AND TRUTH (y)

$$\text{LOSS} = \mathcal{L}(\hat{y}, y)$$

$$\text{COST} = J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$$

COST = LOSS FOR THE ENTIRE DATASET

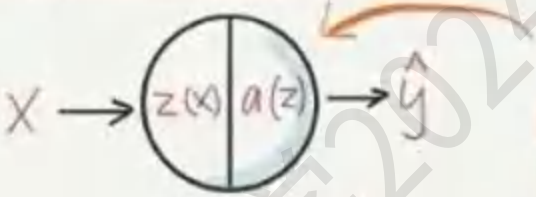


FINDING THE MINIMUM WITH GRADIENT DESCENT

1. FIND THE DOWNHILL DIRECTION (USING DERIVATIVES)
2. WALK (UPDATE w & b) AT A α LEARNING RATE

REPEAT UNTIL YOU REACH BOTTOM (CONVERGE)

PUTTING IT ALL TOGETHER



MINI NEURAL NET

$$Z(x) = wx + b$$

$$\hat{y} = a(z) = \sigma\,\text{SIGMOID}(z)$$

1. FORWARD PROPAGATION • CALCULATE $\hat{y}$
2. BACKWARD PROPAGATION • GRADIENT DESCENT + UPDATE w & b

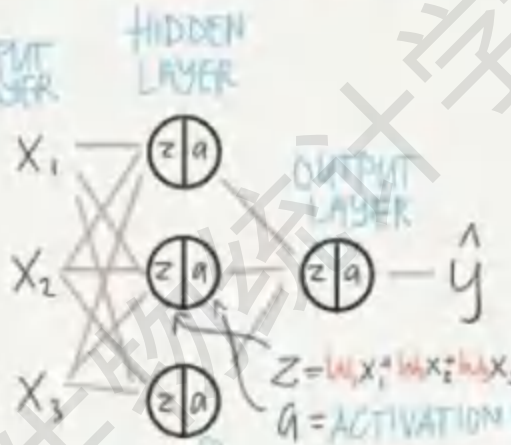
REPEAT UNTIL IT CONVERGES

知乎 @Sophia
©Jessica Mendez

NEURAL NETWORKS & DEEP LEARNING • COURSERA • ANDREW NG • WK 2

SHALLOW NEURAL NETS

2 LAYER NEURAL NET



WHY ACTIVATION FUNCTIONS?

EX. WITH NO ACTIVATION – $a = z$

$$a^{[1]} = z^{[1]} = W^{[1]}x + b^{[1]} \quad \text{LAYER 1}$$
$$a^{[2]} = z^{[2]} = W^{[2]}a^{[1]} + b^{[2]} \quad \text{LAYER 2}$$

PLUG IN $a^{[1]}$

$$a^{[2]} = W^{[2]}(W^{[1]}x + b^{[1]}) + b^{[2]}$$
$$= \underbrace{W^{[2]}W^{[1]}}_{W'}x + \underbrace{W^{[2]}b^{[1]} + b^{[2]}}_{b'}$$

$W'x + b'$ ← LINEAR FUNCTION

WE COULD JUST AS WELL HAVE SKIPPED THE WHOLE NEURAL NET & USED LIN. REGR.

ACTIVATION FUNCTIONS

SIGMOID σ
BINARY CLASSIFIER
- ONLY USED FOR OUTPUT LAYER

SLOW GRAD DESCENT SINCE SLOPE IS SMALL FOR LARGE/SMALL VAL

TANH
NORMALIZED ⇒ GRADIENT DESCENT IS FASTER

RELU
DEFAULT CHOICE FOR ACTIVATION

SLOPE = 1/0

LEAKY RELU
AVOIDS UNDEF SLOPE AT Ø BUT RARELY USED IN PRACTICE

INITIALIZING W+b

WHAT IF: INIT TO Ø

THIS WILL CAUSE ALL THE UNITS TO BE THE SAME AND LEARN EXACTLY THE SAME FEATURES

SOLUTION: RANDOM INIT

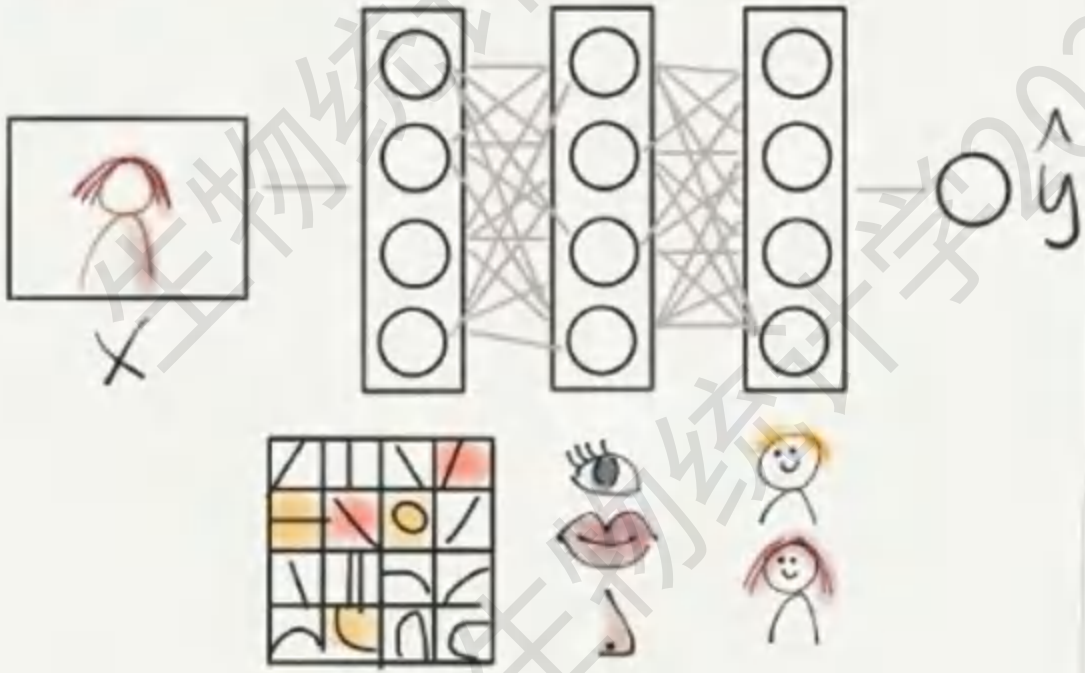
BUT ALSO WANT THEM SMALL SO RAND * 0.01 HYPERPARAM

NEURAL NETWORKS + DEEP LEARNING • COURSERA • ANDREW NG • WK 2
24 生物统计学2024 生物统计学2024 生物统计学2024
知乎 @Sophia
@essenceofsoftware

DEEP
NEURAL NETS

THERE ARE FUNCTIONS A SMALL DEEP NET CAN COMPUTE THAT SHALLOW NETS NEED EXP. MORE UNITS TO COMP.

WHY DEEP NEURAL NETS?



VERY DATA HUNGRY

NEED LOTS OF COMPUTER POWER

ALWAYS VECTORIZE
VECTOR MULT · CHEAPER THAN FOR LOOPS

COMPUTE ON GPUs

LOTS OF HYPERPARAMS

LEARNING RATE α
ITERATIONS
HIDDEN LAYERS
HIDDEN UNITS
CHOICE OF ACTIVATION
MOMENTUM
MINI-BATCH SIZE
REGULARIZATION

LOW LEVEL AUDIO WAVE FEATURES ↗↘ PITCH — PHONEMES — WORDS — SENTENCES

CAT

知乎 @Sophia
@Jessie Mendez

NEURAL NETWORKS & DEEP LEARNING • COURSERA • ANDREW NG • WK 4

SETTING UP YOUR ML APP

CLASSIC ML

100 - 10000 SAMPLES

TRAIN	DEV	TEST
60%	20%	20%

ALL FROM SAME PLACE (DISTRIBUTION)

DEEP LEARNING

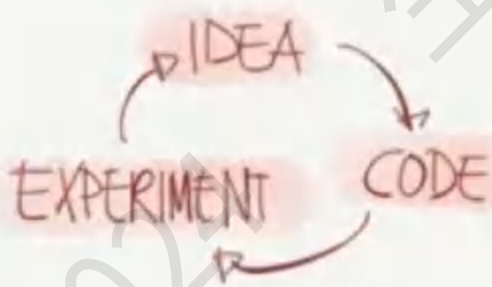
1M SAMPLES

TRAIN	D	T
98%	1%	1%

EX: TRAIN PRO CAT PICS FROM INTERNET

DEV/TEST BLURRY CAT PICS FROM APP

TIP
DEV & TEST SHOULD COME FROM SAME DISTRIBUTION

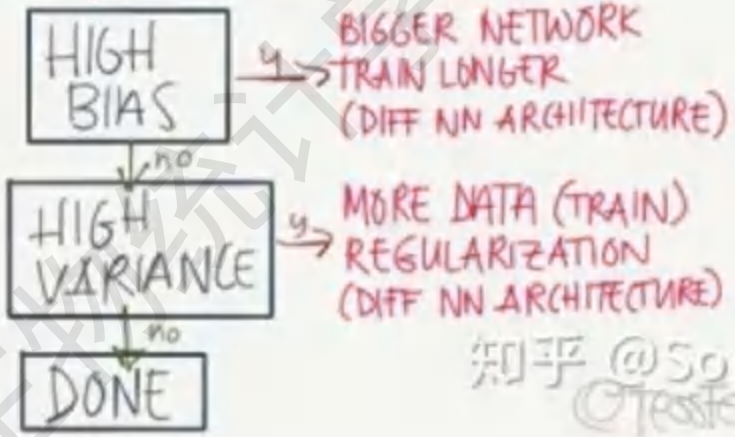


BIAS/VARIANCE

ERROR				
TRAIN	1%	15%	15%	0.5%
TEST	11%	16%	30%	1%
	HIGH VARIANCE	HIGH BIAS	HIGH BIAS & VARIANCE	LOW BIAS & VARIANCE

(ASSUMING HUMANS GET 0% ERROR)

THE ML RECIPE



知乎 @Sophia
@TessFerrandez

REGULARIZATION

PREVENTING OVERFITTING

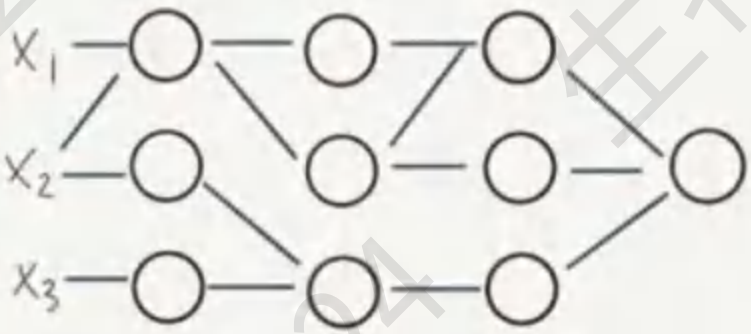
L2 REGULARIZATION

COST: $J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}(\hat{y},y) + \frac{\lambda}{2m}\|w\|_2^2$ ← EUCLIDEAN NORM

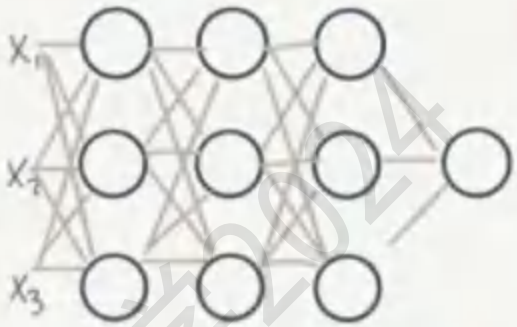
L1 REGULARIZATION

COST: $J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}(\hat{y},y) + \frac{\lambda}{m}\|w\|_1$

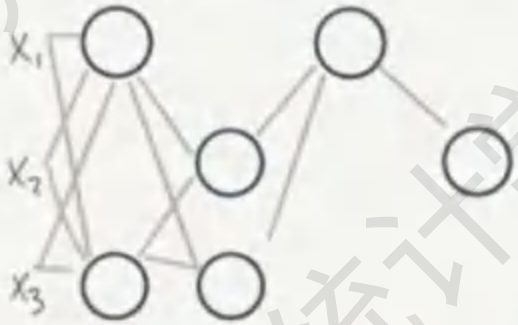
BOTH PENALIZE LARGE WEIGHTS ⇒ SOME WILL BE CLOSE TO Ø ⇒ SIMPLER NETWORKS



DROPOUT



FOR EACH ITERATION & SAMPLE SOME NODES ARE RANDOMLY DROPPED (BASED ON KEEP_PROB)

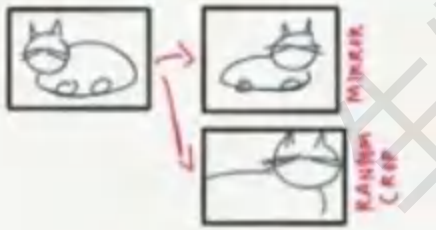


WE GET SIMPLER NWs & LESS CHANCE TO RELY ON SINGLE FEATURES

OTHER REGULARIZATION TECHNIQUES

DATA AUGMENTATION

GENERATE NEW PICS FROM EXISTIN



EARLY STOPPING

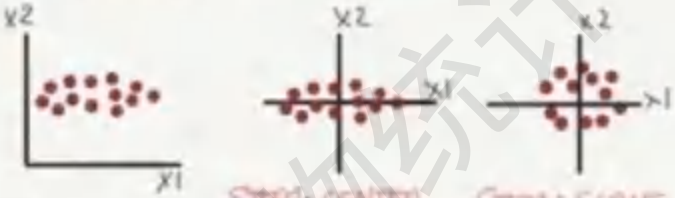
PROBLEM: AFFECTS BOTH BIAS & VARIANCE

知乎 @Sophia ©Tessellanda

IMPROVING DEEP NEURAL NETS COURSERA

OPTIMIZING TRAINING

NORMALIZING INPUTS

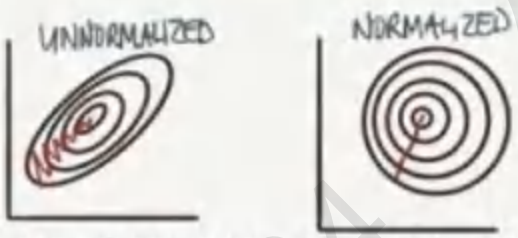


STEP1: CENTER AROUND 0,0

STEP2: SCALE SO VARIANCE IS SAME
ex $-1 \to 1$

TIP
USE SAME AVG/VAR TO NORMALIZE DEV/TEST

WHY DO WE DO THIS?



IF WE NORMALIZE, WE CAN USE A MUCH LARGER LEARNING RATE α

DEALING WITH VANISHING/EXPLODING GRADIENTS

EX: DEEP NW (L LAYERS)

$$\hat{y} = W^{[L-1]} W^{[L-2]} \cdots W^{[1]} x + b$$

IF $W = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \Rightarrow 0.5^{L-1} \Rightarrow$ VANISHING

OR $W = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix} \Rightarrow 1.5^{L-1} \Rightarrow$ EXPLODING

IN BOTH CASES GRADIENT DESCENT TAKES A VERY LONG TIME

PARTIAL SOLUTION: CHOOSE INITIAL VALUES CAREFULLY

$w^{[l]} = \text{rand} * \sqrt{\frac{2}{n^{[l-1]}}}$ (FOR RELU)

#inputs

XAVIER $\sqrt{\frac{1}{n^{[l-1]}}}$ (FOR TANH)

SETS THE VARIANCE

GRADIENT CHECKING

IF YOUR COST DOES NOT DECREASE ON EACH ITER YOU MAY HAVE A BACKPROP BUG.

GRADIENT CHECKING APPROXIMATES THE GRADIENTS SO YOU CAN VERIFY CALC.

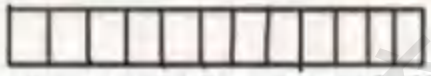
NOTE ONLY USE WHEN DEBUGGING SINCE IT'S SLOW

IMPROVING DEEP NEURAL NETS: COURSERA

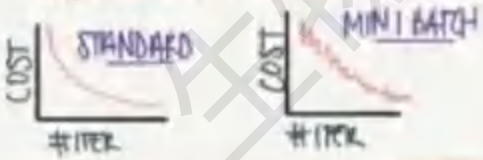
知乎 @Sophia ©Tessellate

OPTIMIZATION ALGORITHMS

MINI-BATCH GRAD. DESCENT

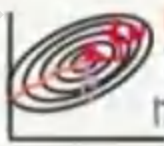


SPLIT YOUR DATA INTO MINI-BATCHES & DO GRAD DESCENT AFTER EACH BATCH
THIS WAY YOU CAN PROGRESS AFTER JUST A SHORT WHILE



CHOOSING THE MINIBATCH SIZE

SIZE = m → BATCH GRAD DESC.
SIZE = 1 → STOCHASTIC GRAD DESC.



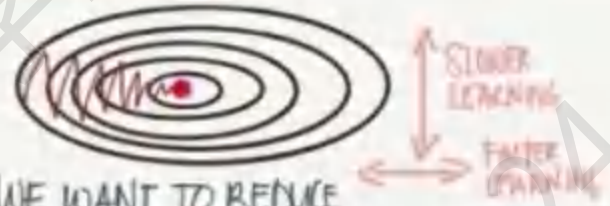
BATCH — TOO LONG PER ITERATION

MIXTURE

STOCHASTIC — LOSE ALMOST ALL SPEED FROM VECTORIZATION
- SHORT ITER.
- USES VECT.

TIP
IF YOU HAVE < 2000 SAMPLES USE SIZE = 2000
OTHERWISE, USE 64, 128, 256...
SO X+Y FITS IN CPU/GPU CACHE

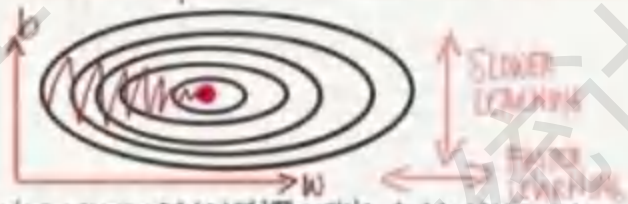
GRADIENT DESCENT W. MOMENTUM



WE WANT TO REDUCE OSCILLATION ↕ SO WE GET TO THE GOAL FASTER

SOLUTION: SMOOTH OUT THE CURVE BY TAKING AN EXPONENTIALLY WEIGHTED AVERAGE OF THE DERIVATIVES (i.e. LAST ONE HAS MORE IMPORTANCE)

RMSProp - ROOT MEAN SQUARED



NORMALIZE GRADIENT USING A MOVING AVG.

$S_{dw} = \beta S_{dw} + (1-\beta)\, dw^2$

$S_{db} = \beta S_{db} + (1-\beta)\, db^2$

$w = w - \alpha \frac{dw}{\sqrt{S_{dw}}} \quad b = b - \alpha \frac{db}{\sqrt{S_{db}}}$

ADAM OPTIMIZATION

COMBO OF GD W MOMENTUM & RMSProp

LEARNING RATE DECAY

IDEA: USE A LARGE α IN THE BEGINNING THEN DECREASE AS WE GET CLOSER TO GOAL

OPTION 1: $\alpha = \frac{1}{1 + \text{DECAYRATE} \cdot \text{EPOCH}} \alpha_0$

EXPONENTIAL: $\alpha = 0.95^{\text{EPOCH}} \alpha_0$

OPTION 3: $\alpha = \frac{k}{\sqrt{\text{EPOCH}}} \alpha_0$

OPTION 4: $\alpha = \frac{k}{\sqrt{t}} \alpha_0$

OPTION 5: DISCRETE STAIRCASE

OPTION 6: MANUAL

EPOCH = 1 PASS THROUGH THE DATA

IMPROVING DEEP NEURAL NETS COURSERA
知乎 @Sophia
生物会计学2024

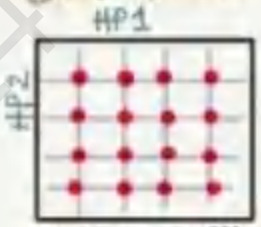
HYPERPARAM TUNING

WHICH HYPERPARAMS ARE MOST IMPORTANT?

- α LEARNING RATE
- # HIDDEN UNITS
- MINIBATCH SIZE
- β MOMENTUM TERM = 0.9
- # LAYERS
- LEARNING RATE DECAY
- $\beta_1 = 0.9$ $\beta_2 = 0.999$ $\varepsilon = 10^{-8}$ (ADAM)

TESTING VALUES

CLASSIC ML



GRID SEARCH

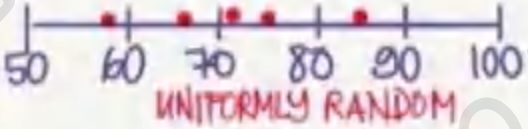
PROBLEM: ONE ITERATION TAKES A LONG TIME & IN 16 GO'S WE HAVE ONLY TRIED 4 α - BUT 4 DIFF ε ← NOT AS IMPORTANT

SOLUTION

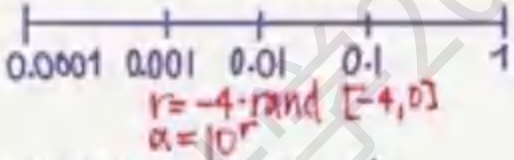
RANDOM SEARCH + COARSE → DENSE

USE AN APPROPRIATE SCALE

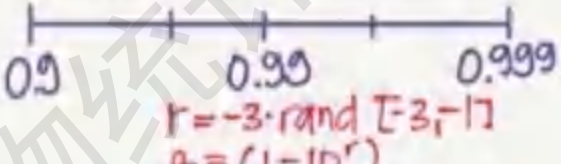
HIDDEN UNITS



α LEARNING RATE



β EXP WEIGHT AVE



TIP

RE-EVALUATE YOUR HYP-PARAMS EVERY FEW MONTHS

PANDA VS CAVIAR

BABYSIT ONE MODEL & TUNE

SPAWN LOTS OF MODELS W/ DIFF HP

GOOD IF YOU HAVE LOTS OF SPARE COMP POWER

MY PANDA IS ACTUALLY A MISCLASSIFIED CAT BECAUSE I DON'T DO BAMBOO

MISC. EXTRAS

BATCH NORMALIZATION

NORMALIZE LAYER OUTPUT

- SPEEDS UP TRAINING
- MAKES WEIGHTS DEEPER IN NW MORE ROBUST (COVARIATE SHIFT)
- SLIGHT REGULARIZING EFFECT

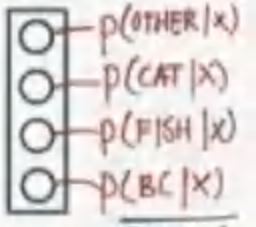
MULTICLASS CLASSIFIC

C = # CLASSES = 4

SOFTMAX ACTIVATION

$$t = e^{(z^{[L]})}$$

$$a^{[L]} = \frac{t}{\sum t_i}$$



SUM: 1

EX: $z^{[L]} = \begin{bmatrix} 5 \\ 2 \\ -1 \\ 3 \end{bmatrix}$ $t = \begin{bmatrix} e^5 \\ e^2 \\ e^{-1} \\ e^3 \end{bmatrix} = \begin{bmatrix} 148.4 \\ 7.4 \\ 0.4 \\ 20.1 \end{bmatrix}$ 176.3

$\Rightarrow a^{[L]} = \frac{t}{\sum t_i} = \begin{bmatrix} 0.842 \\ 0.042 \\ \vdots \end{bmatrix}$ ← 11.4% PROB THIS IS A CAT

2024 生物统计学2024 生物统计学2024
知乎 @Testicone

STRUCTURING YOUR ML PROJECTS

SETTING YOUR GOAL

★ GOAL SHOULD BE A SINGLE #

	PRECISION	RECALL
A	95%	90%
B	98%	85%

IS A OR B BEST?

	PRECISION	RECALL	F1
A	95%	90%	92.4%
B	98%	85%	91%

A IS BEST

F1 = HARMONIC MEAN BETW. RECALL & PRECISION

★ DEFINE OPTIMIZING VS SATISFICING METRICS

	ACCURACY	RUNTIME
A	90%	80ms
B	92%	95ms
C	95%	1500ms

MAXIMIZE ACC. GIVEN TIME ≤ 100ms

ACCURACY = OPTIMIZING

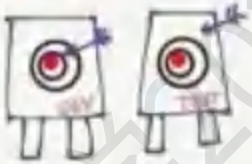
RUNTIME = SATISFICING

SELECTING YOUR DEV/TEST SETS

DATA

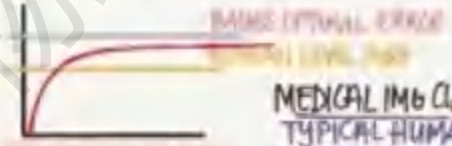
US
UK
EUROPE
S.AM
INDIA
CHINA
AUST.

OPTION 1:
DEV = UK, US, EUR
TEST = REST



IF DEV & TEST ARE DIFF & WE OPTIMIZE FOR DEV WE WILL MISS THE TEST-TARGET

HUMAN LEVEL PERF



WHY DOES ACC SLOW DOWN WHEN WE SURPASS HUMAN LEVEL PERF?

MEDICAL IMG CLASS
TYPICAL HUMAN 3%
TYPICAL DOCTOR 1%
EXPERIENCED DR. 0.7%
TEAM OF EXP DRS. 0.5%

↑ HUMAN LEV PERF (PROXY FOR BAYES)

1. OFTEN CLOSE TO BAYES
2. A HUMAN CAN NO LONGER HELP IMPROVE (INSIGHTS)
3. DIFFICULT TO ANALYSE BIAS/VARIANCE

CAT CLASSIFICATION

	A	B
HUMAN	1%	7.5%
TRAIN ERR	8%	8%
DEV ERR	10%	10%

B: BLURRY

7.5% → AVOIDABLE BIAS

8% → VARIANCE

FOCUS ON BIAS — FOCUS ON VARIANCE

HUMAN
| AVOIDABLE BIAS
TRAIN
| VARIANCE
DEV

- TRAIN BIGGER NETW.
- TRAIN LONGER / BETTER OPT. ALGOS (RMSPROP, ADAM)
- CHANGE NN ARCH OR HYPERPARAM
- MORE DATA (TRAIN)
- REGULARIZATION
- NN ARCHITECTURE

	A	B
HUMAN	0.5	0.5
TRAIN ERR	0.6	0.3
DEV ERR	0.8	0.4
AVOID. BIAS	0.1	?

0.5 → AVOIDABLE BIAS

0.3 → VARIANCE

? — DON'T KNOW IF WE OVERFIT OR IF WE'RE CLOSE TO BAYES

OPTIONS TO PROCEED ARE UNCLEAR

STRUCTURING ML PROJECTS · COURSERA
知乎 @Sophia @TestDriveme

ERROR
ANALYSIS

YOU HAVE 10% ERRORS, SOME ARE DOGS MIS-CLASSIFIED AS CATS. SHOULD YOU TRAIN ON MORE DOG PICS?

1. PICK 100 MIS-LABLED
2. COUNT ERROR REASONS

	DOG	BLURRY	INSTA FILTER	BIG CAT	...
1	1		1		
2				1	
3		1			
...					
100			1		
	5	...			

5% OF ALL ERRORS

FOCUSING ON DOGS: THE BEST WE CAN HOPE FOR IS 9.5% ERROR

YOU FIND SOME INCORR. LABLED DATA IN THE DEV SET. SHOULD YOU FIX IT?

DL ALGORITHMS ARE PRETTY ROBUST TO RANDOM ERRORS. BUT NOT TO SYSTEMATIC ERR.

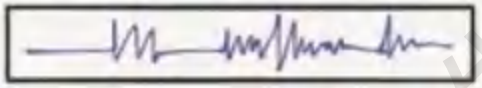
(EX. ALL WHITE CATS INCORR. LABELED AS MIKE)

ADD EXTRA COL. IN ERROR ANALYSIS AND USE SAME CRITERIA

NOTE IF YOU FIX DEV YOU SHOULD FIX TEST AS WELL.

FOR NEW PROJ: BUILD 1st SYSTEM QUICK & ITERATE

EX: SPEECH RECOGNITION



WHAT SHOULD YOU FOCUS ON?

NOISE
ACCENTS
FAR FROM MIKE

1. START QUICKLY
   DEV/TEST METRICS
2. GET TRAIN-SET
3. TRAIN
4. BIAS/VARIANCE ANAL
5. ERROR ANALYSIS
6. PRIORITIZE NEXT STEP

STRUCTURING AI PROJECTS - COURSERA
生物统计学2024
@Sophie @Tessfranklez

TRAIN vs DEV/TEST MISMATCH

AVAILABLE DATA

200k PRO CAT PICS FROM INTERNET

10k BLURRY CAT PICS FROM APP

WHAT WE CARE ABT

HOW DO WE SPLIT → TRAIN/DEV/TEST?

OPTION 1: SHUFFLE ALL

205k (TRAIN)	D	T

2.5k

PROBLEM: DEV/TEST IS NOW MOSTLY WEB IMG (NOT REPRES. OF END SCENARIO)

SOLUTION: LET DEV/TEST COME FROM APP. THEN SHUFFLE 5k OF APP PICS W/ WEB FOR TRAIN

205k	2.5	2.5

WEB+APP APP APP

BIAS & VARIANCE W/ MISMATCHED TRAIN/DEV

HUMANS ~0%
TRAIN 1%
DEV ERR 10%

IS THIS DIFF DUE TO THE MODEL NOT GENERALIZING OR IS DEV DATA MUCH HARDER

A: CREATE A TRAIN-DEV SET THAT WE DON'T TRAIN ON

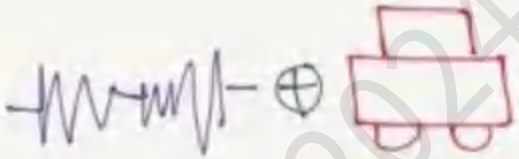
TRAIN	TD	D	T

	A	B	C	D
TRAIN	1%	1%	10%	10%
TRAIN-DEV	9%	15%	11%	11%
DEV	10%	10%	12%	20%
	VARIANCE	TRAIN/DEV MISMATCH	BIAS	BIAS + DATA MISMATCH

ADDRESSING DATA MISMATCH

EX. CAR GPS · TRAINING DATA IS 10,000H OF GENERAL SPEECH DATA

1. CARRY OUT MANUAL ERROR ANALYSIS TO UNDERSTAND THE DIFFERENCE (EX NOISE, STREET NUMBERS)
2. TRY TO MAKE TRAIN MORE SIMILAR TO DEV OR GATHER MORE DEV-LIKE TRAIN-DATA



⇒ AUDIO W CAR NOISE

NOTE: BE CAREFUL · IF YOU ONLY HAVE 1 HR OF CAR NOISE & APPLY IT TO 10K HR SPEECH YOU MAY OVERFIT TO THE CAR NOISE

知乎 @Sophia ©TestDaily

STRUCTURING ML PROJECTS · COURSERA

EXTENDED LEARNING

TRANSFER LEARNING

PROBLEM: YOU WANT TO CLASSIFY SOME MEDICAL IMG. YOU HAVE AN NN THAT CLASSIFIES CATS

NN TO CLASSIFY CATS/DOGS/CARS

OPTION 1: YOU ONLY HAVE A FEW RADIOLOGY IMAGES

SOLUTION: INIT W. WEIGHTS FROM CAT NN
ONLY RETRAIN LAST LAYER(S) ON RADIOLOGY IMAGES

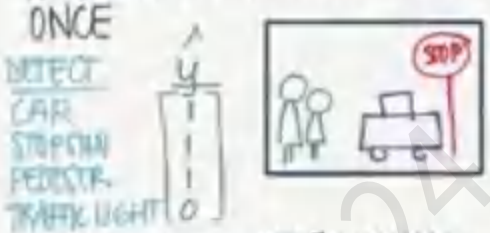
OPTION 2 YOU HAVE LOTS OF RADIOLOGY IMG.

SOLUTION: INIT WITH WEIGHTS FROM CAT NN
RETRAIN ALL LAYERS

THIS IS MICROSOFT CUSTOM VISION

MULTI TASK LEARNING

TRAINING ON MULT. TASKS AT ONCE



UNLIKE SOFMAX · MANY THINGS CAN BE TRUE

COST: $J(W,B) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{C} \mathcal{L}(\hat{y}_j^{(i)}, y_j^{(i)})$

SUMMING OVER ALL OUTP OPTIONS

WE COULD HAVE JUST TRAINED 4 NNS INSTEAD BUT... MT LEARNING MAKES SENSE WHEN

A. THE LEARNING DATA YOU HAVE FOR THE DIFF TASKS IS QUITE SIMILAR – & THE AMOUNTS (EG. 1K CARS, 1K STOP SIGNS)

B. THE SUM OF THE DATA ALLOWS YOU TO TRAIN A BIG ENOUGH NN TO DO WELL ON ALL TASKS

IN REALITY TRANSFER LEARNING IS USED MORE OFTEN

END-TO-END LEARNING

FROM X-RAY OF CHILDS HAND TELL ME THE AGE OF THE CHILD

TYPICAL SOLN:

1. LOCATE BONES TO FIND LENGTHS USING ML
2. TRAIN MODEL TO PREDICT AGE BASED ON BONELENGTH

END-TO-END

RADIOLOGY IMG → CHILD AGE

PROS:

- LET'S THE DATA SPEAK (MAYBE IT FINDS RELATIONS WE'RE UNAWARE OF)
- LESS HAND-DESIGNING OF COMPONENTS NEEDED

CONS:

- NEEDS LARGE AMTS OF LABELED DATA (X→Y)
- EXCLUDES POTENTIALLY USEFUL HANDMADE COMPONENTS

STRUCTURING ML PROJECTS · COURSERA
24生物统计学2024 生物统计学2024 生物统计学2024
© TessFerrandez

CONVOLUTION

FUNDAMENTALS

COMPUTER VISION

IMAGE CLASSIFICATION — CAT OR NOT-CAT

OBJECT DETECTION — WHERE IS THE CAR?

NEURAL STYLE TRANSFER — PAINT ME LIKE PICASSO

PROBLEM: IMAGES CAN BE BIG

1000×1000×3 (RGB) = 3M

WITH 1000 HIDDEN UNITS WE NEED 3M * 1000 = 3B PARAMS

SOLUTION: USE CONVOLUTIONS

IT'S LIKE SCANNING OVER YOUR IMG WITH A MAGNIFYING GLASS OR FILTER

TESS NOTE: ALSO SOLVES THE PROBLEM THAT THE CAT IS NOT ALWAYS IN THE SAME LOCATION IN THE IMG

CONVOLUTION

$3+1+2+0+0+0-1-8-2=-5$

(3×1)

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

INPUT 6×6 IMAGE

$*$

1	0	-1
1	0	-1
1	0	-1

FILTER 3×3

CONVOLUTION

$=$

-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

OUTPUT 4×4 IMAGE

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

INPUT 6×6 IMAGE

$*$

VERTICAL EDGE DETECTOR

1	0	-1
1	0	-1
1	0	-1

FILTER 3×3

$=$

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0

OUTPUT 4×4 IMAGE

DETECTED EDGE IN THE MIDDLE

THIS IS LIKE ADDING AN 'INSTA' FILTER THAT JUST SHOWS OUTLINES

WE COULD HARD-CODE FILTERS · JUST LIKE WE CAN HARD-CODE HEURISTIC RULES... BUT.... A MUCH BETTER WAY IS TO TREAT THE FILTER # AS PARAMS TO BE LEARNED

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

CONVOLUTIONAL NEURAL NETS · COURSERA

知乎 @Sophia
@TessFerrandez

CONVOLUTIONAL NEURAL NETS · COURSERA

PADDING

PROBLEM: IMAGES SHRINK
6x6 → 3x3 → 4x4
PROBLEM: EDGES GET LESS 'LOVE'

SOLUTION: PAD W. A BORDER OF 0s BEFORE CONVOLVING

TWO COMMONLY USED PADDING OPTIONS
(HOW MUCH TO PAD)

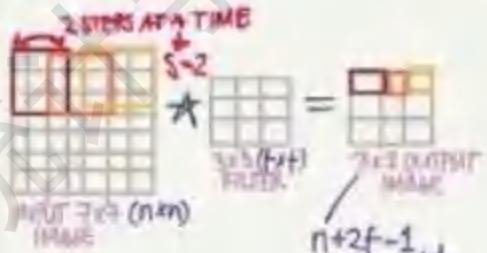
'VALID' ⇒ $P = 0$ NO PADDING

'SAME' ⇒ $P = \frac{f-1}{2}$ OUTPUT SIZE = INPUT SIZE

FILTER SIZE

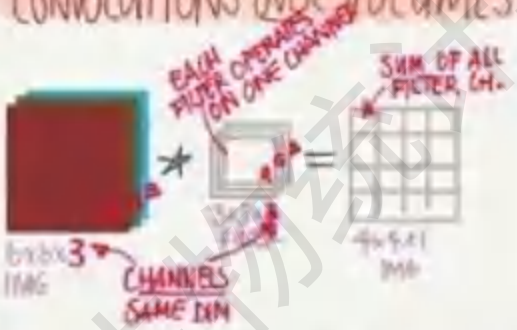
STRIDE

WHAT PACE YOU SCAN WITH



$$\frac{n+2p-f}{s}+1$$

CONVOLUTIONS OVER VOLUMES

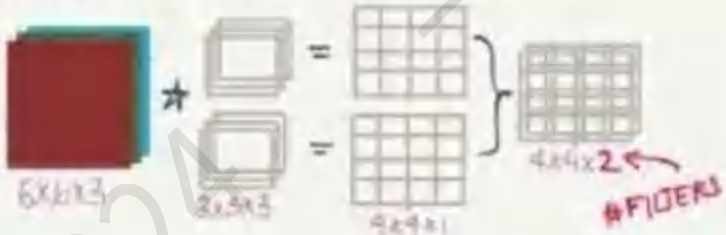


THIS ALLOWS US TO DETECT FEATURES IN COLOR IMAGES FOR EXAMPLE

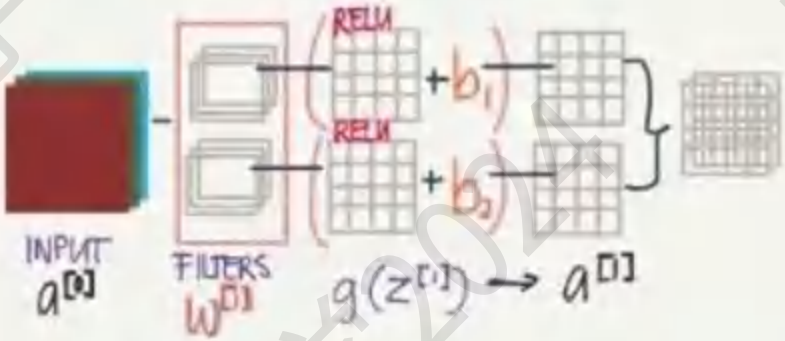
MAYBE WE WANT TO FIND ALL EDGES OR MAYBE ORANGE BLOBS

MULTIPLE FILTERS

DETECTING MULTIPLE FEATURES AT A TIME



ONE CONV NET LAYER



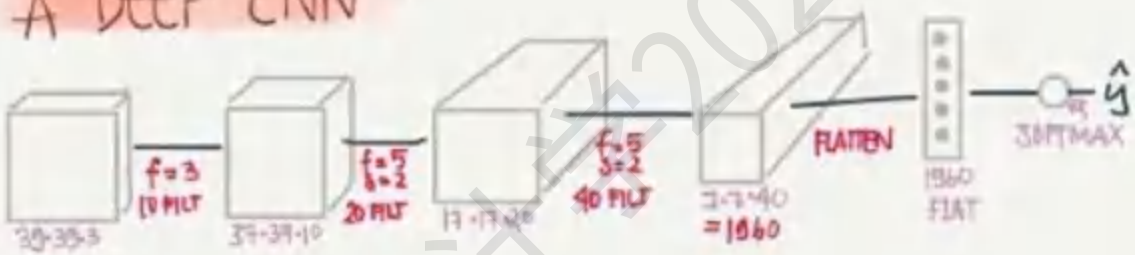
NOTE IT DOESN'T MATTER HOW BIG THE INPUT IS - THE LEARNABLE PARAMS W & b ONLY DEPEND ON THE # OF FILTERS AND THEIR SIZES.

$W = 3 \cdot 3 \cdot 3 \cdot 2 = 54$
$b = 2$
56 PARAMS TO LEARN

NOTE: ALL CONVOLUTION IDEAS CAN BE APPLIED TO 1D AS WELL LIKE EKG SIGNALS · AND 3D LIKE CT-SCANS

知乎 @Sophia
@Rosanne

A DEEP CNN



A LOT OF THE WORK IS FIGURING OUT HYPERPARAMS
= #FILTERS, STRIDE, PADDING ETC

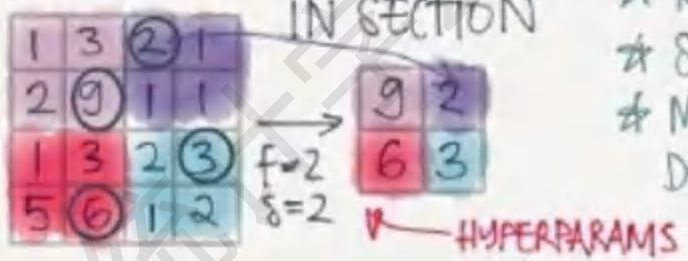
TYPICALLY SIZE → TREND DOWN
FILTERS → TREND UP

TYPICAL CONV.NET LAYERS

CONVOLUTION
POOLING
FULLY CONNECTED

POOLING (MAX)

FIND MAX VAL IN SECTION

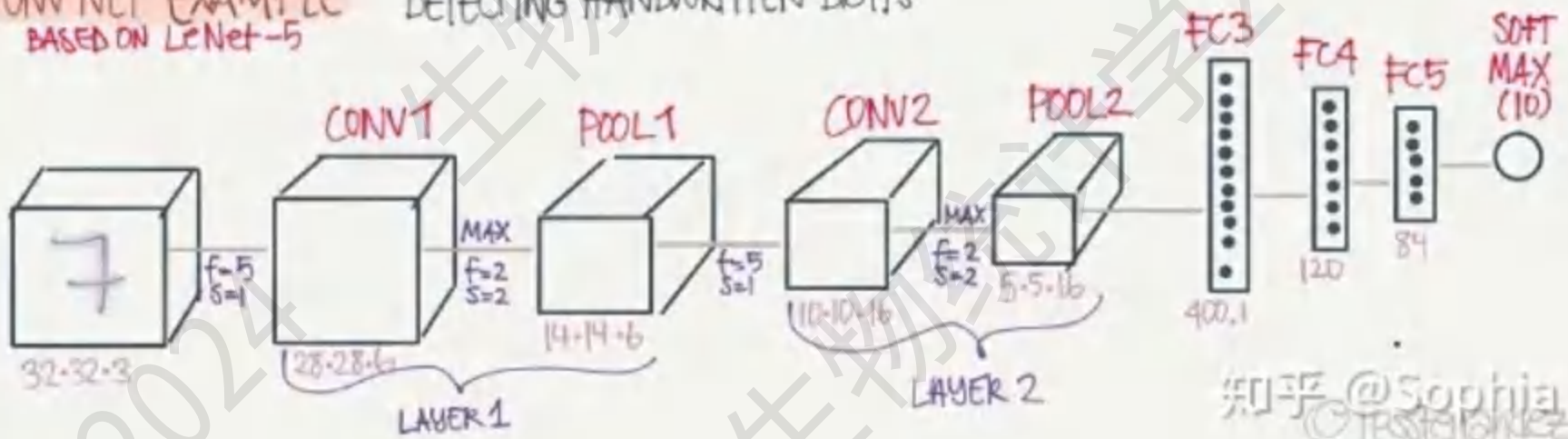


- ★ REDUCES SIZE OF REPRES.
- ★ SPEEDS UP COMPUTATION
- ★ MAKES SOME OF THE DETECTED FEAT. MORE ROBUST

CONV NET EXAMPLE

BASED ON LeNet-5

DETECTING HANDWRITTEN DIGITS



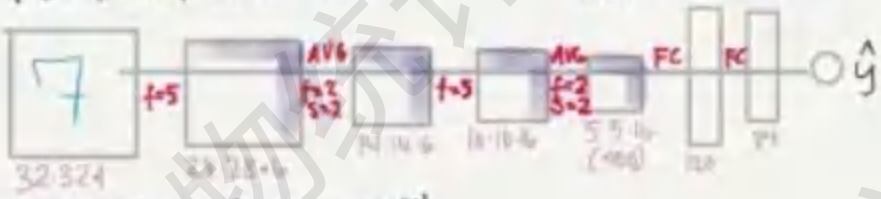
知乎 @Sophia
©TESSIONKEE

CONVOLUTIONAL NEURAL NETS · COURSERA

CLASSIC CONV. NETS

LeNet 5

DOCUMENT CLASSIFICATION ≈60k PARAMETERS



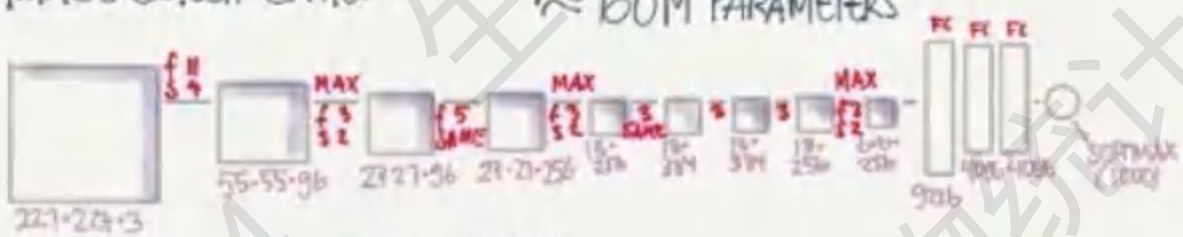
TRENDS: HEIGHT/WIDTH GO DOWN
CHANNELS GO UP

COMMON PATTERN: A COUPLE OF CONV(1')/POOL LAYERS FOLLOWED BY A FEW FC

OLD STUFF: USED AVG POOLING INST. OF MAX
PADDING WAS NOT VERY COMMON
IT USED SIGMOID/TANH INST OF RELU

AlexNet

IMAGE CLASSIFICATION ≈60M PARAMETERS



- SIMILAR TO LeNet BUT MUCH BIGGER
- USES RELU
- THE NN THAT GOT RESEARCHERS INTERESTED IN VISION AGAIN

VGG-16

ALL CONV. LAYERS HAVE SAME # PARAMS
f=3x3 S=1 P=SAME
AND POOLING LAYER 2x2 S=2

≈138 M PARAMETERS

- VERY DEEP
- EASY ARCHITECTURE
- # FILTERS DOUBLE 64, 128, 256, 512

CONVOLUTIONAL NEURAL NETS · COURSERA

知乎 @Sophia
© Tessa Gonzalez

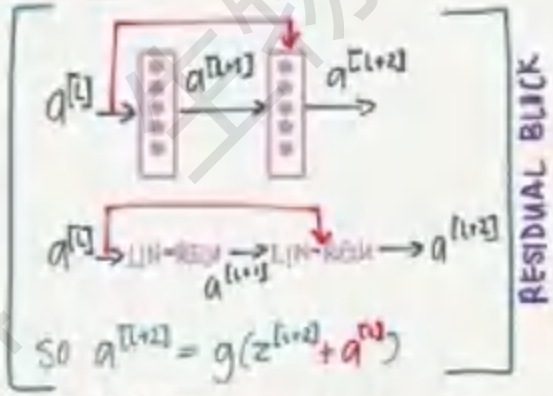
CONVOLUTIONAL NEURAL NETS · COURSERA

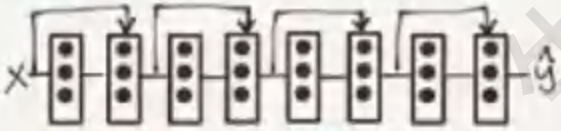
SPECIAL NETWORKS

ResNets

PROBLEM: DEEP NN OFTEN SUFFER PROBLEMS W VANISHING OR EXPLODING GRADIENTS

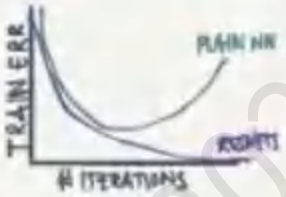
SOLUTION RESIDUAL NETS





VERY DEEP NN

$a^{[l+2]} = g(z^{[l+2]} + a^{[l]})$
$= (W^{[l+2]} a^{[l+1]} + b^{[l+2]} + a^{[l]})$



WORST CASE: W & b HAVE VANISHED TO 0 ⇒ $a^{[l+2]} = g(a^{[l]}) = a^{[l]}$ SO EXTRA LAYERS LEARNT NOTHING BUT DIDN'T HURT

BEST CASE: LAYERS LEARNED SOMETHING

NETWORK IN NETWORK (1×1 CONVOLUTION)

6	5	3	2
4	1	9	5
5	8	2	4
0	3	6	1

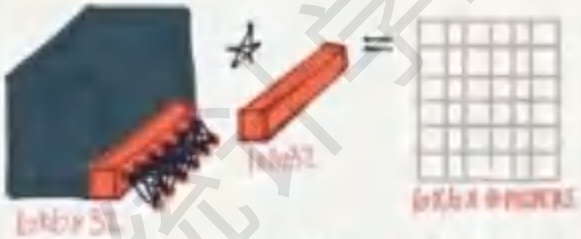
$\star$ $\boxed{2}$ =

12	10	6	4
8	2	18	10
10	16	4	8
0	6	12	2

1×1 CONVOLUTION

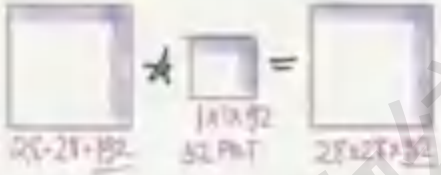
IT SEEMS PRETTY USELESS, BUT IT ACTUALLY SERVES 2 PURPOSES

1. NETWORK IN A NETWORK



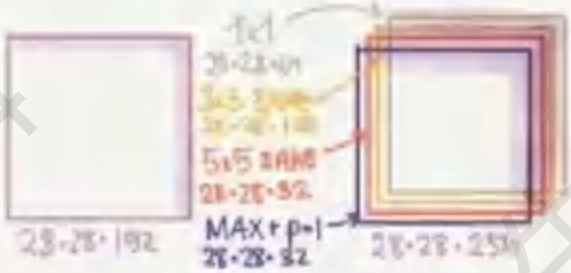
LEARNS COMPLEX, NON-LINEAR RELATIONSHIPS ABOUT A SLICE OF A VOLUME

2 REDUCING # CHANNELS



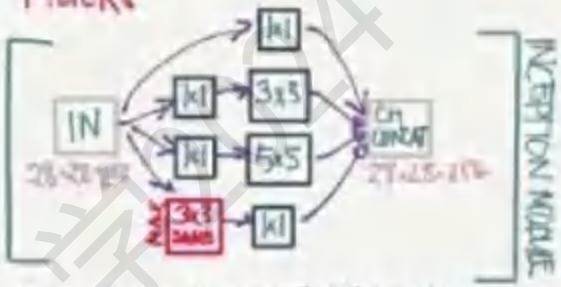
INCEPTION NETWORKS

INSTEAD OF CHOOSING A 1×1, 3×3, 5×5 OR A POOLING LAYER - CHOOSE ALL



PROBLEM: VERY EXPENSIVE TO COMPUTE

SOLUTION: SHRINK THE # CHANNELS W/ A 1×1 CONV BEFORE APPLYING ALL THE FILTERS



TO BUILD AN INCEPTION NETWORK YOU MAINLY STACK A BUNCH OF INCEPTION MODULES

INCEPTION THE MOVIE

2024 生物统计学 @Sophia

PRACTICAL ADVICE

USE OPEN SOURCE IMPLEMENTATIONS

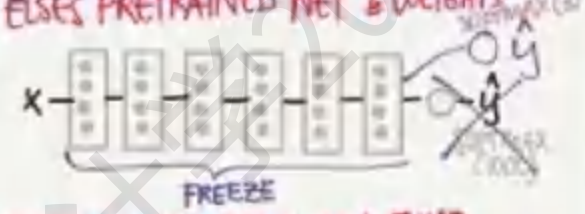
SOME OF THE PAPERS ARE HARD TO IMPLEMENT FROM SCRATCH - USING OS YOU CAN REUSE OTHER PPLS WORK

DON'T FORGET TO CONTRIBUTE

TRANSFER LEARNING

WANT TO TRAIN A CLASSIFIER FOR YOUR CATS BUT DON'T HAVE ENOUGH PICTURES

SOLUTION DOWNLOAD SOMEONE ELSES PRETRAINED NET & WEIGHTS



FREEZE THE PARAMS, AND JUST REPLACE THE SOFTMAX LAYER WITH YOUR OWN & TRAIN

IF YOU HAVE MORE PICS - RETRAIN A MORE OF THE LATER LAYERS (MAYBE INITIALIZING WITH THE PRETRAINED WEIGHTS)

STATE OF COMPUTER VISION

WE HAVE LOT OF DATA

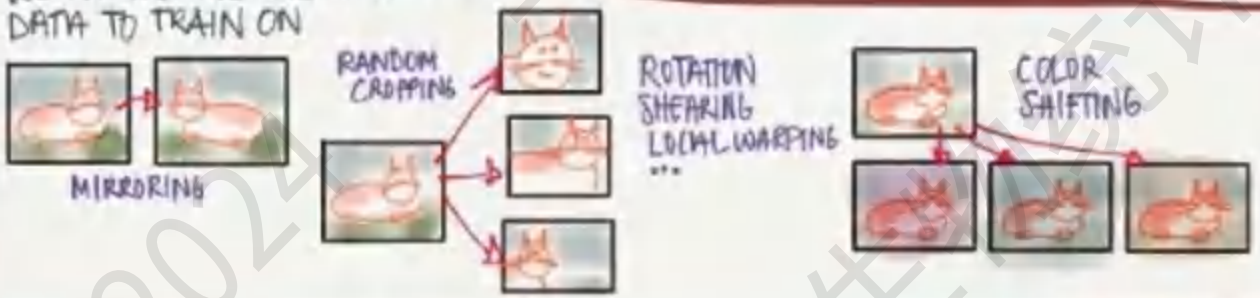
- SPEECH RECOG.
- IMAGE RECOGNITION
- OBJECT DETECTION
  (BAD W/ LIMITED DATA)

WE HAVE LITTLE LABELED DATA

MORE HAND ENGINEERING

DATA AUGMENTATION

WE ALMOST ALWAYS NEED MORE DATA TO TRAIN ON



TIPS FOR DOING WELL ON BENCHMARKS/COMPETITIONS

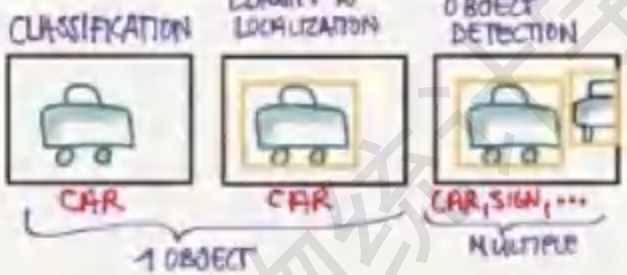
- ENSEMBLING
  AVG OUTPUTS FROM MULT NN
- MULTI-CROP AT TEST TIME
  AVG OUTPUTS FROM MULTIPLE CROPS OF THE IMAGE

IN PRACTICE THEY ARE NOT USED IN PRODUCTION BECAUSE THEY ARE COMPUT & MEM EXPENSIVE

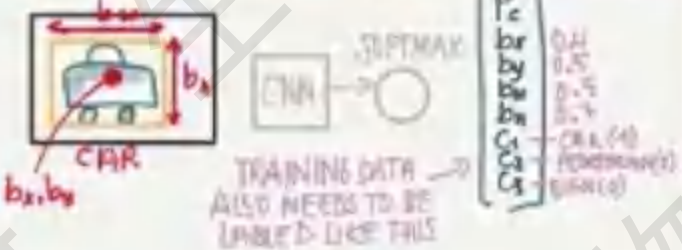
CONVOLUTIONAL NEURAL NETS · COURSERA

知乎 @Sophia

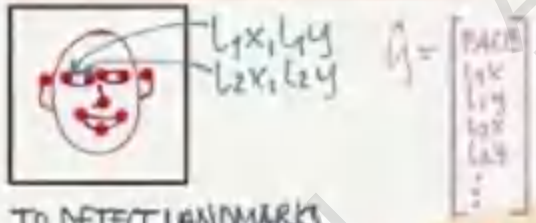
DETECTION ALGORITHMS



OBJECT LOCALIZATION



LANDMARK DETECTION



TO DETECT LANDMARKS IN THE FACE (CORNER OF MOUTH ETC) LABEL THE X,Y COORDS OF THE LANDMARK

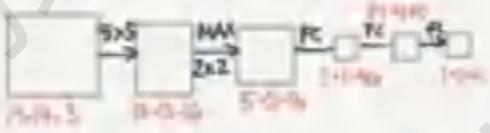
USED FOR SENTIMENT ANALYSIS & FOR EFFECTS LIKE PLACING CROWN ON HEAD ETC.

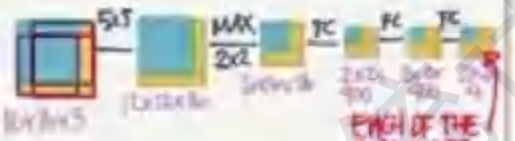
SLIDING WINDOWS DETECTION

1. CREATE TIGHTLY CROPPED IMG OF CARS (LOTS)
2. SLIDE A WINDOW OVER THE IMG. & CLASSIFY THIS WINDOW CAR (1/0) AGAINST YOUR OTHER CARS
3. REPEAT WITH SLIGHTLY LARGER WINDOW SIZE

PROBLEM: VERY EXPENSIVE (TO COMPUTE)

SINCE ALL WINDOWS SHARE A LOT OF THE COMPUTATIONS WE CAN DO THIS MUCH CHEAPER W/ CONVOLUTIONS





NOW WE JUST PASS THROUGH ONCE AND CALC ALL AT THE SAME TIME

YOLO - You Only Look Once

1. SPLIT IMG INTO X(9) GRID CELLS
2. FOR EACH CELL, SAY IF IT CONTAINS CAR + BOUNDING BOX (IF CELL CONTAINS THE MID POINT)

SAME AS OBJECT LOCALIZATION

IN PRACTICE WE MIGHT HAVE A 19x19 GRID

$X \times 9 = 3 \times 3 \times 8$

HOW DO YOU KNOW HOW GOOD IT IS?

HOW GOOD IS THE RED SQUARE?

$IOU = \frac{\text{SIZE OF}}{\text{SIZE OF}}$

INTERSECTION OVER UNION

GENERALLY - IF IOU ≥ 0.5 IT IS REGARDED AS CORRECT

WHAT IF MULTIPLE SQUARES CLAIM THE SAME CAR?

NON MAX SUPRESSION

IF TWO BOUNDING BOXES HAVE A HIGH IOU - PICK THE ONE W HIGHEST P_c - GET RID OF THE REST.

ANCHOR BOXES

ANCHOR BOXES LET YOU ENCODE MULTIPLE OBJECTS IN THE SAME SQUARE

CONVOLUTIONAL NEURAL NETS · COURSERA
2024 生物统计学 @Sophia @Tesla Mike

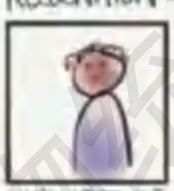
FACE RECOGNITION

FACE VERIFICATION

IS THIS PETE?

99% ACC ⇒ PRETTY GOOD

FACE REIDENTITION



WHO IS THIS? (OUT OF K PERSONS)

IF K = 100 NEED MUCH HIGHER THAN 99%

ONE SHOT LEARNING

NEED TO BE ABLE TO RECOGNIZE A PERSON EVEN THOUGH YOU ONLY HAVE ONE SAMPLE IN YOUR DB.

YOU CAN'T TRAIN A CNN WITH A SOFTMAX (EACH PERSON) BECAUSE

A) YOU DON'T HAVE ENOUGH SAMPLES

B) IF A NEW PERSON JOINS YOU NEED TO RETRAIN THE NETWORK.

SOLUTION LEARN A SIMILARITY FUNCTION

$d(\text{img1}, \text{img2}) = \text{degree of difference}$

BUT HOW DO YOU LEARN THIS?

SIAMESE NETWORK DeepFace

$$d(x^{(1)}, x^{(2)}) = \|f(x^{(1)}) - f(x^{(2)})\|_2^2$$

LEARN THE PARAMS OF THE NN SUCH THAT

- IF $x^{(1)}, x^{(2)}$ ARE THE SAME PERSON · $d(x^1, x^2)$ ⇒ SMALL
- IF $x^{(1)}, x^{(2)}$ ARE DIFFERENT PEOPLE · $d(x^1, x^2)$ ⇒ LARGE

WE CAN ACCOMPLISH THIS WITH THE TRIPLET LOSS FUNCTION

TRIPLET LOSS FaceNet

TIP PRECOMPUTE ENCODINGS FOR PPL IN YOUR DB, SO YOU DONT HAVE TO SAVE IMAGES & COMPUTE ENCODINGS AT RUN-TIME

WANT $\underbrace{\|f(A)-f(P)\|^2}_{d(A,P)} \le \underbrace{\|f(A)-f(N)\|^2}_{d(A,N)} \Rightarrow d(A,P) - d(A,N) \le 0$

BUT WE WANT A GOOD MARGIN, SO...

$d(A,P) - d(A,N) + \alpha \le 0$

HOW DO WE CHOOSE TRIPLETS TO TRAIN ON?

- IF A/P ARE VERY SIMILAR, & A/N ARE VERY DIFFERENT TRAINING IS VERY EASY.

SELECT A/N THAT ARE PRETTY SIMILAR TO TRAIN A GOOD NET

SOME BIG COMPANIES HAVE ALREADY TRAINED NETWORKS ON LARGE AMTS OF PHOTOS SO YOU MAY JUST WANT TO REUSE THEIR WEIGHTS

CONVOLUTIONAL NEURAL NETS · COURSERA

知乎 @Sophia
© TechFanatic

CONVOLUTIONAL NEURAL NETS · COURSERA

NEURAL STYLE TRANSFER

CONTENT (C) + STYLE (S) = GENERATED (G)

PICTURE OF CHARTRES | STARRY NIGHT - VAN GOGH | CHARTRES IN THE STYLE OF VAN GOGH

WE CAN VISUALIZE WHAT A NETWORK LEARNS BY LOOKING AT WHAT IMAGES (PARTS) ACTIVATED EACH UNIT MOST

LAYER 1 - SIMPLE FEATURES LIKE LINES/COLORS

LAYER 2 - MORE COMPLEX PATTERNS - BLOB LEGS, HONEYCOMB PATTERNS

LAYER 5 - EVEN MORE COMPLEX - WHALE, DOG, NET, A CAT DETECTOR

BUT HOW DOES THIS HELP US GENERATE AN IMAGE IN THE STYLE OF ANOTHER?

IDEA:

1. GENERATE A RANDOM IMG
2. OPTIMIZE THE COST FUNCTION

$$J(G) = \alpha J_{content}(C,G) + \beta J_{style}(S,G)$$

HOW SIMILAR ARE C & G | HOW SIMILAR ARE S & G

3. UPDATE EACH PIXEL

CONTENT COST FUNCTION

- USE A PRE-TRAINED CONVNET (EX VGG)
- SELECT A HIDDEN LAYER SOMEWHERE IN THE MIDDLE

LATER → CAPTURES LARGER FEATURES

- LET $a^{[l](C)}$ & $a^{[l](G)}$ BE THE ACTIVATIONS
- IF $a^{[l](C)}$ & $a^{[l](G)}$ ARE SIMILAR THEY HAVE SIMILAR CONTENT

BECAUSE THEY BOTH TRIGGER THE SAME HIDDEN UNITS

HOW DO WE TELL IF THEY ARE SIMILAR?

$$J_{content}(C,G) = \frac{1}{2}\|a^{[l](C)} - a^{[l](G)}\|^2$$

CAPTURING THE STYLE

USING THE STYLE IMG AND THE ACTIVATIONS IN A LAYER.
LOOK THROUGH THE ACTIVATIONS IN THE DIFFERENT CHANNELS TO SEE HOW CORRELATED THEY ARE

WHEN WE SEE PATTERNS LIKE THIS

DO WE USUALLY SEE IT WITH PATCHES LIKE THESE?

STYLE MATRIX

CREATE A MATRIX OF HOW CORRELATED THE ACTIVATIONS ARE, FOR EACH POS (x,y) & CHANNEL PAIR (K, K') FOR THE STYLE IMG & GENERATED

$$G_{kk'} = \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} a_{ijk} \cdot a_{ijk'}$$

THE STYLE COST FUNCTION

$$J(S,G) = \|G^{(S)} - G^{(G)}\|_F^2$$

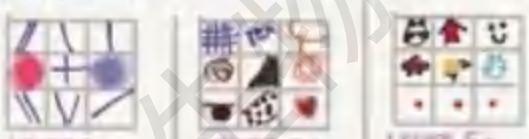
FROBENIUS NORM

TO GET MORE VISUALLY PLEASING IMAGES IF YOU CALC J(S,G) OVER MULTIPLE LAYERS

知乎 @Sophia Chen

NEURAL STYLE TRANSFER

WE CAN VISUALIZE WHAT A NETWORK LEARNS BY LOOKING AT WHAT IMAGES (PARTS) ACTIVATED EACH UNIT MOST



LAYER 1 - SIMPLE FEATURES LIKE LINES/COLORS

LAYER 2 - MORE COMPLEX FEATURES - RIVER LINES, HONEYCOMB PATTERNS

LAYER 5 - EVEN MORE COMPLEX - HUMAN, DOG, PERSON, CAT DETECTOR

BUT HOW DOES THIS HELP US GENERATE AN IMAGE IN THE STYLE OF ANOTHER?

IDEA:

1. GENERATE A RANDOM IMG
2. OPTIMIZE THE COST FUNCTION

$$J(G) = \alpha J_{content}(C,G) + \beta J_{style}(S,G)$$

HOW SIMILAR ARE C & G — HOW SIMILAR ARE S & G

3. UPDATE EACH PIXEL

CONTENT COST FUNCTION

- USE A PRE-TRAINED CONVNET (EX VGG)
- SELECT A HIDDEN LAYER SOMEWHERE IN THE MIDDLE
  LATER ⇒ CAPTURES LARGER FEATURES
- LET $a^{[l](C)}$ & $a^{[l](G)}$ BE THE ACTIVATIONS
- IF $a^{[l](C)}$ & $a^{[l](G)}$ ARE SIMILAR THEY HAVE SIMILAR CONTENT
  BECAUSE THEY BOTH TRIGGER THE SAME HIDDEN UNITS

HOW DO WE TELL IF THEY ARE SIMILAR?

$$J_{content}(C,G) = \frac{1}{2}\|a^{[l](C)} - a^{[l](G)}\|^2$$

CAPTURING THE STYLE

USING THE STYLE IMG AND THE ACTIVATIONS IN A LAYER.
LOOK THROUGH THE ACTIVATIONS IN THE DIFFERENT CHANNELS TO SEE HOW CORRELATED THEY ARE

WHEN WE SEE PATTERNS LIKE THIS

DO WE USUALLY SEE IT WITH PATCHES LIKE THESE?

STYLE MATRIX

CREATE A MATRIX OF HOW CORRELATED THE ACTIVATIONS ARE, FOR EACH POS (x,y) & CHANNEL PAIR (K,K') FOR THE STYLE IMG & GENERATED

$$G_{kk'} = \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} a_{ijk} \cdot a_{ijk'}$$

THE STYLE COST FUNCTION

$$J(S,G) = \|G^{(S)} - G^{(G)}\|_F^2$$

FROBENIUS NORM

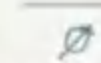
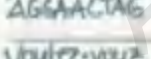
TO GET MORE VISUALLY PLEASING IMAGES IF YOU CALC J(S,G) OVER MULTIPLE LAYERS

CONVOLUTIONAL NEURAL NETS · COURSERA

知乎 @Sophia @Tesfandez

RECURRENT
NEURAL NETWORKS

SEQUENCE PROBLEMS

IN	OUT	PURPOSE
	THE QUICK BROWN FOX JUMPED...	SPEECH RECOGNITION
∅		MUSIC GENERATION
THERE IS NOTHING TO LIKE IN THIS MOVIE		SENTIMENT CLASSIFICATION
AGCCCCTGTG AGGAACTAG	AGCCCCTGTG AGGAACTAG	DNA SEQUENCE ANALYSIS
VOULEZ-VOUS CHANTER AVEC MOI?	DO YOU WANT TO SING WITH ME?	MACHINE TRANSLATION
	RUNNING	VIDEO ACTIVITY RECOGNITION
YESTERDAY HARRY POTTER MET HERMIONE GRANGER	YESTERDAY HARRY POTTER MET HERMIONE GRANGER	NAME ENTITY RECOGNITION

NAME ENTITY RECOGNITION

X = HARRY POTTER AND HERMIONE GRANGER INVENTED A NEW SPELL

$x^{<1>}$ $x^{<2>}$ ... $T_x = 9$ (9 WORDS)

$y = \underset{y^{<1>}}{1}\ \underset{y^{<2>}}{1}\ \underset{\cdots}{0}\ 1\ 1\ 0\ 0\ 0$ $T_y = T_x$

EXAMPLE FOR A PROBLEM WHERE EVERY $x^{<t>}$ HAS AN OUTPUT $y^{<t>}$

HOW DO WE REPRESENT WORDS?

CREATE A VOCABULARY (EG 10K MOST COMMON WORDS IN YOUR TEXTS - OR DOWNLOAD EXISTING)

a, aaron, and, harry, potter, zulu

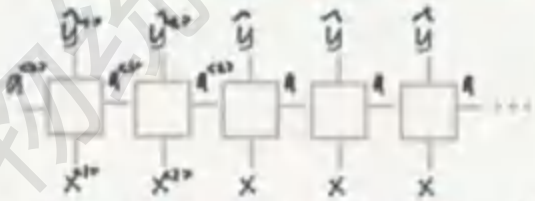
EACH WORD IS A **ONE-HOT** VECTOR

HARRY = [0, 0, ..., 1, ..., 0] (4075)

WE COULD USE A STANDARD NETWORK BUT...

1. INPUT & OUTPUT CAN HAVE DIFFERENT LENGTHS IN DIFF EXAMPLES
2. WE DON'T SHARE FEATURES LEARNED ACROSS DIFFERENT POSITIONS

RECURRENT NEURAL NET (RNN)



PREVIOUS RESULTS ARE PASSED IN AS INPUTS SO WE GET CONTEXT.

$a^{<t>} = g_1(W_a[a^{<t-1>}, x^{<t>}] + b_a)$ TANH/RELU

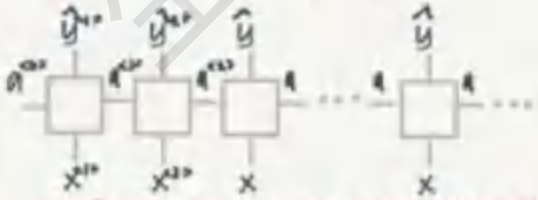
$\hat{y}^{<t>} = g_2(W_{ya} a^{<t>} + b_y)$ SIGMOID

THE SAME W & b ARE USED IN ALL TIME STEPS

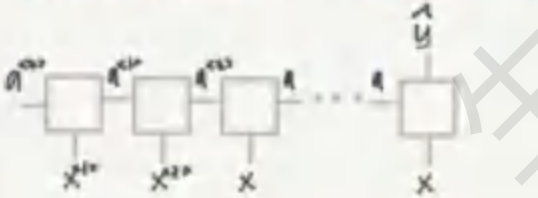
THE LOSS WE OPTIMIZE IS THE SUM OF $\mathcal{L}(\hat{y}, y)$ FROM 1–T

DIFFERENT TYPES OF RNN

MANY-TO-MANY $T_x = T_y$

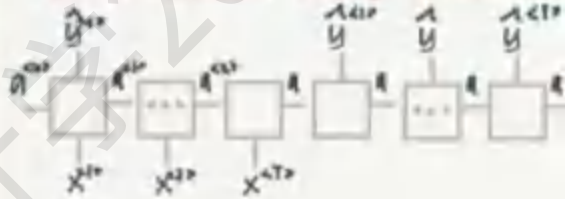


MANY-TO-ONE EX. SENTIMENT ANALYSIS



ONE-TO-MANY - MUSIC GENERATION

MANY-TO-MANY $T_x \neq T_y$ TRANSLATION



SEQUENCE MODELS · COURSERA
知乎 @Sophia
@Yesfranciz

NLP

WORD EMBEDDINGS

MAN IS TO WOMAN AS KING IS TO QUEEN

PROBLEM: THE ONE-HOT REPR O_{6257} OF APPLE HAS NO INFO ABOUT ITS RELATIONSHIP TO O_{6257} ORANGE

I WANT A GLASS OF ORANGE ____
I WANT A GLASS OF APPLE ____

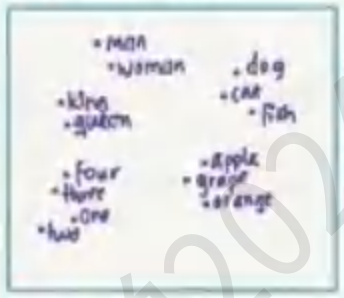
SOLUTION: CREATE A MATRIX OF FEATURES TO DESCRIBE THE WORDS

WORD EMBEDDINGS

	MAN (5391)	WOMAN (9853)	KING (4914)	QUEEN (7157)	APPLE (456)	ORANGE (6257)
GENDER	-1	1	-0.95	0.97	0.00	0.01
ROYAL	0.01	0.02	0.93	0.95	-0.01	0.00
AGE	0.03	0.02	0.7	0.69	0.03	-0.02
FOOD	0.09	0.01	0.02	0.01	0.94	0.96
⋮						

e_{5391}

IN REALITY · THE FEATURES ARE LEARNED & NOT AS STRAIGHTFWD AS GENDER/AGE



t-SNE

VISUAL REPRESENT OF 300D WORD EMBEDDINGS

USING WORD EMBEDDINGS

EX. NAME/ENTITY RECOGN

SALLY SMITH IS AN ORANGE FARMER
ROBERT LIN IS AN APPLE FARMER
BOB JONES IS A DURIAN CULTIVATOR

WITH WORD EMBEDDINGS WE UNDERSTAND THAT AN ORANGE FARMER IS A PERSON ⇒ SALLY SMITH = NAME

- APPLE ~ ORANGE ⇒ APPLE FARMER = PERSON
- USING WORD EMBEDDINGS TRAINED ON LRGE OF TEXT WE ALSO GET EMB. FOR MORE UNCOMMON WORDS (DURIAN, CULTIVATOR)

EX. MAN IS TO WOMAN AS KING IS TO ?

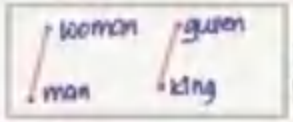
E = EMBEDDING MATRIX

e_{man}

	MAN (5391)	WOMAN (9853)	KING (4914)	QUEEN (7157)
GENDER	-1	1	-0.95	0.97
ROYAL	0.01	0.02	0.93	0.95
AGE	0.03	0.02	0.7	0.69
FOOD	0.09	0.01	0.02	0.01

$e_{man} - e_{woman}$ $\quad$ $e_{king} - e_{queen}$

$\begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ VERY SIMILAR $\begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$



FIND MAN → WOMAN KING → ?

FIND(w):

$\text{ARG.MAX}\ \text{SIM}(e_w, e_{king} - e_{man} + e_{woman})$

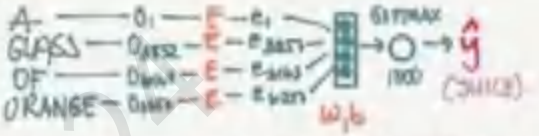
$\text{SIM}(u,v) = \frac{u^T v}{\|u\|\|v\|}$ COSINE SIMILARITY

LEARNING WORD EMBEDDINGS

HOW DO WE LEARN THE EMBEDDING MATRIX E?

IDEA 1: USING A NEURAL LANG MODEL

I WANT A GLASS OF ORANGE $\hat{y}$



WE CAN USE DIFFERENT CONTEXTS THAN THE LAST 4 WORDS

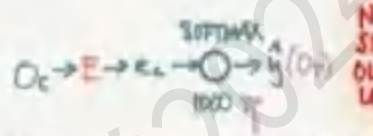
- LAST 4 WORDS
- 4 WORDS LEFT + RIGHT
- LAST 1 WORD
- NEARBY 1 WORD ← SKIP-GRAM: RANDOM WITHIN EX 5 WORDS

IDEA 2: SKIP-GRAMS WORD2VEC

I WANT A GLASS OF ORANGE JUICE TO GO ALONG WITH MY CEREAL

PICK RANDOM CONTEXT/TARGET PAIRS (WITHIN EX 5 WORDS)

CONTEXT	TARGET
ORANGE	JUICE
ORANGE	GLASS
ORANGE	MY
...	...



NOTE: WHILE THIS SIMPLE NN PREDICTS O_t OUR REAL GOAL IS TO LEARN E

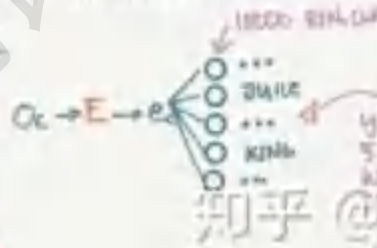
THIS IS VERY COMPUTATIONALLY EXPENSIVE BECAUSE CAN ESTIMATE BY USING A HIERARCHICAL SOFTMAX CLASSIFIER

IDEA: NEGATIVE SAMPLING

1. PICK A CONTEXT/TARGET PAIR AS A POSITIVE EXAMPLE
2. PICK A FEW NEG EXAMPLES CONTEXT + RANDOM

CONTEXT	WORD	TARGET	
ORANGE	JUICE		1
ORANGE	APPLE		0
ORANGE	BOOK		0
ORANGE	THE		0
ORANGE	...		0

NOTE: SOMETIMES BY CHANCE YOU PICK A POS PAIR · BUT IT DOESN'T MATTER



YOU ONLY TRAIN 5 FOR EACH CONTEXT WORD ⇒ EFFICIENT

知乎 @Sophia
@TeaTimeDize
SEQUENCE MODELS · COURSERA

WORD
EMBEDDINGS
CONTINUED...

GloVe WORD VECTORS

X_{ij} = # TIMES WORD i (TARGET t) APPEARS IN THE CONTEXT OF j (CONTEXT c)
(WORD RELATED THEY ARE)

$$\text{MINIMIZE} \sum_{i=1}^{10k} \sum_{j=1}^{10k} f(x_{ij})(\theta_i^T e_j + b_i + b_j - \log x_{ij})^2$$

f IS AN EXPONENT
(DON'T WEIGH WORDS VERY FREQUENT (THE, OF...) & VERY INFREQUENT (DURIAN))

EVERYTHING LED UP TO THIS VERY SIMPLE ALGORITHM

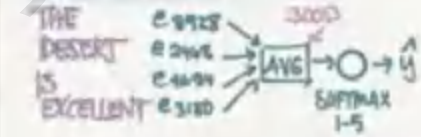
SENTIMENT CLASSIFICATION

X	y
THE DESSERT IS EXCELLENT	★★★★★
SERVICE WAS QUITE SLOW	★★
GOOD FOOD, QUICK MEAL BUT NOTHING SPECIAL	★★★
COMPLETELY LACKING IN GOOD TASTE GOOD SERVICE AND GOOD AMBIENCE	★

PROBLEM: YOU MAY NOT HAVE A LARGE DATASET

BUT YOU CAN USE AN EMBEDDING MATRIX E THAT IS ALREADY PRE-TRAINED

IDEA: SIMPLE CLASSIFICATION

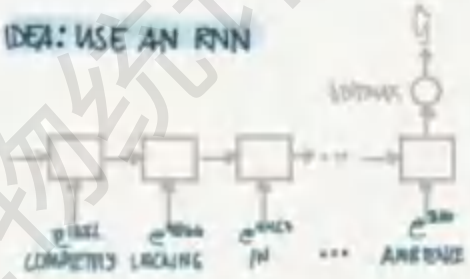


WORKS WELL FOR SHORT SENTENCES BUT DOESN'T TAKE ORDER INTO ACCOUNT

"COMPLETELY LACKING IN GOOD TASTE, GOOD SERVICE AND GOOD AMBIENCE"

THIS MAY BE SEEN AS A PN REVIEW

IDEA: USE AN RNN



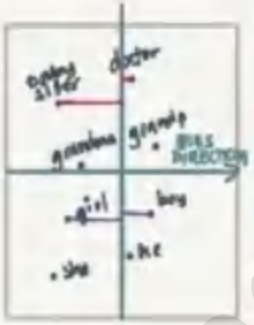
THIS CAN NOW TAKE INTO ACCOUNT THAT COMPLETELY LACKING NEGATES THE WORD GOOD

ELIMINATING BIAS IN WORD EMBEDDINGS

MAN IS TO COMPUTER PROGRAMMER AS WOMAN IS TO HOME MAKER

SOMETIMES THE TEXT CONTAINS & ALGOS LEARN A GENDER, RACE, AGE... BIAS WE DON'T WANT OUR MODELS TO HAVE - EX. HIRING BASED ON GENDER, SENTENCING BASED ON RACE ETC.

ADDRESSING BIAS



1. IDENTIFY BIAS DIRECTION
   $e_{he} - e_{she}$
   $e_{male} - e_{female}$
2. NEUTRALIZE
   FOR EVERY WORD THAT IS NOT DEFINITIONAL (GIRL, BOY, HE, SHE...) PROJECT TO GET RID OF BIAS
3. EQUALIZE PAIRS
   THE ONLY DIFF BETWEEN EX GIRL/BOY SHOULD BE GENDER

HOW DO YOU KNOW WHICH WORDS TO NEUTRALIZE?

DOCTOR, BEARD, SEWING MACHINE?

A: BY TRAINING A CLASSIFIER TO FIND OUT IF A WORD IS DEFINITIONAL

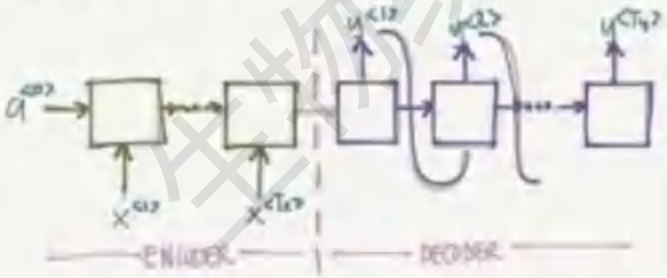
TURNS OUT THE # OF PAIRS IS FAIRLY SMALL SO YOU CAN EVE HAND PICK THEM

SEQUENCE MODELS · COURSERA
知乎 @Sophia
@TessFerrandez

SEQUENCE TO SEQUENCE

BASIC MODELS

JANE VISITE L'AFRICE EN SEPTEMBRE → JANE IS VISITING AFRICA IN SEPTEMBER



→ THIS IS A CAT ON A CHAIR

CNN → RNN

HOW DO YOU PICK THE MOST LIKELY SENTENCE?

$$P(\overset{\text{ENGLISH}}{y^{<1>}, \dots, y^{<T_y>}} \mid \overset{\text{FRENCH}}{x})$$

WE DON'T WANT A RANDOMLY GENERATED SENTENCE (WE WOULD SOMETIMES GET A GOOD, SOMETIMES BAD)

INSTEAD WE WANT TO MAXIMIZE

$$\text{ARG MAX } P(y^{<1>}, \dots, y^{<T_y>} \mid x)$$

IDEA: USE GREEDY SEARCH

1. PICK THE WORD WITH THE BEST PROBABILITY
2. REPEAT UNTIL <EOS>

WITH THIS WE COULD GET

- JANE IS GOING TO BE VISITING AFRICA THIS SEPTEMBER

INSTEAD OF

- JANE IS VISITING AFRICA THIS SEPTEMBER

$\boxed{\text{SOLUTION}}$ OPTIMISE THE PROB OF THE WHOLE SENTENCE INSTEAD

BEAM SEARCH

1. PICK THE FIRST WORD

   PICK THE B (EX 3) BEST ALTERNATIVES (IN, JANE, SEPTEMBER)

2. FOR EACH B WORDS PICK THE NEXT WORD AND EVALUATE THE PAIRS TO END UP W/ B PAIRS

   $P(y^{<1>}, y^{<2>} \mid x) = P(y^{<1>} \mid x)\, P(y^{<2>} \mid x, y^{<1>})$

NOTE: KEEP TRACK OF THE P FOR THE SENTENCES OF EACH LENGTH - AFTER X TIME (X MAX WORDS) PICK THE BEST

(IN SEPTEMBER, JANE IS, JANE VISITS)

3. REPEAT TIL DONE

$$\text{ARG MAX} \prod_{t=1}^{T_y} P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>})$$

PROBLEM: MULTIPLYING PROBABILITIES RESULTS IN A VERY SMALL NUMBER (OVERFLOWS!)

PROBLEM: IF WE OPTIMIZE FOR THE MULT WE WILL PREFER SHORT SENTENCES. SINCE EACH WORD WILL REDUCE PROB

INSTEAD WE CAN OPTIMIZE FOR THIS

$$\frac{1}{T_y^{\alpha}} \sum_{t=1}^{T_y} \log(P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>}))$$

HOW DO WE PICK B?

LARGE B: BETTER RESULT, SLOWER

SMALL B: WORSE RESULT, BETTER

IN PRID YOU MIGHT SE B=10. 100 IS PROBABLY A BIT TOO HIGH - BUT ITS DOMAIN DEPENDENT

ERROR ANALYSIS IN BEAM S

HUMAN: JANE VISITS AFRICA IN SEPT... y^*

ALGO: JANE VISITED AFRICA LAST SEPTEMBER $\hat{y}$

HOW DO WE KNOW IF ITS OUR RNN OR OUR BEAM SEARCH WE SHOULD WORK ON?

LET THE RNN GIVE $P(y^* \mid x) = P(y^*)$ & $P(\hat{y}) = P(\hat{y})$

IF $P(y^*) > P(\hat{y})$:

BEAM PICKED THE WRONG ONE TRY A HIGHER B

ELSEIF

THE RNN PICKED THE WRONG PROBABILITIES (OR BROKEN)

SEQUENCE MODELS · COURSERA
2024 生物统计学
知乎 @Sophia
@TessFerrandez

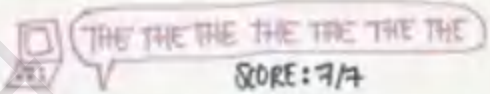
SEQUENCE TO SEQUENCE

FRENCH: LE CHAT EST SUR LE TAPIS
HUMAN1: THE CAT IS ON THE MAT
HUMAN2: THERE IS A CAT ON THE MAT

HOW DO YOU EVALUATE THE MACHINE TRANSLATION WHEN MULTIPLE ARE RIGHT?

BLEU SCORE

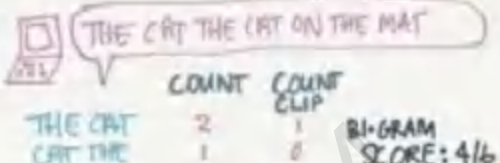
IDEA: CHECK IF THE WORDS ME APPEAR IN THE REAL TRANSLATION



SCORE: 7/7

IDEA: ONLY GIVE CREDIT FOR A WORD THE MAX # TIMES IT APPEARS IN A TARGET SENTENCE

SCORE: 2/7 — COUNT CLIP



BIGRAMS	COUNT	COUNT CLIP
THE CAT	2	1
CAT THE	1	0
CAT ON	1	1
ON THE	1	1
THE MAT	1	1
	6	4/6

BI-GRAM SCORE: 4/6

COMBINED BLEU SCORE

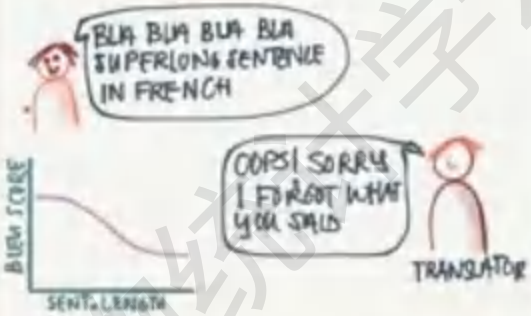
$$BP \cdot \exp\left(\frac{1}{4}\sum_{n=1}^{4} P_n\right)$$

P_1 → SCORE SINGLE WORD
P_2 → SCORE BIGRAMS
...
BP → BREVITY PENALTY

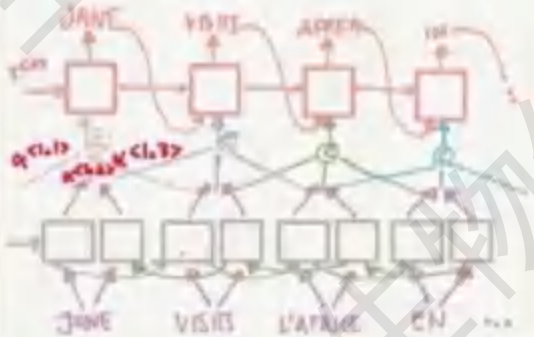
PENALIZES SENTENCES SHORTER THAN THE TARGET

A USEFUL SINGLE NUMBER EVAL METRIC

ATTENTION MODEL



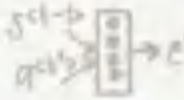
SOLUTION: TRANSLATE A LITTLE AT A TIME USING ONLY PARTS OF THE SENTENCE AS CONTEXT



$\alpha^{<t,t'>}$ = HOW MUCH ATTENTION $y^{<t>}$ SHOULD PAY TO $a^{<t'>}$

$$c^{<2>} = \sum_{t'} \alpha^{<2,t'>} \cdot a^{<t'>} \quad \Big| \quad \sum_{t'} \alpha^{<1,t'>} = 1$$

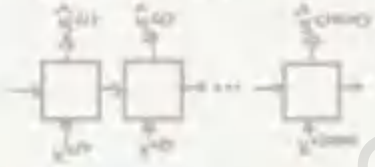
α IS CALCULATED USING A SMALL NEURAL NETWORK



$$\alpha^{<1,t'>} = \frac{\exp(e^{<1,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<1,t'>})}$$

SPEECH RECOGNITION

THE QUICK BROWN FOX ...



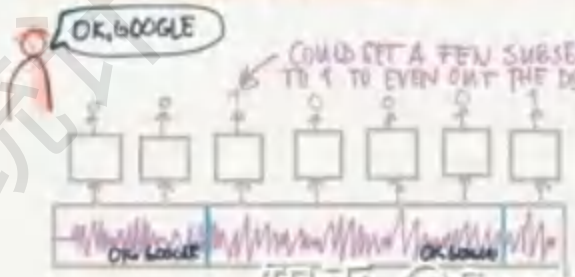
PROBLEM: 10 s CLIP AT 100 Hz = 1000 INPUTS BUT ONLY ≈ 20 OUTPUTS

SOLUTION: USE CTC COST (CONNECTION TEMPORAL CLASSIFICATION)

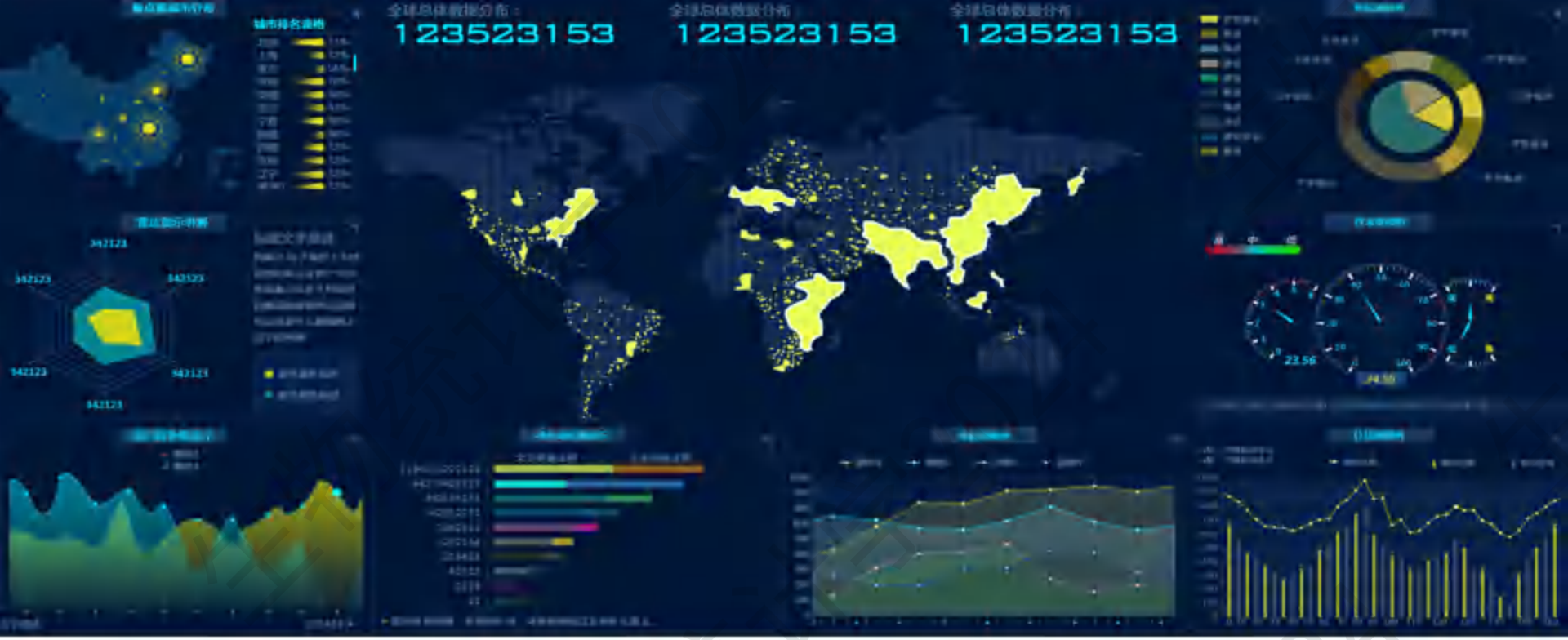
ttt_h_eee ___ll_l___ll___ooo → C

COLLAPSE REPEATED CHARS NOT SEP BY BLANK

TRIGGER WORD DETECTION



SEQUENCE MODELS · COURSERA
知乎 @Sophia
@TessFerrandez


123523153
123523153
123523153

什么是大数据？

生物统计学2024

大数据特征定义

慧聪网
HC360.com

大数据时代要具备大数据思维

维克托·迈尔-舍恩伯格认为：

1-需要全部数据样本而不是抽样;
2-关注效率而不是精确度;
3-关注相关性而不是因果关系。

大数据并不在“大”，而在于“有用”。

价值含量、挖掘成本比数量更为重要。

生物统计学2024

大数据的价值所在？

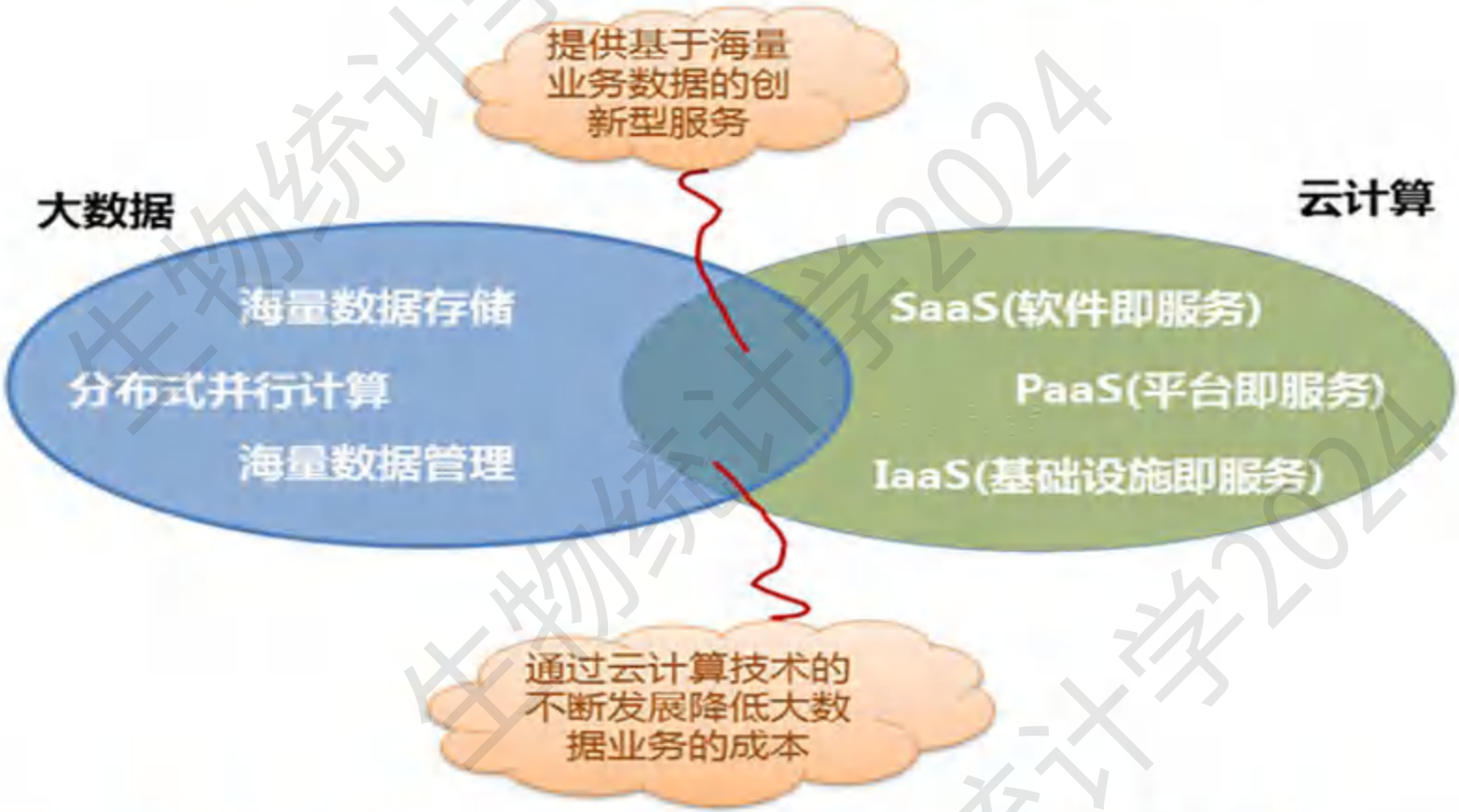
- 如果把大数据比作一种产业，那么这种产业实现盈利的关键，在于提高对数据的“加工能力”，通过“加工”实现数据的“增值”。
- 未来在大数据领域最具有价值的是两种事物：

1-拥有大数据思维的人，这种人可以将大数据的潜在价值转化为实际利益；

2-还未有被大数据触及过的业务领域。这些是还未被挖掘的油井，金矿，是所谓的蓝海。

生物统计学2024

云计算和大数据的关系

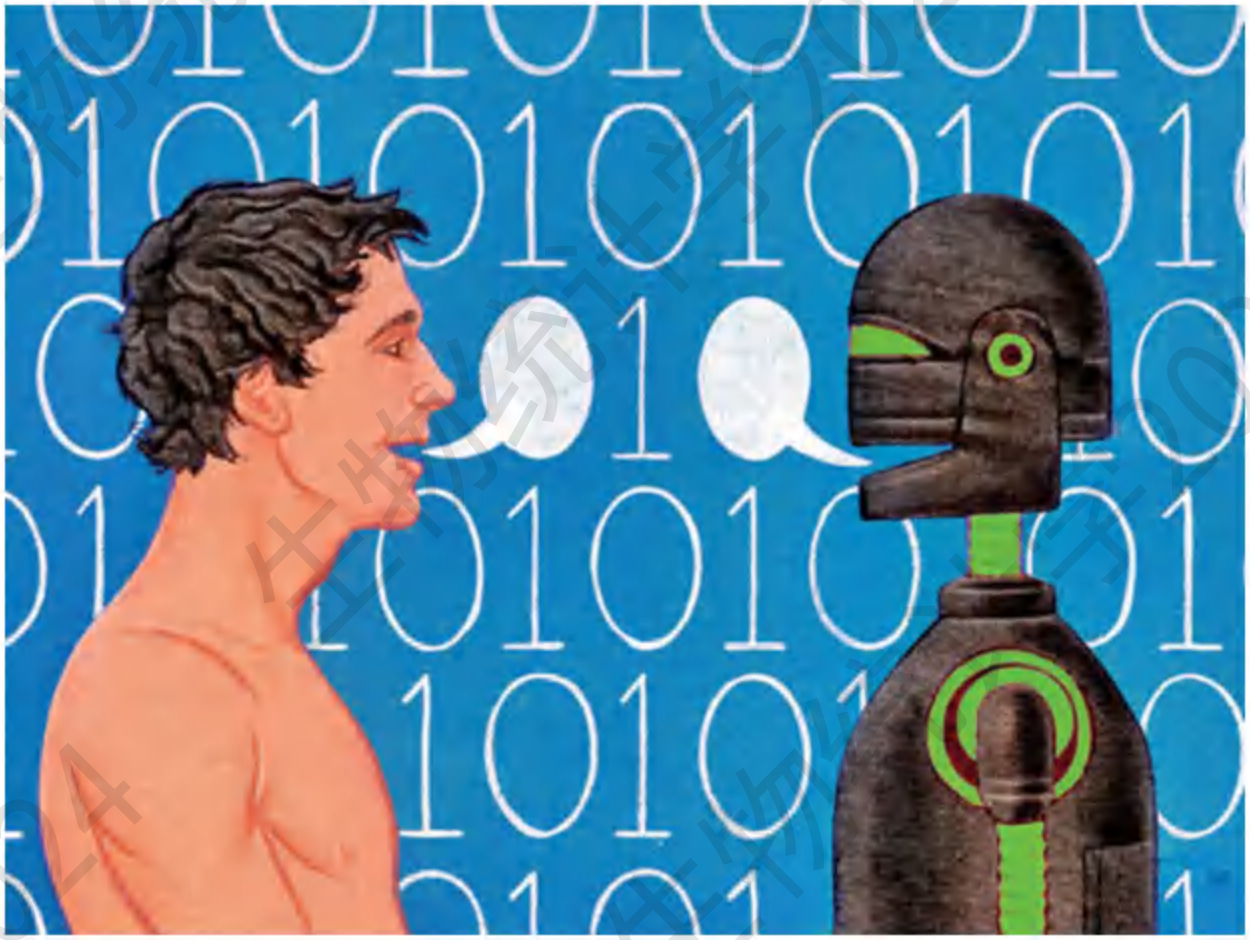


云计算思想：把计算能力作为一种像水和电一样的公用事业提供给用户。

云计算充当了工业革命时期的发动机的角色，而大数据则是电。

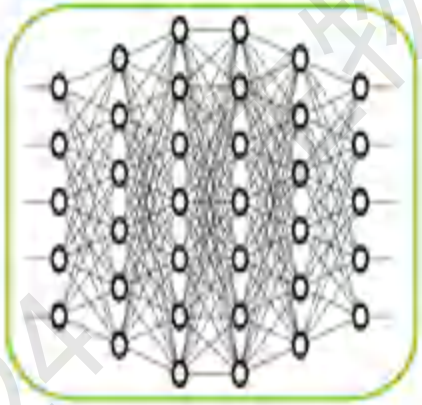
生物统计学2024

机器学习、深度学习和人工智能



生物统计学2024

- 任何通过数据训练的学习算法的相关研究都属于机器学习。比如线性回归（Linear Regression）、K均值（K-means，基于原型的目标函数聚类方法）、决策树（Decision Trees，运用概率分析的一种图解法）、随机森林（Random Forest，运用概率分析的一种图解法）、PCA（Principal Component Analysis，主成分分析）、SVM（Support Vector Machine，支持向量机）以及ANN（Artificial Neural Networks，人工神经网络）。

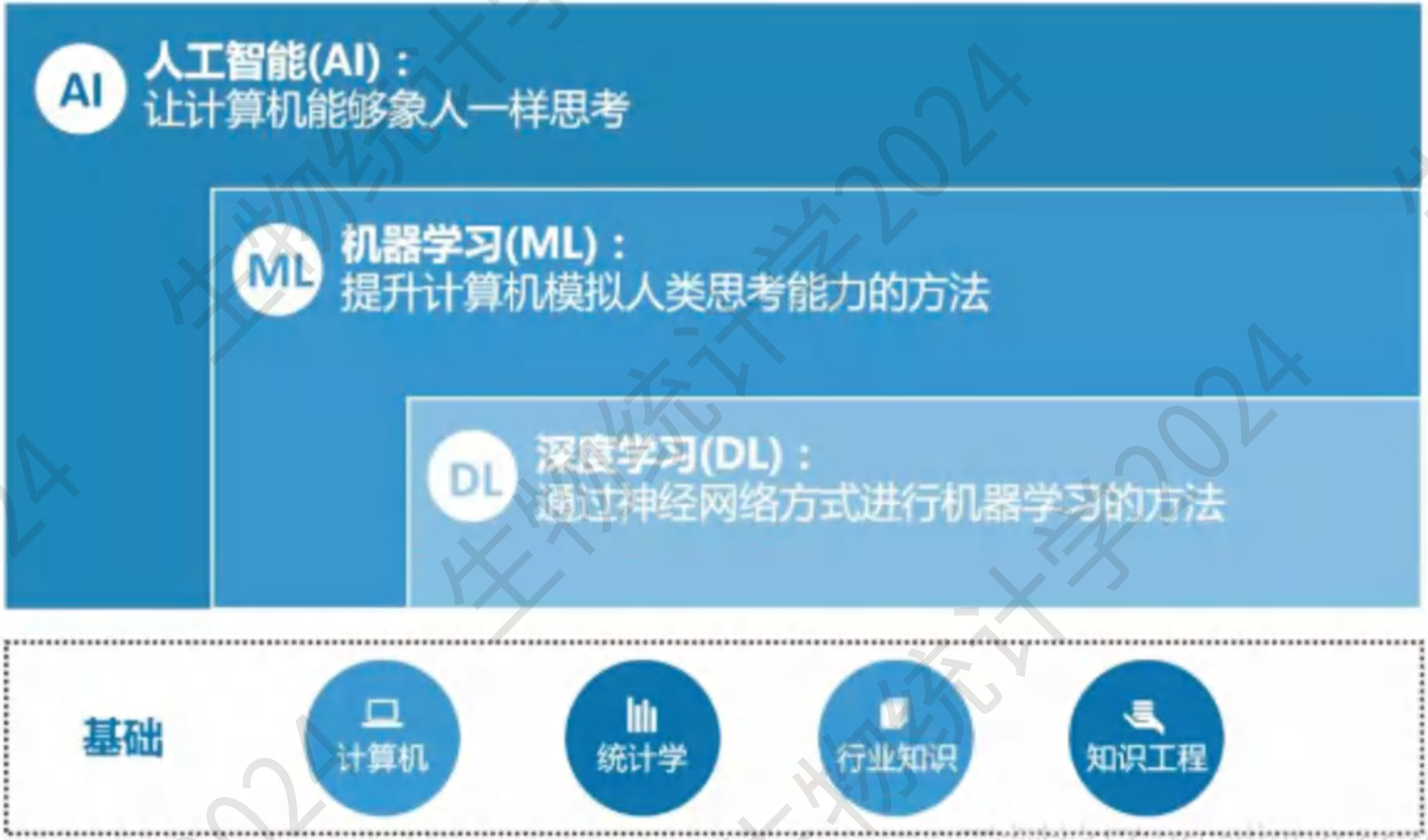


- 深度学习的概念源于人工神经网络的研究，通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。深度学习是机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本。

- 人工智能企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括语音识别、图像识别、机器人、自然语言处理、智能搜索和专家系统等。

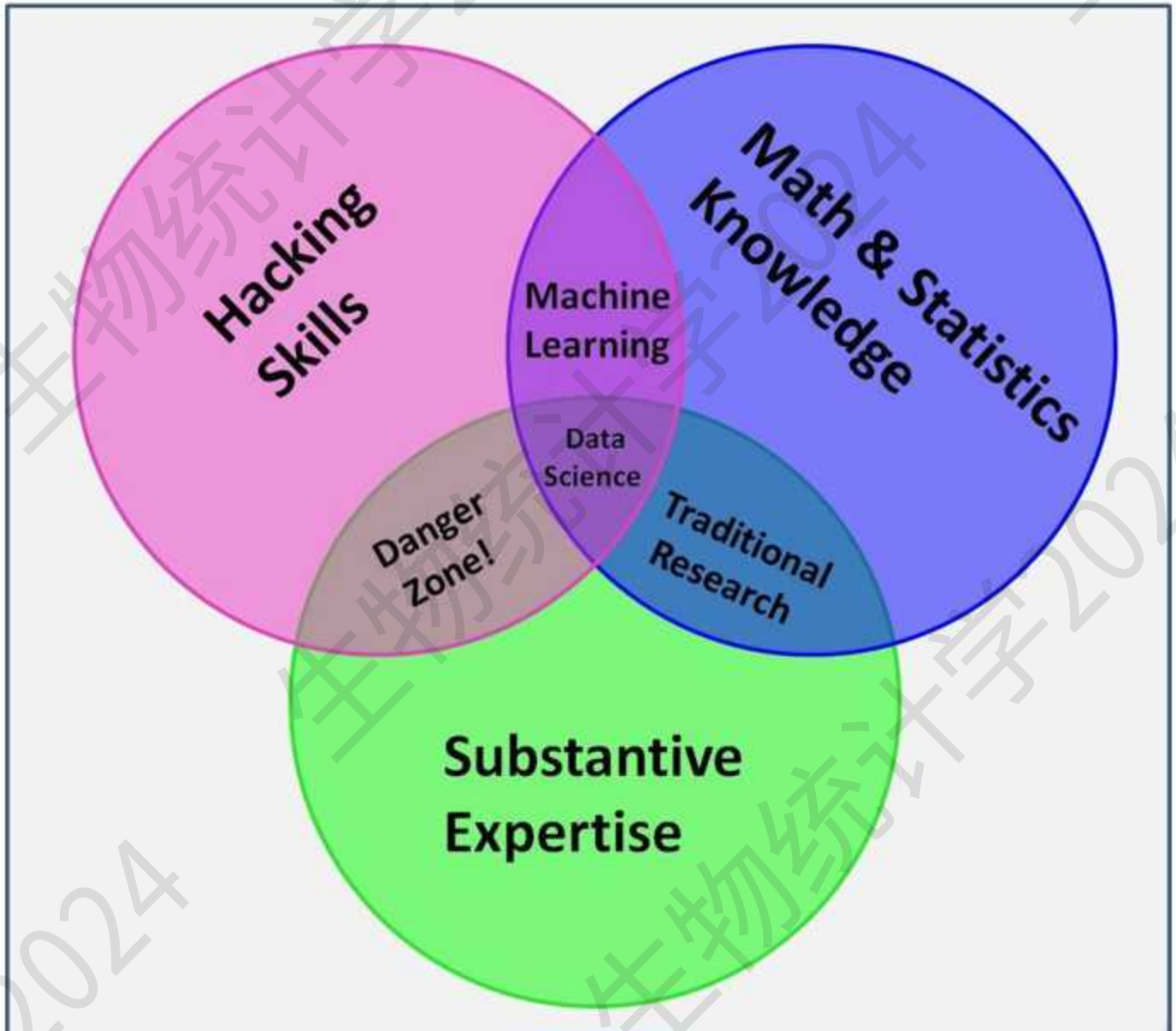
生物统计学2024

人工智能与机器学习、深度学习的关系



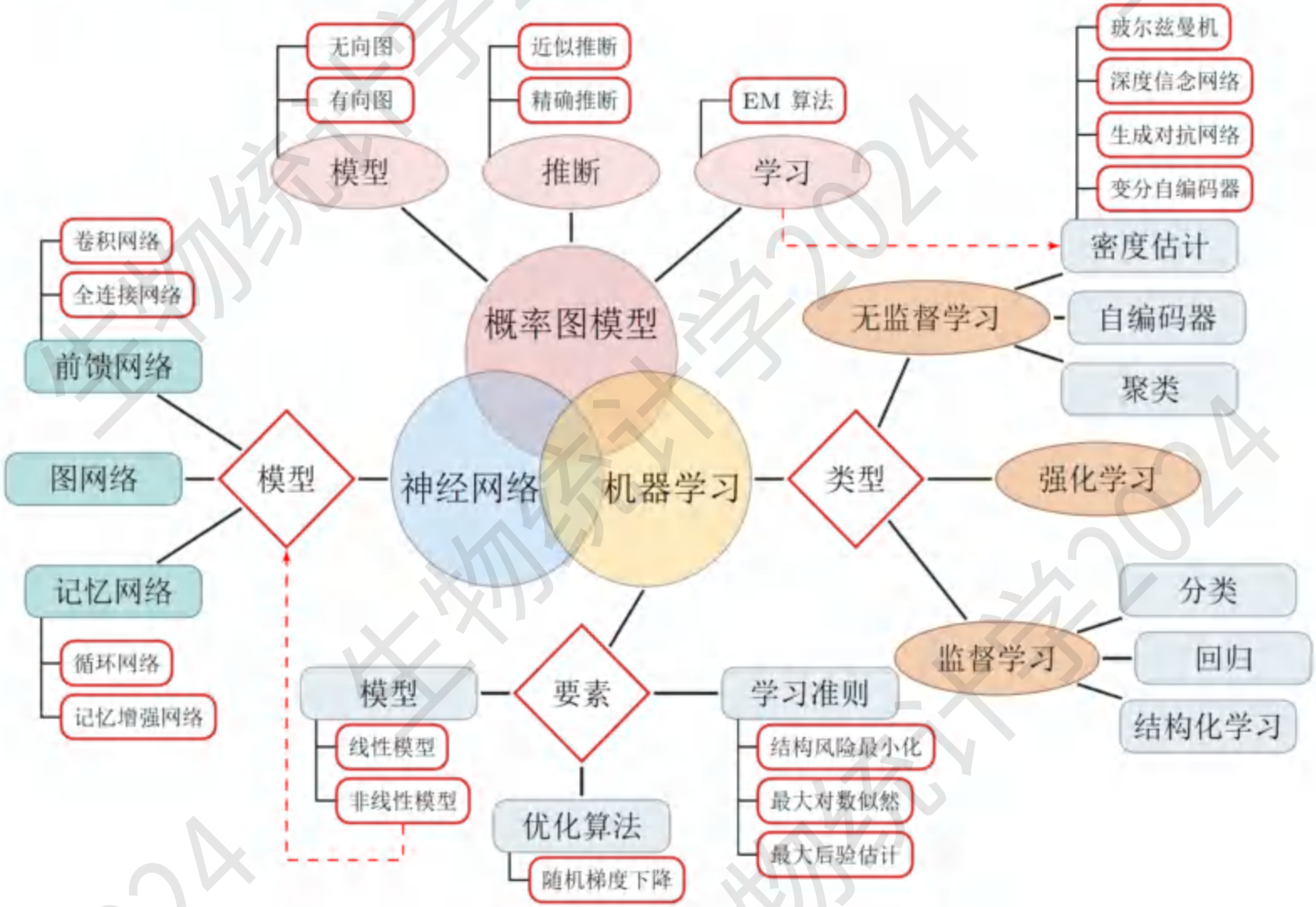
生物统计学2024
云栖社区 yq.aliyun.com

深度学习与数据科学



生物统计学2024

深度学习与数据科学



生物统计学2024

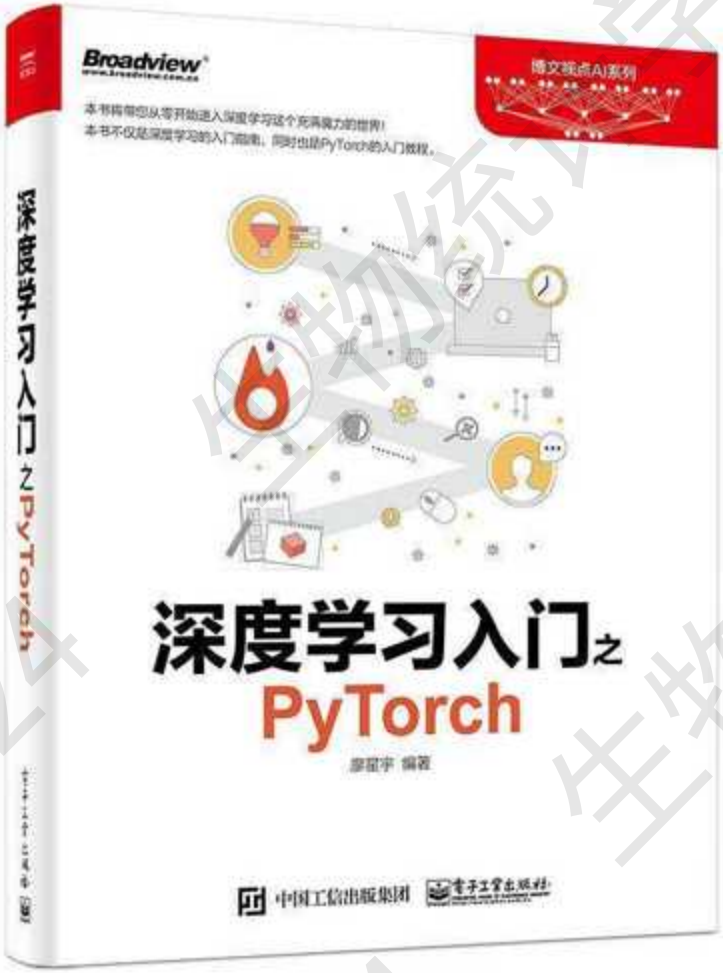
深度学习与数据科学

- 深度学习：一种基于无监督特征学习和特征层次结构的学习方法

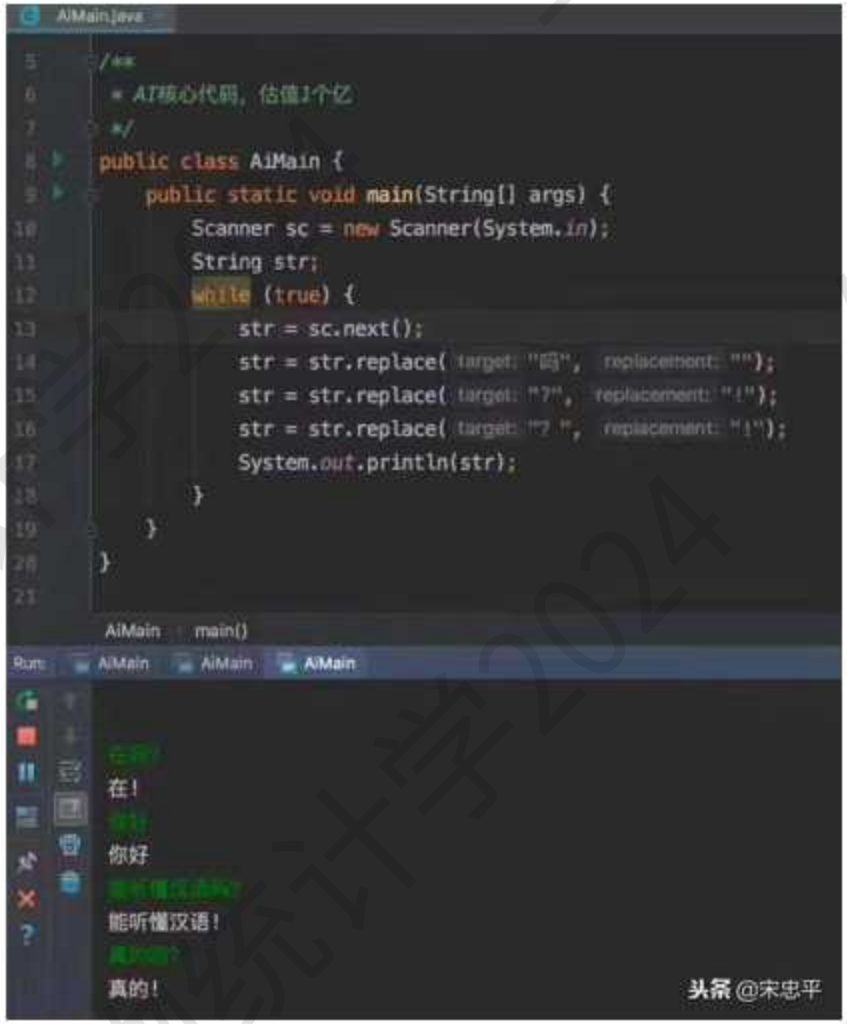
- 可能的的名称：
  - 深度学习
  - 特征学习
  - 无监督特征学习

生物统计学2024

真假深度学习



真

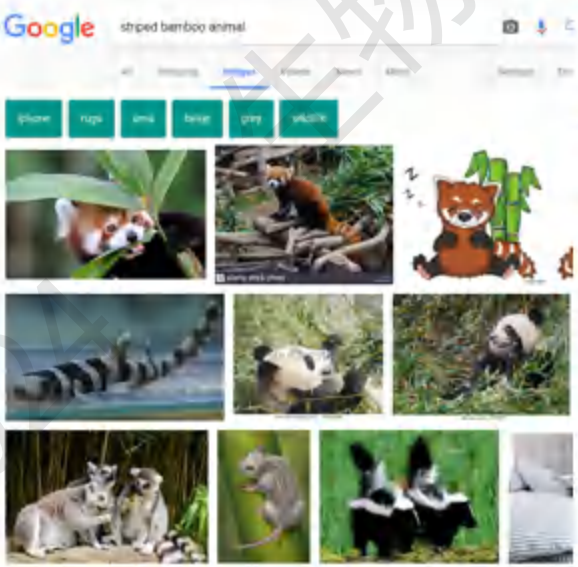


假

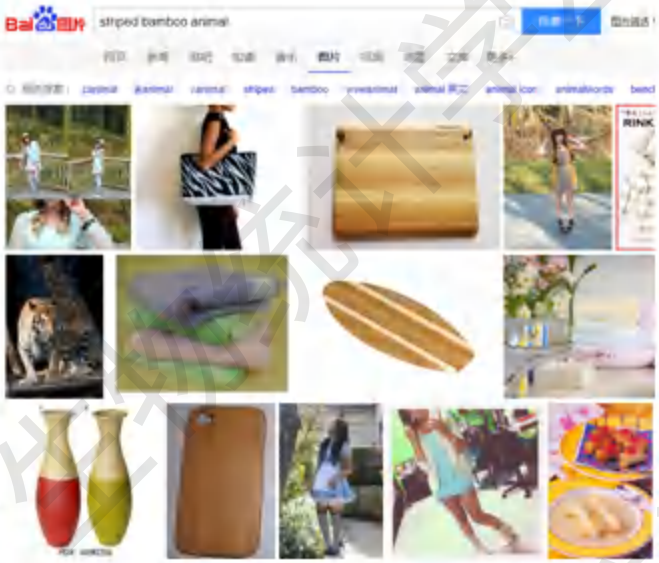
生物统计学2024

搜索引擎的区别

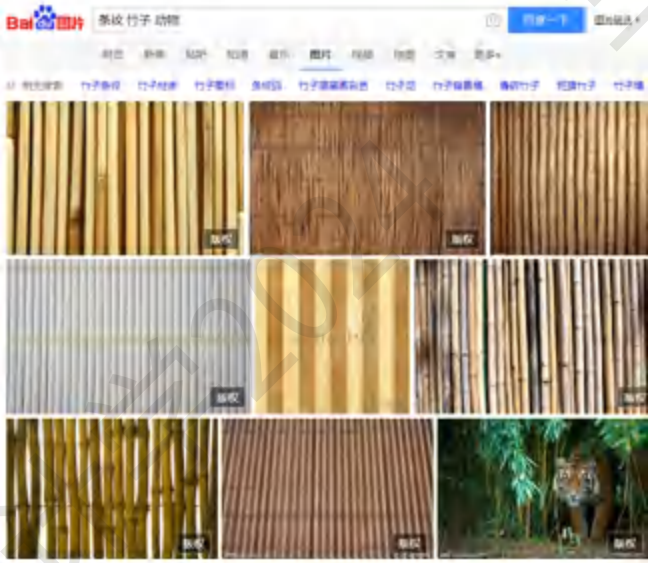
striped bamboo animal



striped bamboo animal

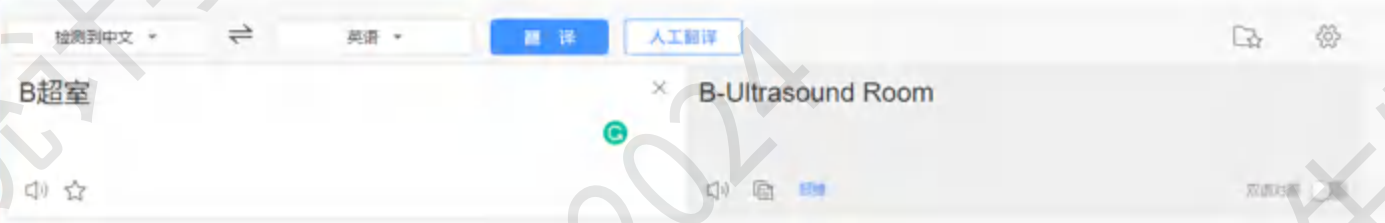


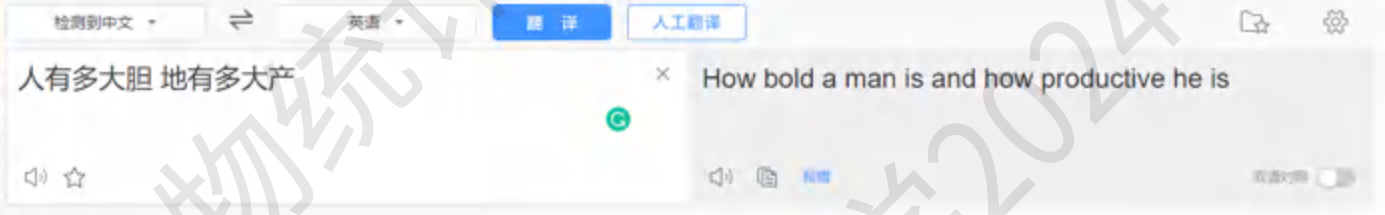
条纹 竹子 动物

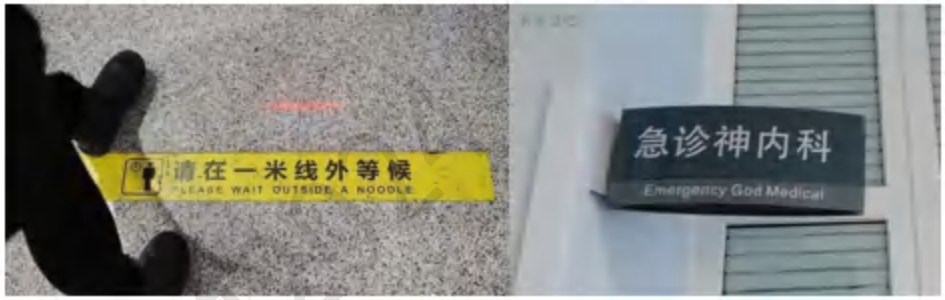


生物统计学2024

翻译引擎和人类的较量







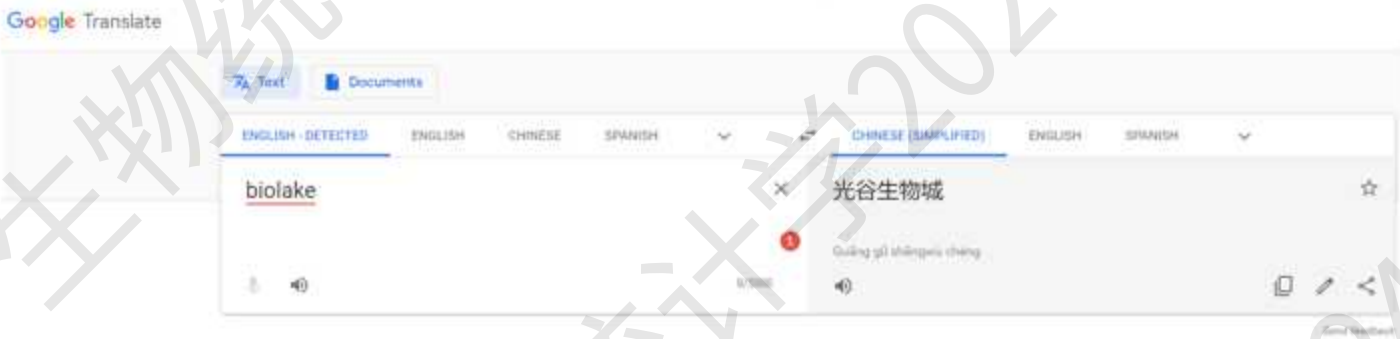
生物统计学2024

翻译引擎和人类的较量

Simple
Florida Georgia Line
We've been there, it's safe to say it ain't our style
我们也曾有过这种处境，但我们可以肯定地说那不是我们的风格
It's just that simple, S-I-M-P-L-E
就是如此简单，鸡一安简，的安单
Simple as can be
尽可能的简单
It's just that simple, S-I-M-P-L-E
就是如此简单，鸡一安简，的安单
Simple as can be
尽可能的简单
不为无益之事 何以遣有涯之生
What's life without whimsy?
撒子
what
馄饨
chaos
灭 烟 台
Destroy Yantai

翻译引擎的区别

biolake



biolake

光谷生物城

生物统计学2024

翻译的区别（互有优劣）

它很难吃。

谷歌：It is hard to eat.

百度：It tastes terrible.

我要下班。

谷歌：I want to get off work from work.

百度：I'm going to get off work。

生物统计学2024

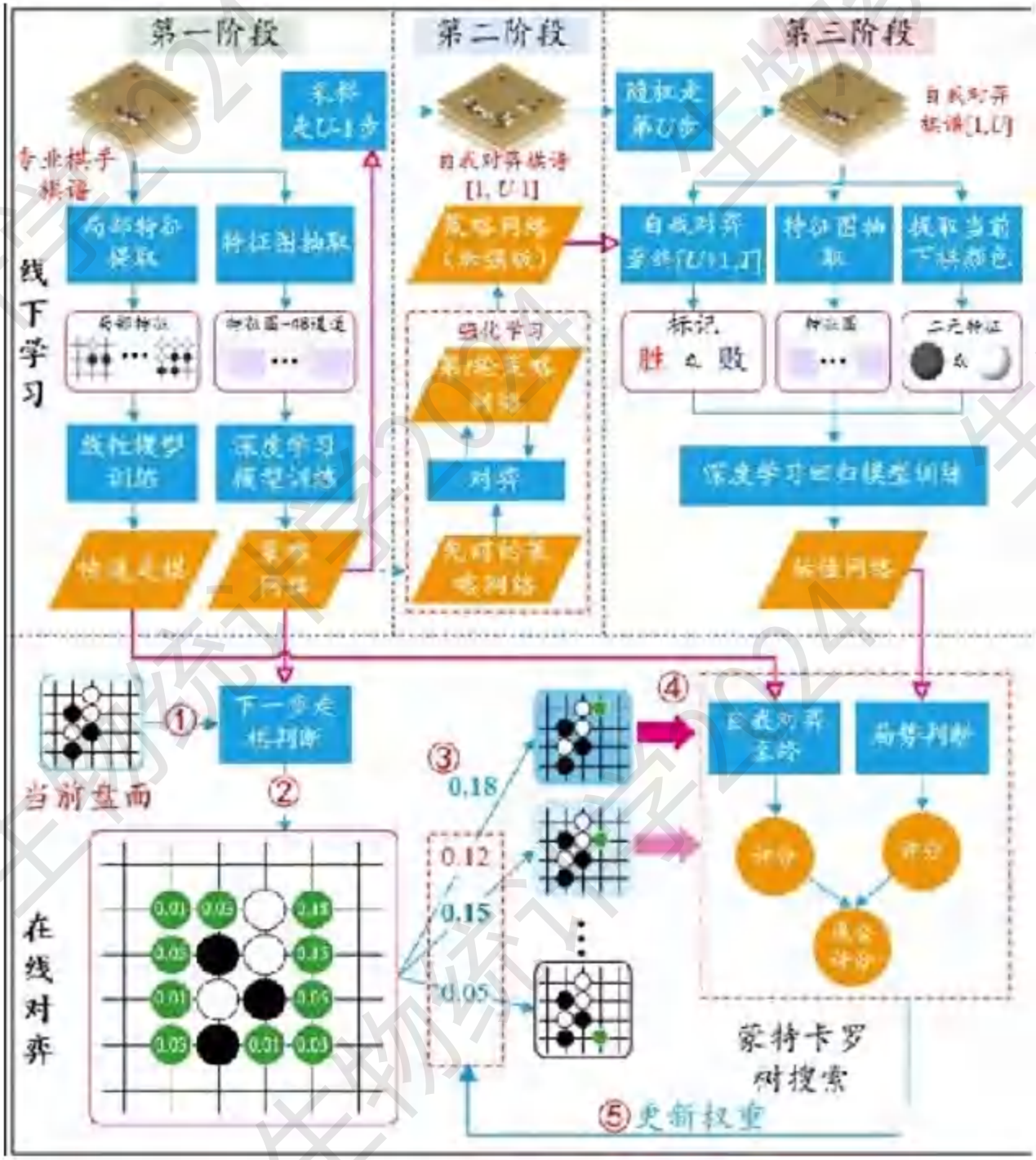
AlphaGo

生物统计学2024

AlphaGo

ARTICLE

Mastering the game of Go with deep neural networks and tree search



AlphaGo vs. AlphaGo Zero

数据很重要！

算法更重要！

Google DeepMind

生物统计学2024

AlphaZero

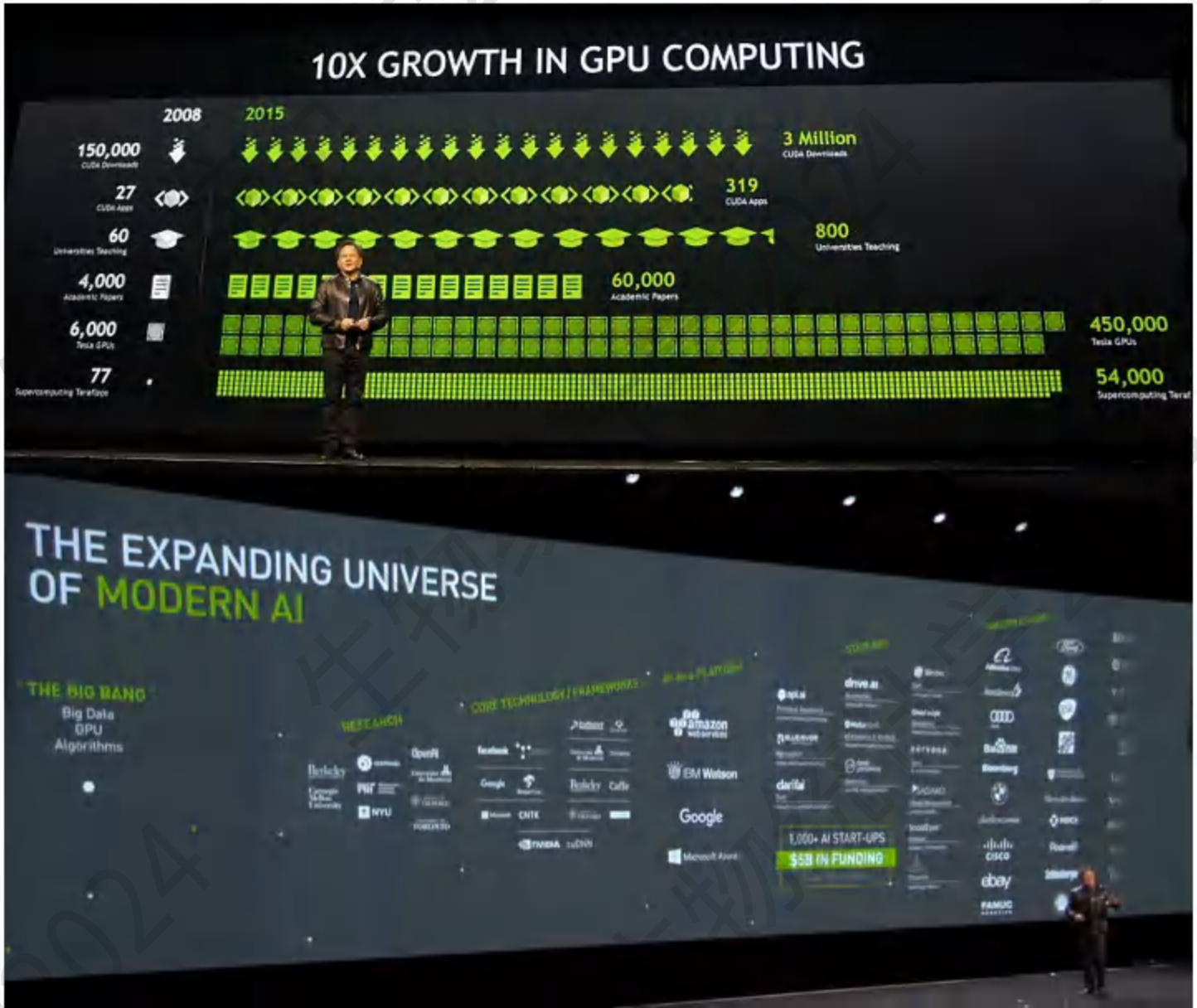
Reinforcement Learning (强化学习)

生物统计学2024

Google DeepMind

Transfer Learning (迁移学习)

Hardware: GPU, HPC...



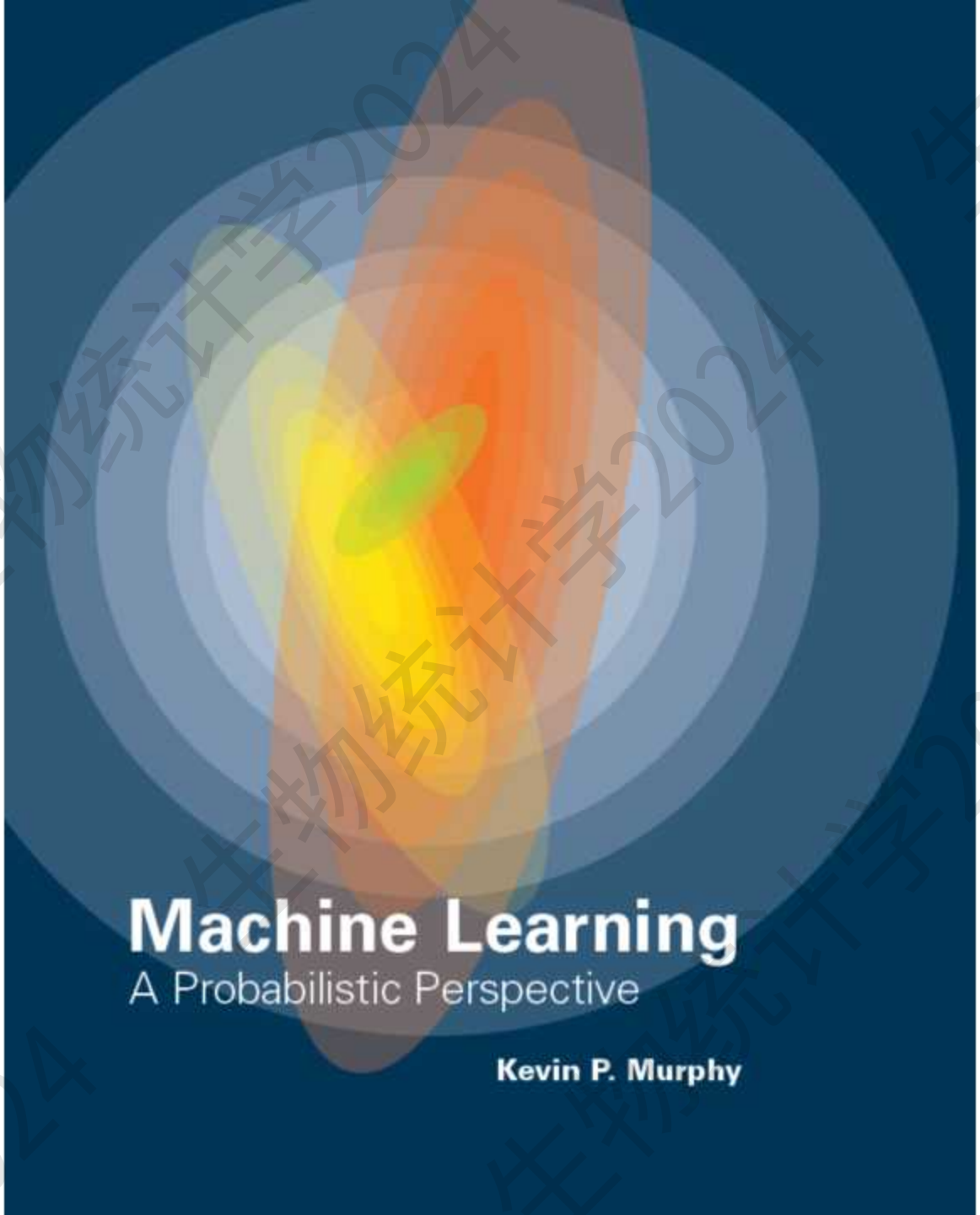
Software: TensorFlow, Caffe...

生物统计学2024

Turing Award, 2019

Yann LeCun
Geoffrey Hinton
Yoshua Bengio
Andrew Ng

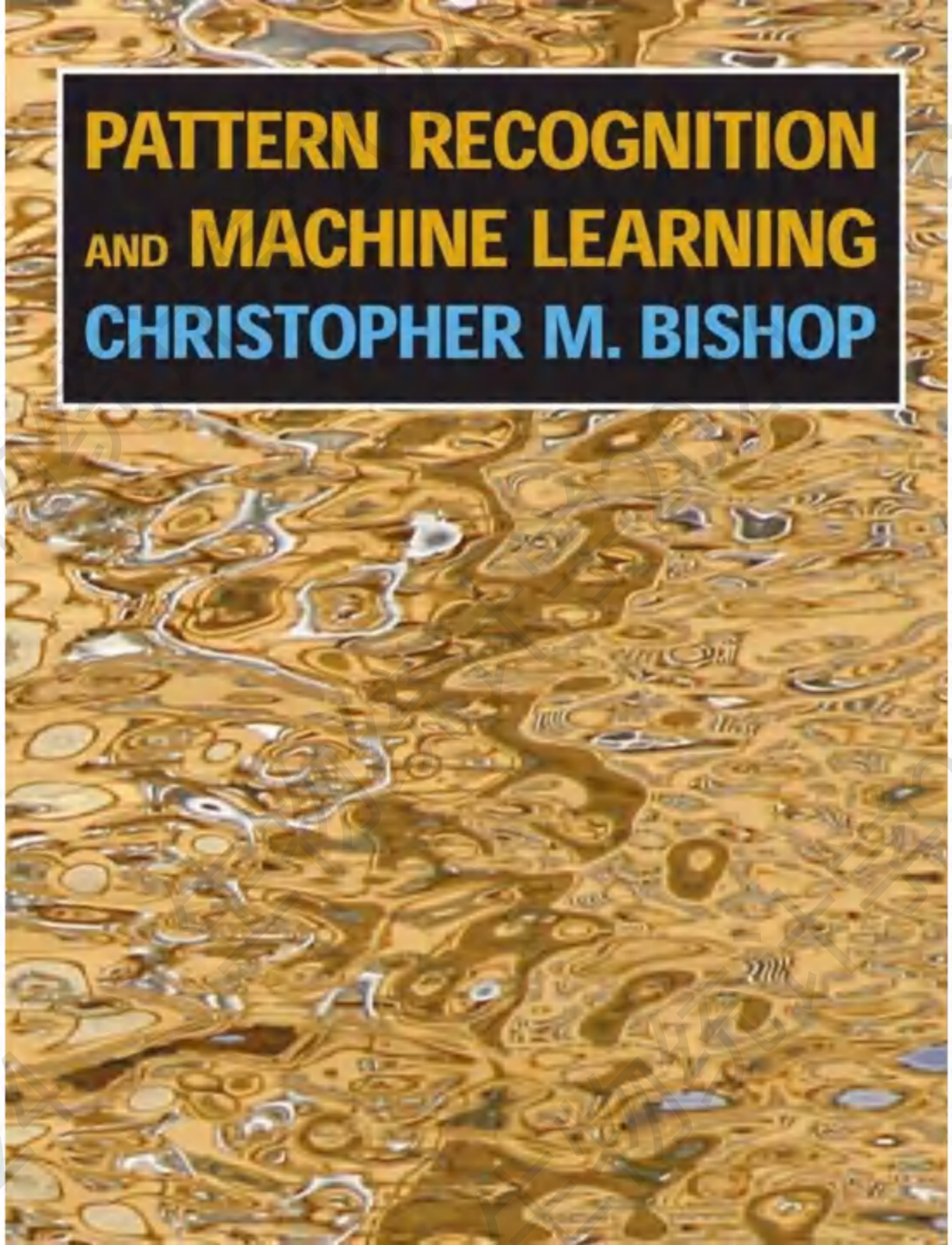
生物计算2024


Machine Learning
A Probabilistic Perspective
Kevin P. Murphy

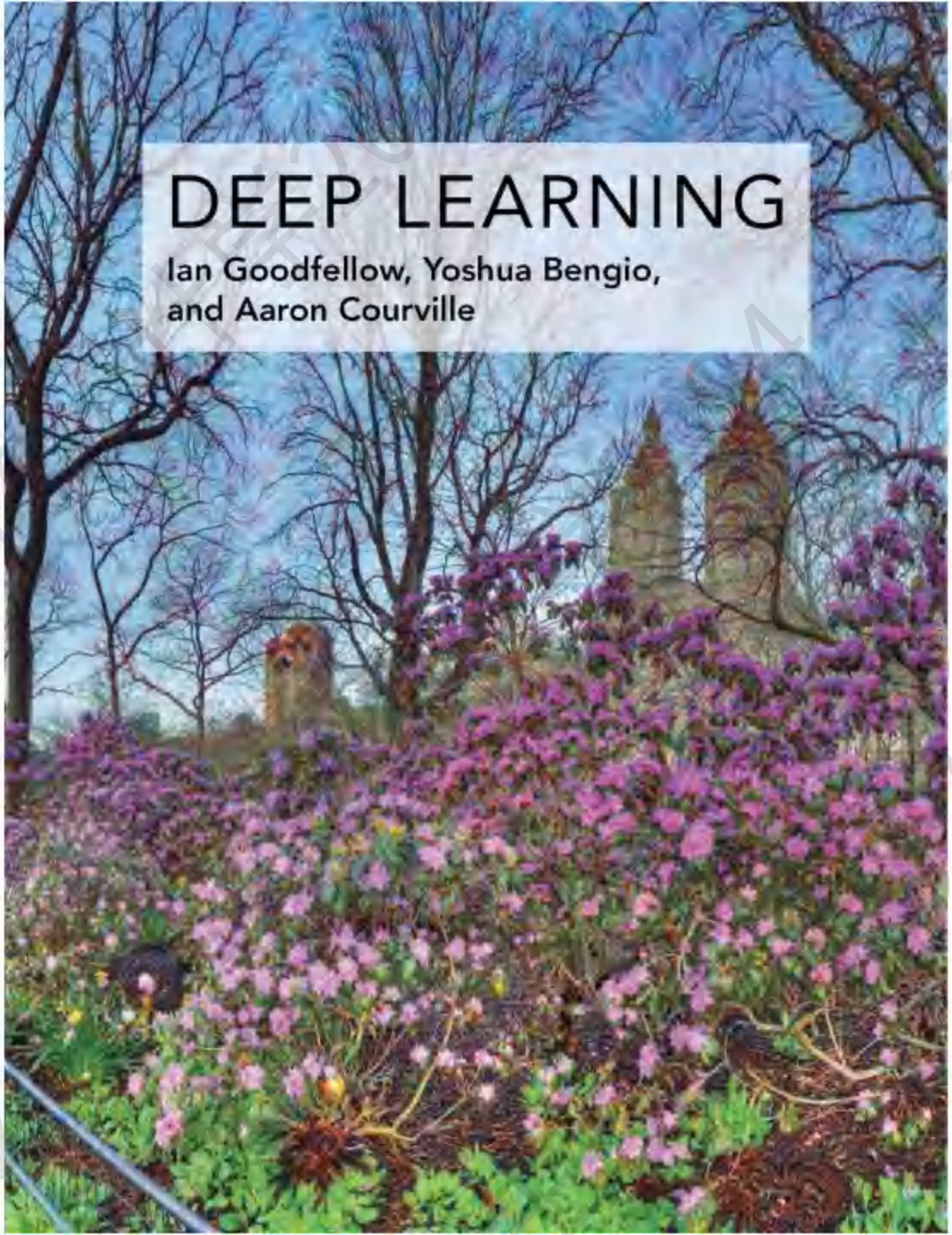
生物统计学2024

PATTERN RECOGNITION AND MACHINE LEARNING

CHRISTOPHER M. BISHOP

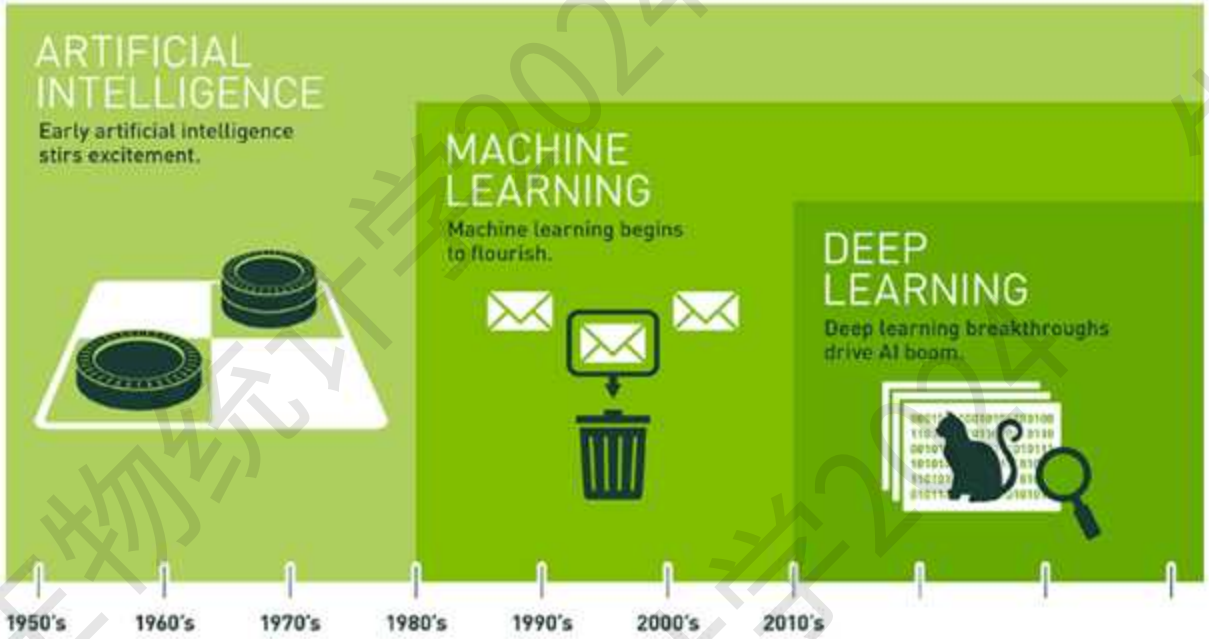


生物医学工程2024


DEEP LEARNING
Ian Goodfellow, Yoshua Bengio,
and Aaron Courville

Introduction

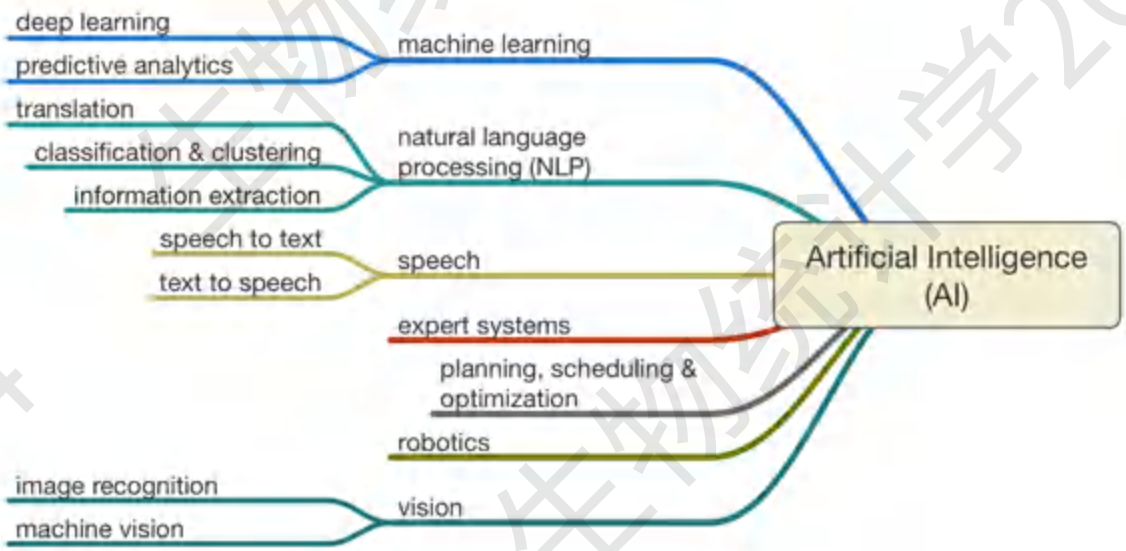
- Artificial Intelligence
- Machine learning
- “Deep” learning



生物统计学2024

Artificial Intelligence

Artificial intelligence (AI, also machine intelligence, MI) is intelligence displayed by machines, in contrast with the natural intelligence (NI) displayed by humans and other animals.

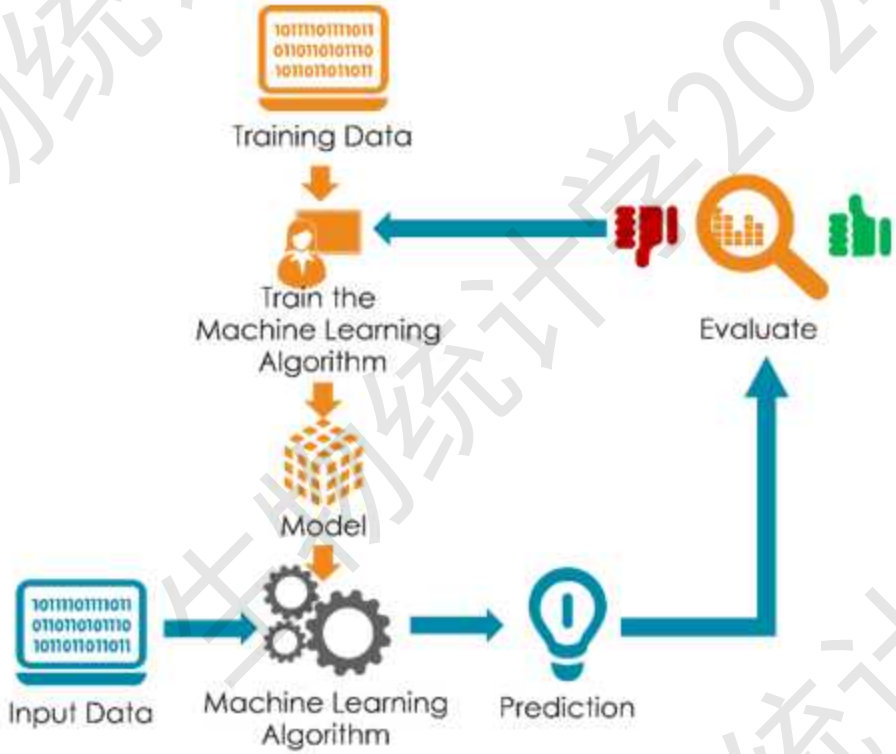


Hatley, L. (2016). Presentation, New Designs for Learning: Games and Gamification
https://en.wikipedia.org/wiki/Applications_of_artificial_intelligence

生物统计学2024

Machine Learning

Machine learning is the science of getting computers to act without being explicitly programmed.

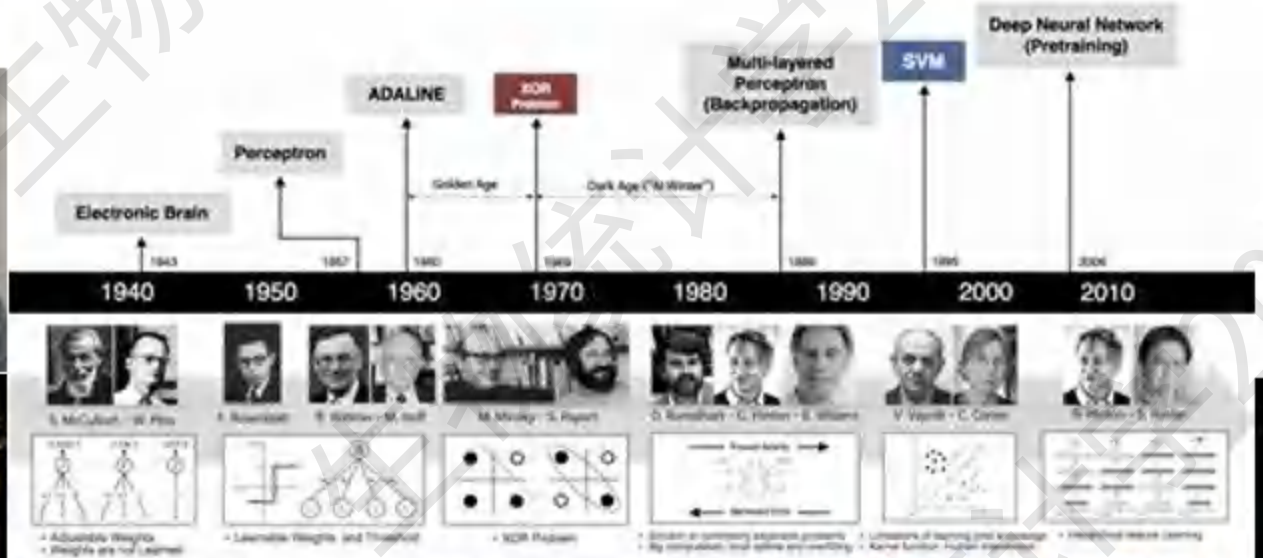


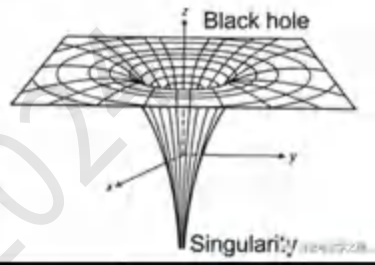
General workflow of Machine Learning

Pat Alvarado et al,2016,Building the Machine Learning Infrastructure

生物统计学2024

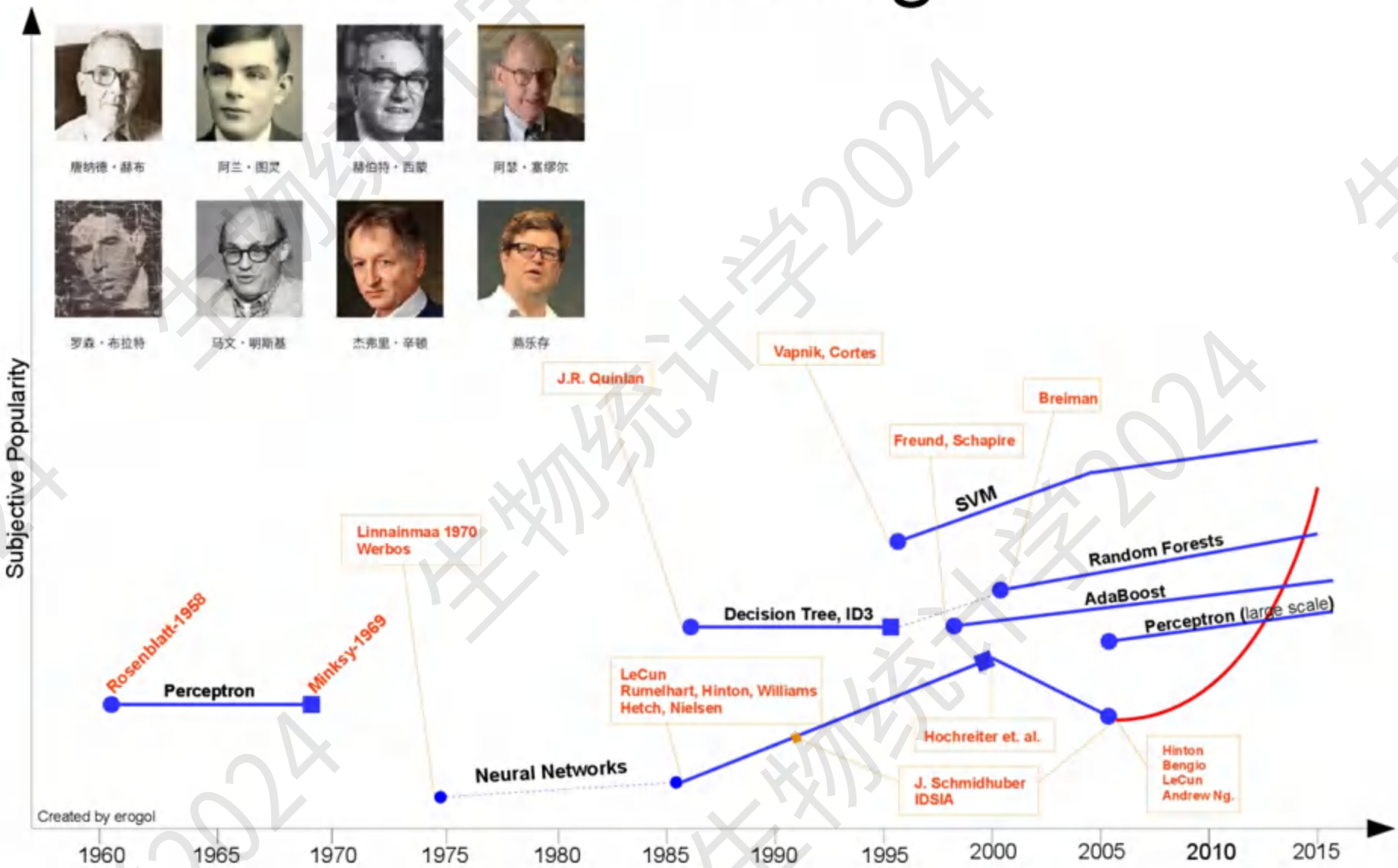
Machine Learning





生物统计学2024

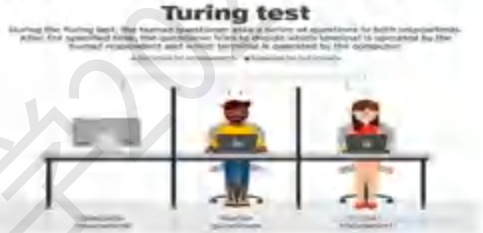
Machine Learning



生物统计学2024

Machine Learning

- 图灵
  - 1912年6月23日-1954年6月7日，英国著名数学家、逻辑学家、密码学家，被誉为“计算机科学之父”、“人工智能之父”，是计算机逻辑的奠基者



- 理查德·费曼
  - 1918年5月11日-1988年2月15日，1965年诺贝尔物理学奖得主

- 冯·诺依曼
  - 1903年12月28日-1957年2月8日，数学家、计算机科学家、物理学家，是20世纪最重要的数学家之一，也是现代计算机、博弈论、核武器和生化武器等领域内的科学全才之一，被后人称为“现代计算机之父”、“博弈论之父”

生物统计学2024

机器学习 ≈ 构建一个映射函数

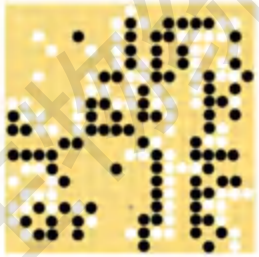
- 语音识别

$f($ $)=$ “你好”

- 图像识别

$f($ $)=$ “9”

- 围棋

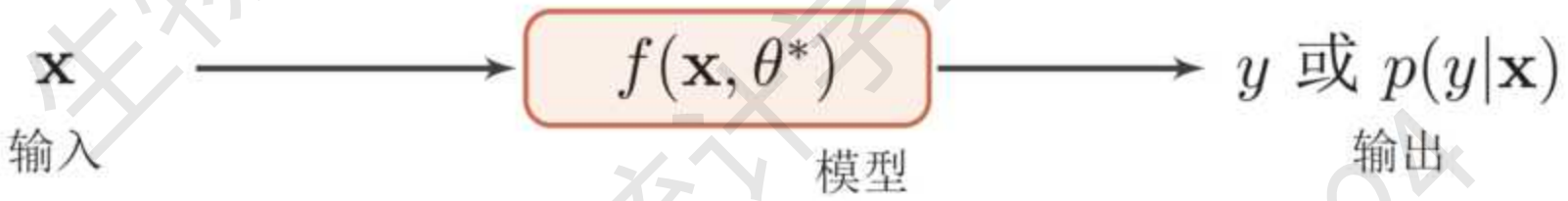
$f($  $)=$ “6-5” (落子位置)

- 机器翻译

$f($ “你好！” $)=$ “Hello!”

生物统计学2024

机器学习概览



生物统计学2024

机器学习的三要素

- 模型
  - 线性方法 $f(\mathbf{x},\theta)=\mathbf{w}^{\mathrm{T}}\mathbf{x}+b$
  - 广义线性方法 $f(\mathbf{x},\theta)=\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x})+b$
    - 如果$\phi(x)$为可学习的非线性基函数，$f(x,\theta)$就等价于神经网络。

- 学习准则
  - 期望风险

$$\mathcal{R}(f)=\mathbb{E}_{(\mathbf{x},y)\sim p(\mathbf{x},y)}[\mathcal{L}(f(\mathbf{x}),y)],$$

- 优化
  - 梯度下降

生物统计学2024

常见的机器学习类型

	监督学习	无监督学习	强化学习
训练样本	训练集 $\{(\mathbf{x}^{(n)}, y^{(n)})\}_{n=1}^{N}$	训练集 $\{\mathbf{x}^{n}\}_{n=1}^{N}$	智能体和环境交互的轨迹 τ 和累积奖励 G_τ
优化目标	$y = f(\mathbf{x})$ 或 $p(y\|\mathbf{x})$	$p(\mathbf{x})$ 或带隐变量 $\mathbf{z}$ 的 $p(\mathbf{x}\|\mathbf{z})$	期望总回报 $\mathbb{E}_\tau[G_\tau]$
学习准则	期望风险最小化 最大似然估计	最大似然估计 最小重构错误	策略评估 策略改进

生物统计学2024

参数学习

- 期望风险未知，通过经验风险近似
  - 训练数据：$\mathcal{D} = \{\mathrm{x}^{(\mathrm{n})}, y^{(\mathrm{n})}\}$, $i \in [1, N]$

$$\mathcal{R}_{\mathcal{D}}^{emp}(\theta) = \frac{1}{N}\sum_{n=1}^{N}\mathcal{L}(y^{(n)}, f(x^{(n)}, \theta))$$

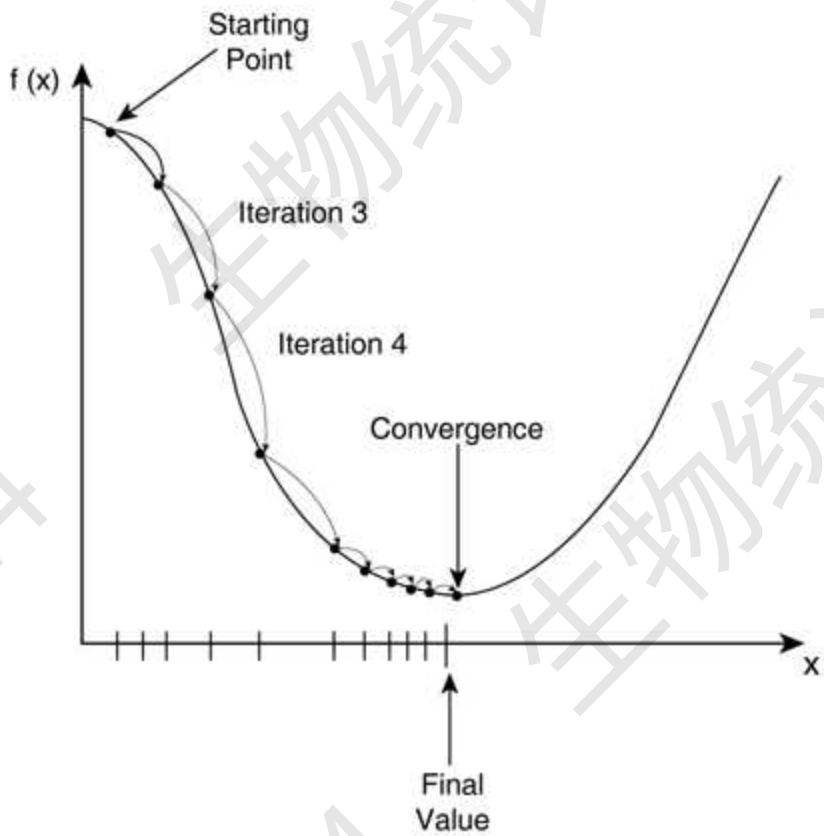
- 经验风险最小化
  - 在选择合适的风险函数后，我们寻找一个参数θ*，使得经验风险函数最小化。

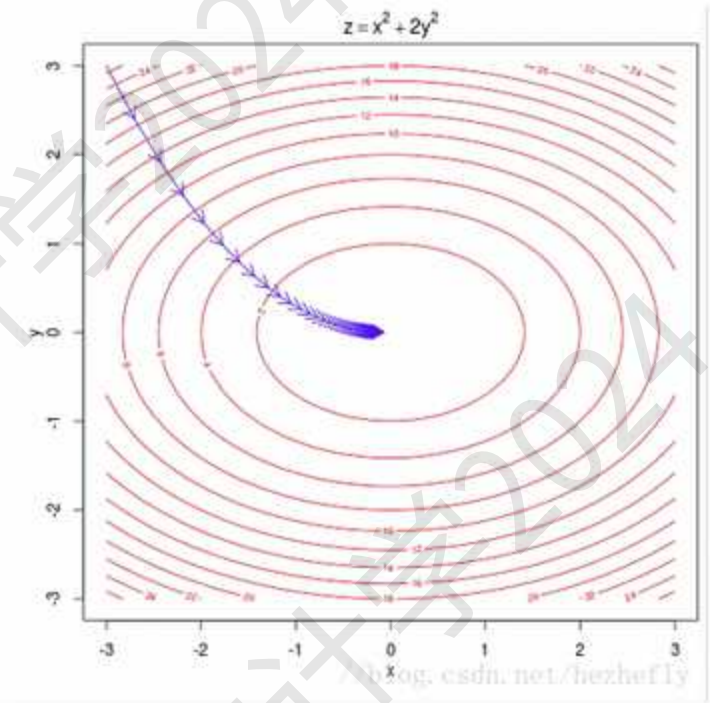
$$\theta^* = \arg\min_{\theta} \mathcal{R}_{\mathcal{D}}^{emp}(\theta)$$

- 机器学习问题转化成为一个最优化问题

生物统计学2024

优化：梯度下降法

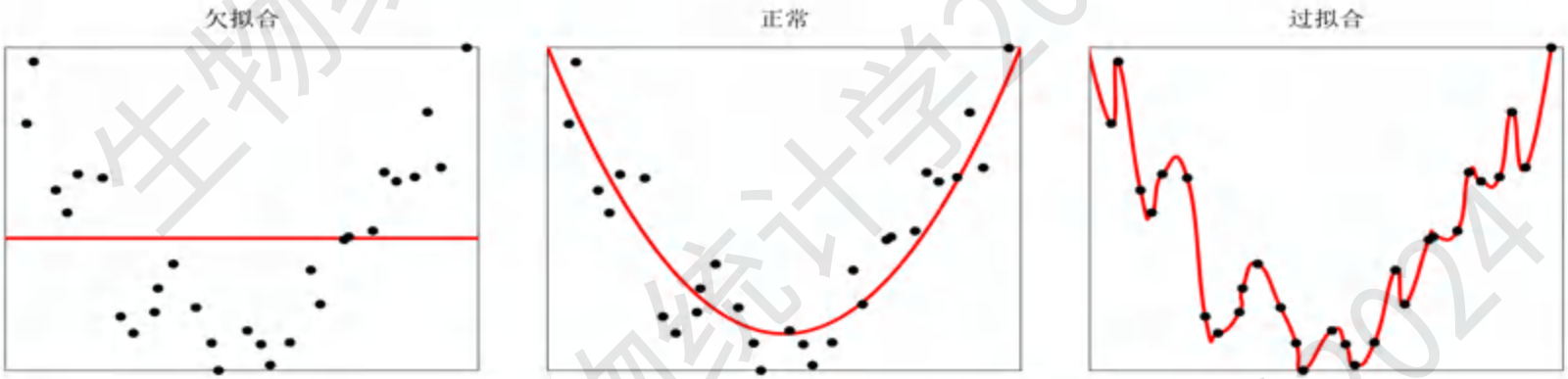




生物统计学2024

机器学习 = 优化？

机器学习 = 优化？ NO！



过拟合：经验风险最小化原则很容易导致模型在训练集上错误率很低，但是在未知数据上错误率很高。

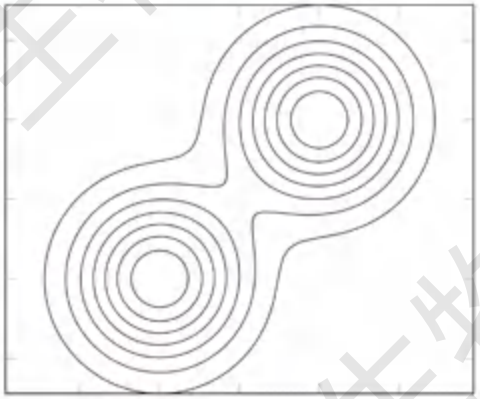
生物统计学2024

泛化错误

期望风险

$$\mathcal{R}(f) = \mathbb{E}_{(\mathbf{x},y)\sim p(\mathbf{x},y)}[\mathcal{L}(f(\mathbf{x}), y)],$$

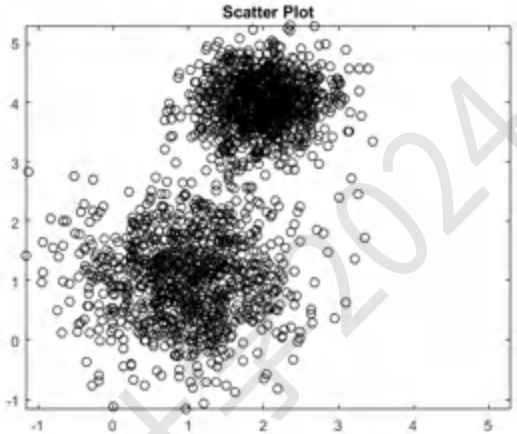
真实分布 p_r



$\neq$

经验风险

$$\mathcal{R}_{\mathcal{D}}^{emp}(\theta) = \frac{1}{N}\sum_{n=1}^{N}\mathcal{L}(y^{(n)}, f(x^{(n)}, \theta))$$



$$\mathcal{G}_{\mathcal{D}}(f) = \mathcal{R}(f) - \mathcal{R}_{\mathcal{D}}^{emp}(f)$$

泛化错误

生物统计学2024

如何减少泛化错误？

优化
经验风险最小

正则化
降低模型复杂度

生物统计学2024

正则化（regularization）

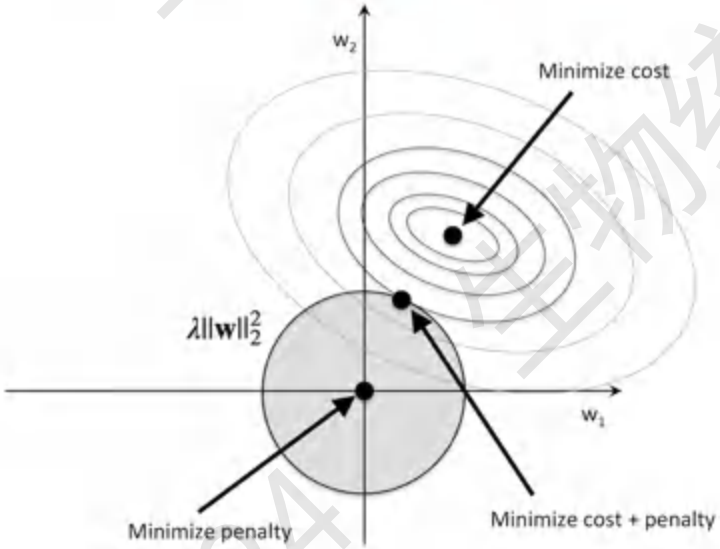
所有损害优化的方法都是正则化。

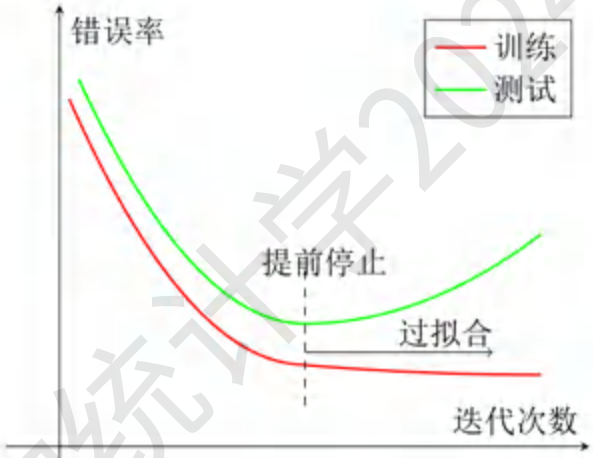
增加优化约束

L1/L2约束、数据增强

干扰优化过程

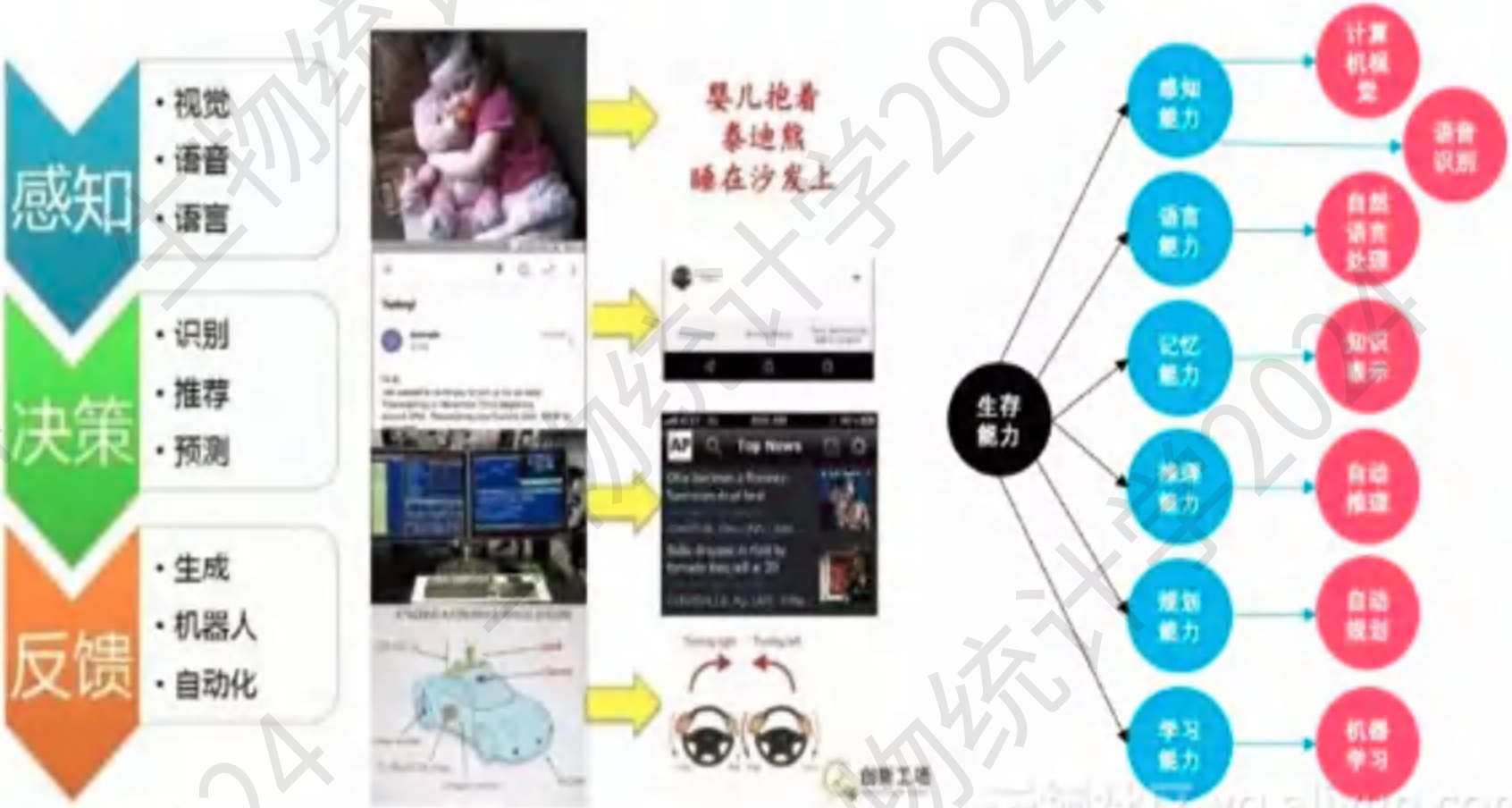
权重衰减、随机梯度下降、提前停止





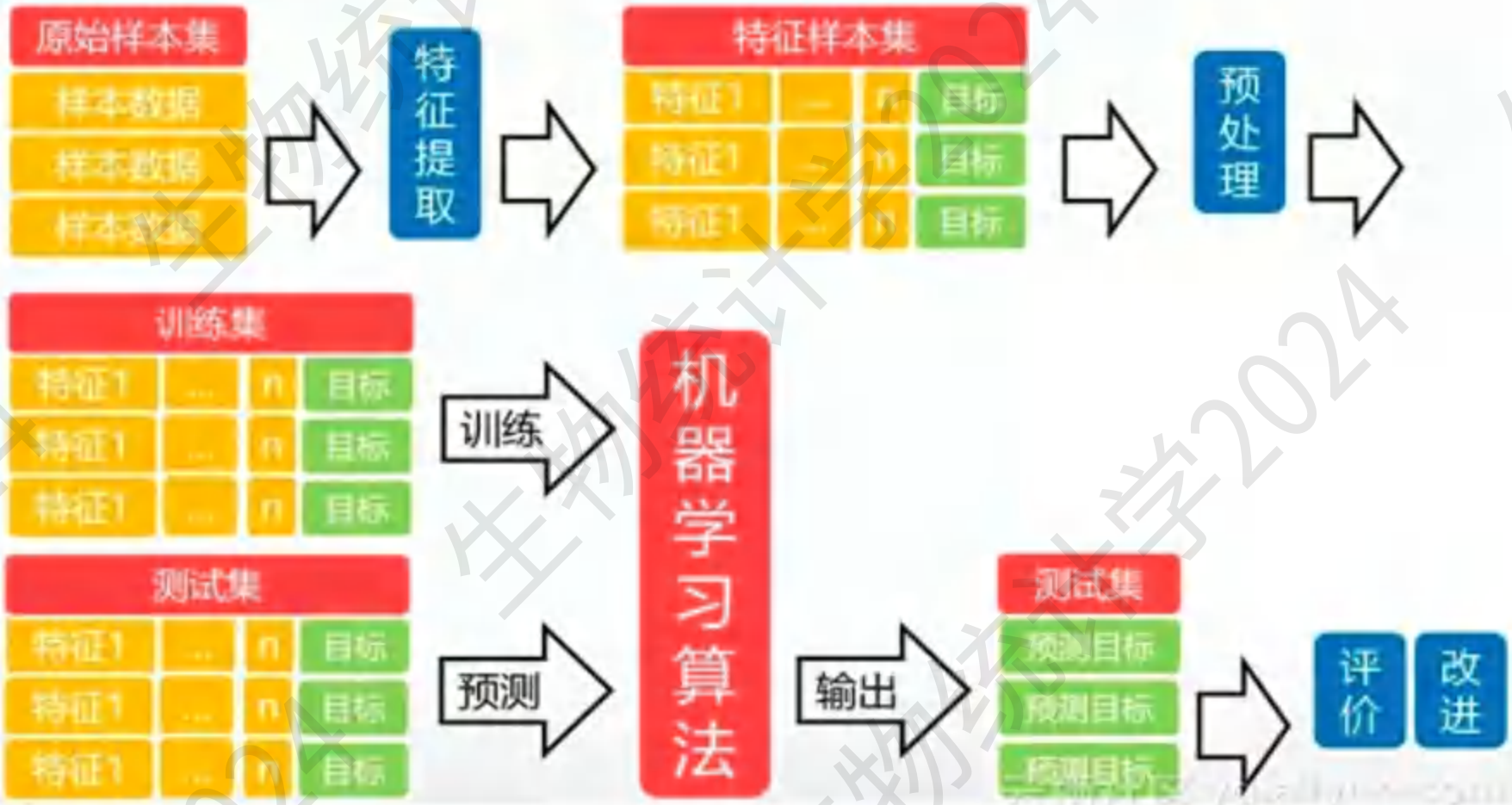
生物统计学2024

拥抱人工智能从机器学习开始



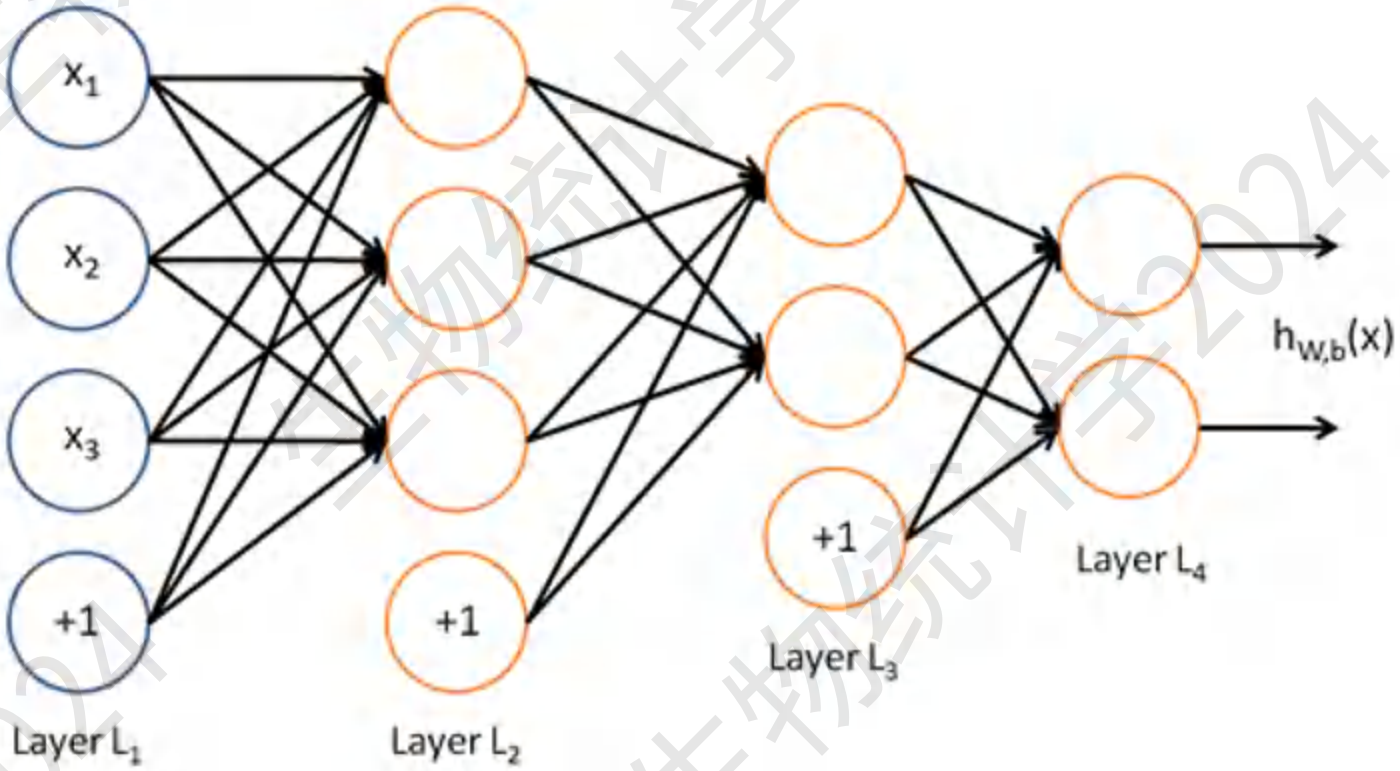
生物统计学2024
云栖社区 yq.aliyun.com

- 机器学习最大的特点是利用数据而不是指令来进行各种工作，其学习过程主要包括：数据的特征提取、数据预处理、训练模型、测试模型、模型评估改进等几部分。



机器学习算法是使计算机具有智能的关键

- 算法是通过使用已知的输入和输出以某种方式“训练”以对特定输入进行响应。代表着用系统的方法描述解决问题的策略机制。人工智能的发展离不开机器学习算法的不断进步。



生物统计学2024

机器学习算法分类

云栖社区 yq.aliyun.com

机器学习算法分类

有监督学习

训练数据集

数据					
标签	车	飞机	美女	帅哥	哥，但不帅

训练

机器学会如何将图片分类

生物统计学2024

机器学习算法分类

无监督学习

数据集（无需训练）

《母猪的产后护理》

《滕王阁序》

《惊！名校博士网恋被骗，对象竟为 AI！》

《生成性 AI 的前世今生》

《建筑学概论》（高等院校建筑学科精品教材）

《了不起的比尔盖茨》

《了不起的盖茨比》

《建筑学概论》（电影）

直接分类

生成的结果

文档类别 1（骈文）

文档类别 2（专业书籍）

文档类别 3（人物传记）

文档类别 4（劲爆新闻）

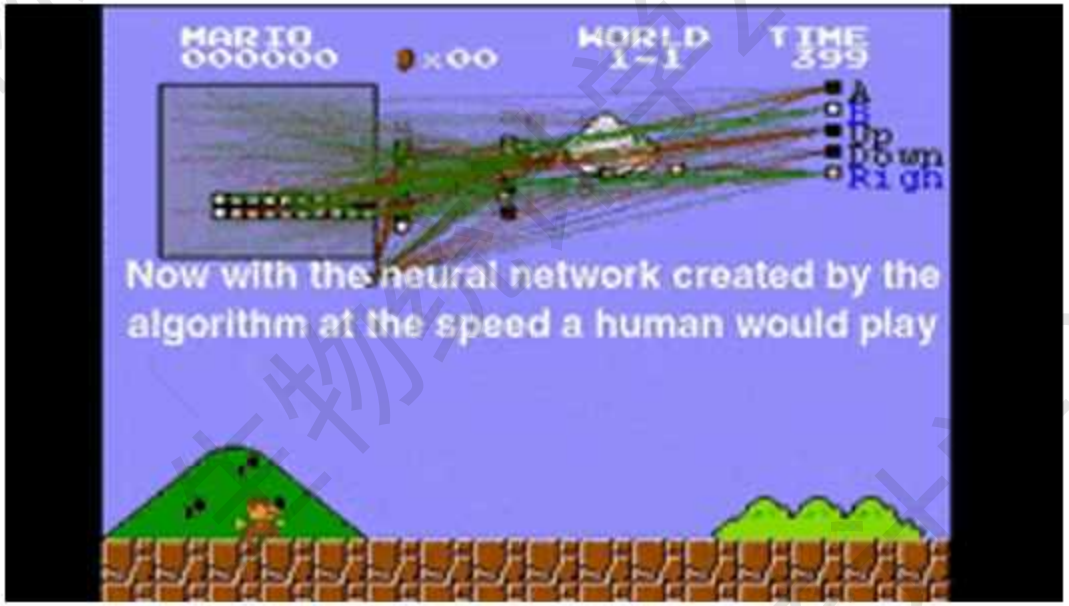
文档类别 5（爱情文学）

文档类别 6（科普）

生物统计学2024

机器学习算法分类

强化学习



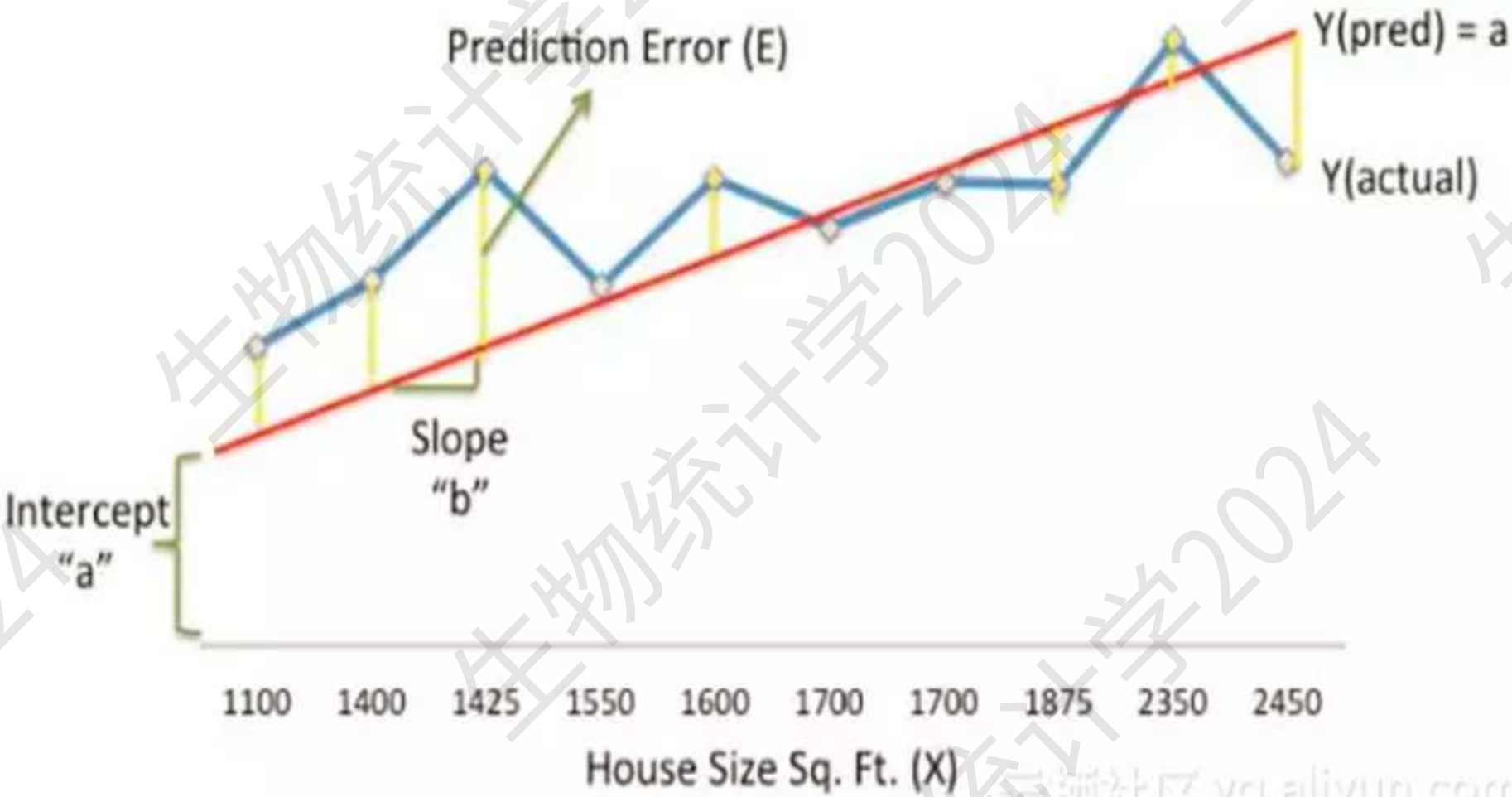
生物统计学2024

1. 线性回归：找到一条直线来预测目标值

- 一个简单的场景：已知房屋价格与尺寸的历史数据，问面积为2000时，售价为多少？

House Size sq.ft (X)	1400	1600	1700	1875	1100	1550	2350	2450	1425	17(
House Price$ (Y)	245,000	312,000	279,000	308,000	199,000	219,000	405,000	324,000	319,000	255,

生物统计学2024



生物统计学2024
云栖社区 yq.aliyun.com

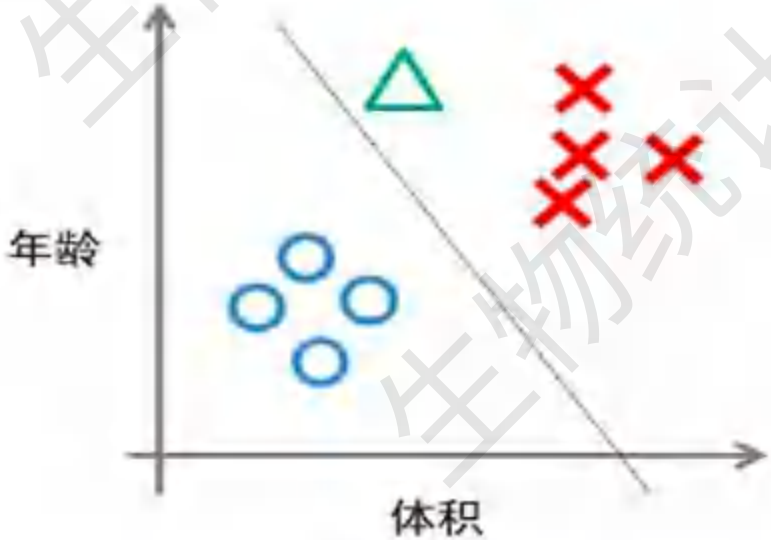
线性回归的应用

- **预测客户终生价值：** 基于老客户历史数据与客户生命周期的关联关系，建立线性回归模型，预测新客户的终生价值，进而开展针对性的活动。
- **机场客流量分布预测：** 以海量机场WiFi数据及安检登机值机数据，通过数据算法实现机场航站楼客流分析与预测。
- **货币基金资金流入流出预测：** 通过用户基本信息数据、用户申购赎回数据、收益率表和银行间拆借利率等信息，对用户的申购赎回数据的把握，精准预测未来每日的资金流入流出情况。
- **电影票房预测：** 依据历史票房数据、影评数据、舆情数据等互联网公众数据，对电影票房进行预测。

生物统计学2024

2. 逻辑回归：找到一条直线来分类数据

- 逻辑回归虽然名字叫回归，却是属于分类算法，是通过Sigmoid函数将线性函数的结果映射到Sigmoid函数中，预估事件出现的概率并分类。



目标：预测肿瘤的性质

输入：肿瘤的体积，
患者的年龄

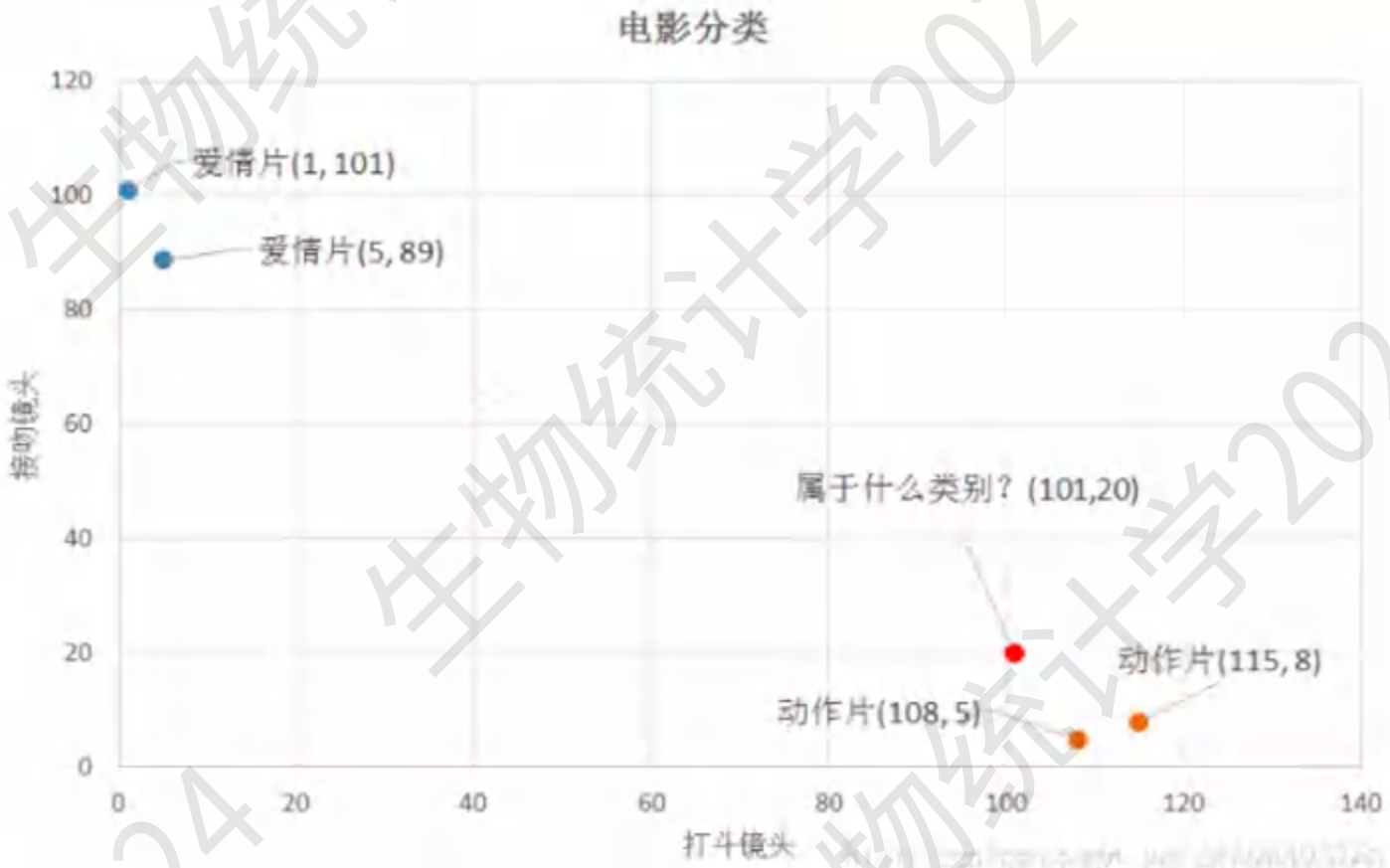
输出：良性或恶性

逻辑回归从直观上来说是画出了一条分类线。位于分类线一侧的数据，概率>0.5,属于分类A；位于分类线另一侧的数据，概率<0.5,属于分类B。

生物统计学2024

3. K-近邻：用距离度量最相邻的分类标签

- 一个简单的场景：已知一个电影中的打斗和接吻镜头数，判断它是属于爱情片还是动作片。当接吻镜头数较多时，根据经验我们判断它为爱情片。那么计算机如何进行判别呢？



生物统计学2024

4. 朴素贝叶斯：选择后验概率最大的类为分类标签

- 一个简单的场景：一号碗(C1)有30颗水果糖和10颗巧克力糖，二号碗(C2)有水果糖和巧克力糖各20颗。现在随机选择一个碗，从中摸出一颗糖，发现是水果糖。
- 问这颗水果糖(X)最有可能来自哪个碗？

似然函数　　先验概率

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

后验概率　　证据因子（evidence）

$$p(x) = \sum_{j=1}^{2} p(x \mid \omega_j) P(\omega_j)$$

生物统计学2024

例如上面的例子中： P(X): 水果糖的概率为5/8

P(X|C1): 一号碗中水果糖的概率为3/4

P(X|C2): 二号碗中水果糖的概率为2/4

P(C1)=P(C2): 两个碗被选中的概率相同，为1/2

则水果糖来自一号碗的概率为:

$P(C1|X)=P(X|C1)P(C1)/P(X)=(3/4)(1/2)/(5/8)=3/5

水果糖来自二号碗的概率为:

P(C2|X)=P(X|C2)P(C2)/P(X)=(2/4)(1/2)/(5/8)=2/5

P(C1|X)＞P(C2|X)

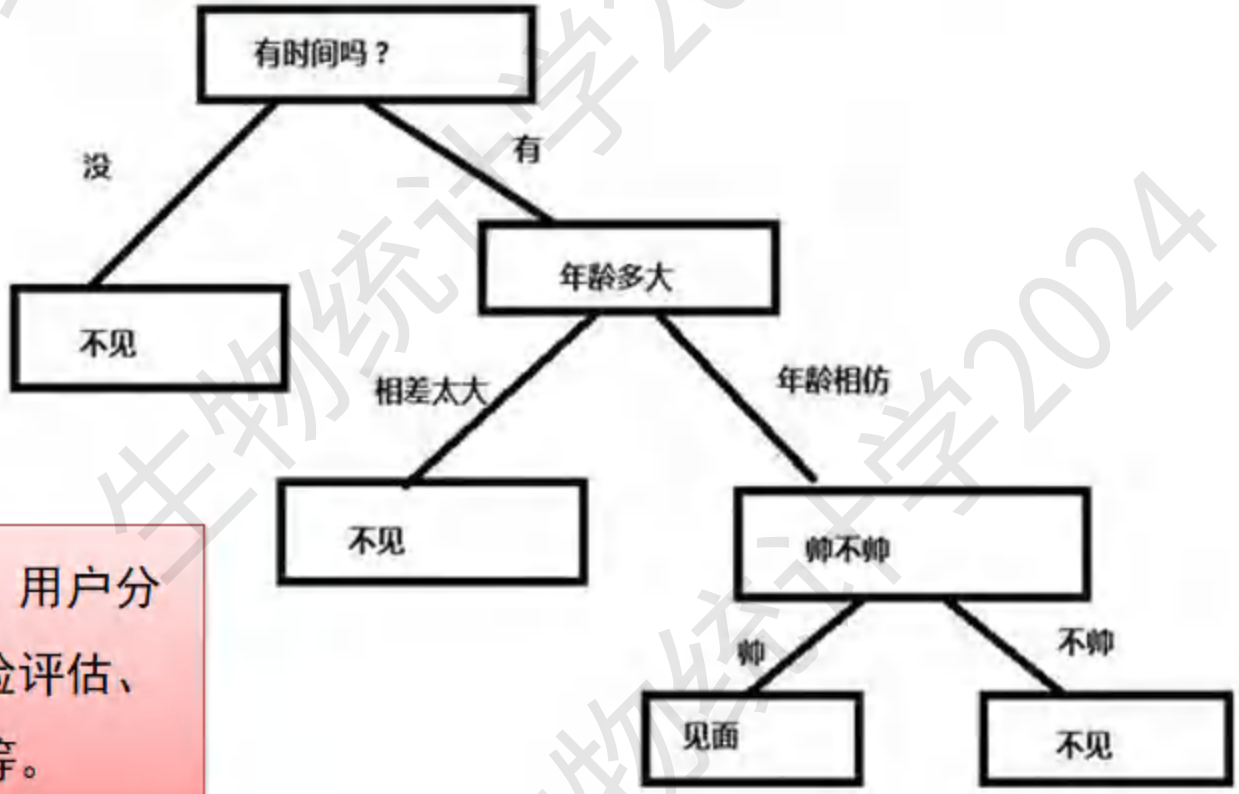
因此这颗糖最有可能来自一号碗。

朴素贝叶斯的主要应用有文本分类、垃圾文本过滤，情感判别，多分类实时预测等。

生物统计学2024

5. 决策树：构造熵值下降最快的分类树

- 一个简单的场景：
  相亲时，可能首先检测相亲对方是否有时间。如果有，则考虑进一步接触，再观察其是否有上进心，如果没有，直接Say Goodbye。如果有，则在看帅不帅，帅的可以列入候选名单。

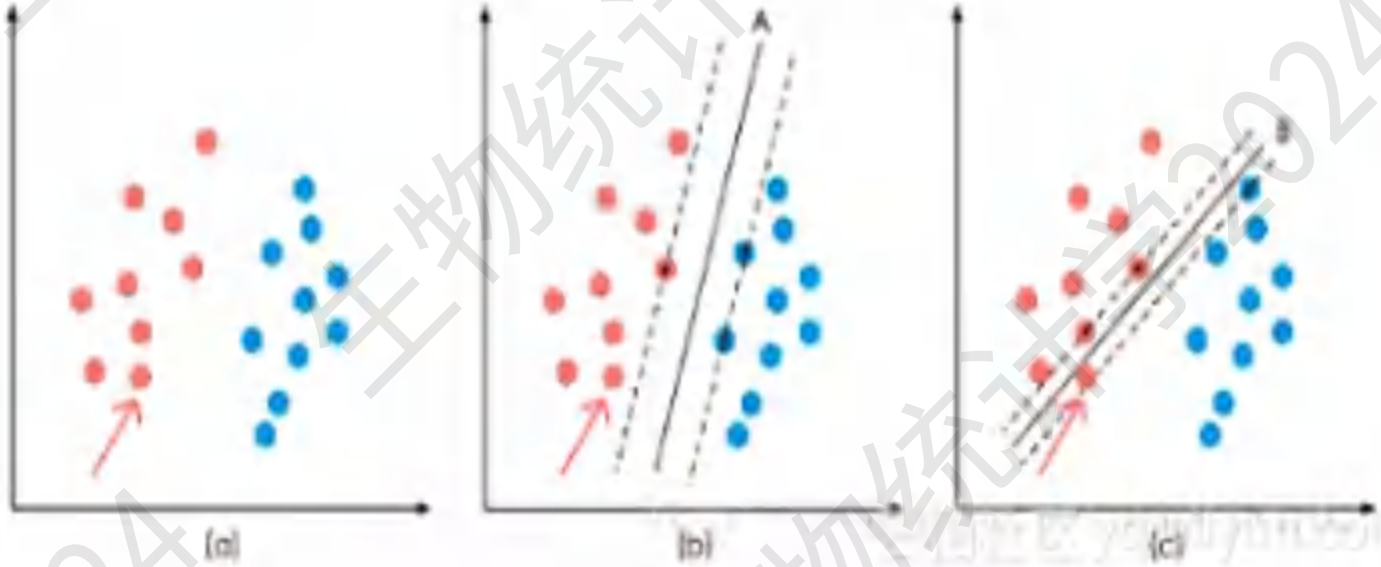


决策树可以应于：用户分级评估、贷款风险评估、选股、投标决策等。

生物统计学2024

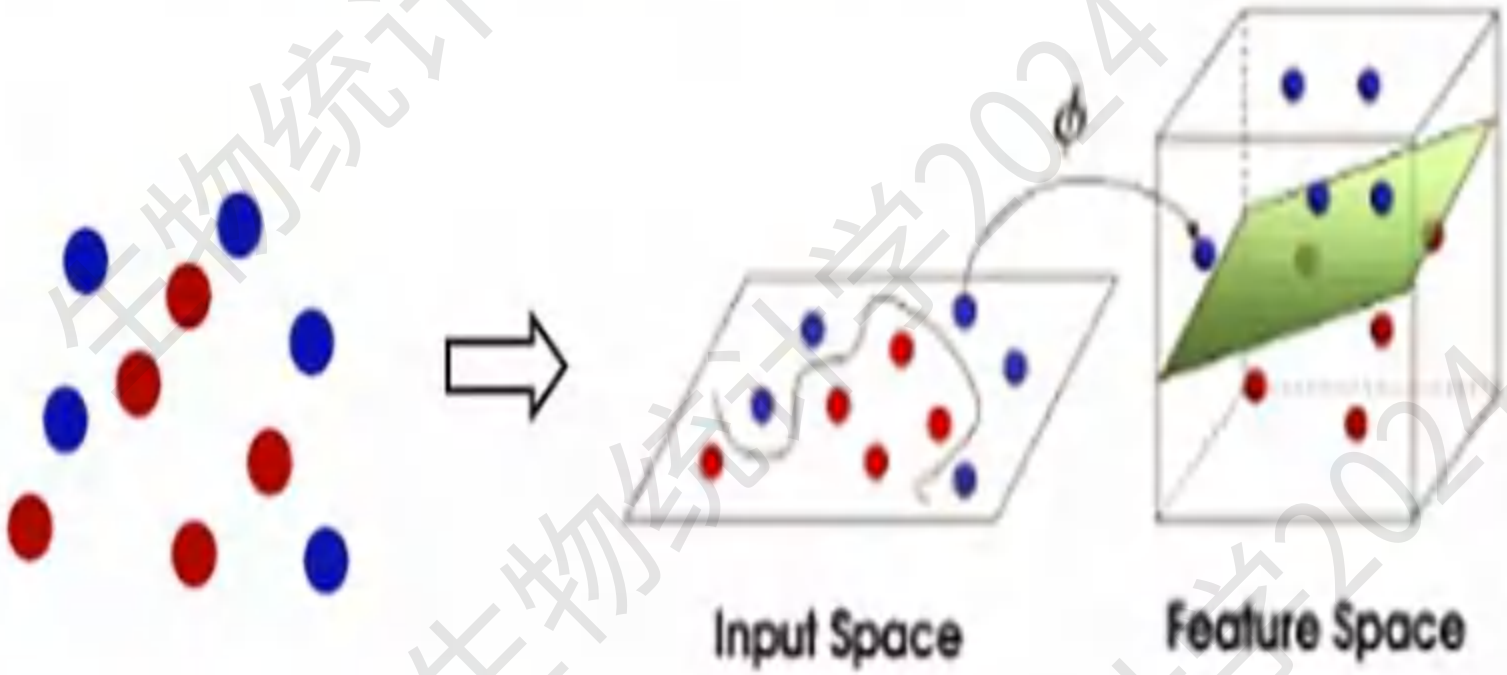
6. 支持向量机（SVM）：构造超平面，分类非线性数据

一个简单的场景：
要求用一根线将不同颜色的球分开，要求尽量在放更多球之后，仍然适用。A、B两条线都可以满足条件。再继续增加球，线A仍可以将球很好的分开，而线B则不可以。



生物统计学2024

进一步增加难度，当球没有明确的分界线，用一条直线已经无法将球分开，该怎么解决？

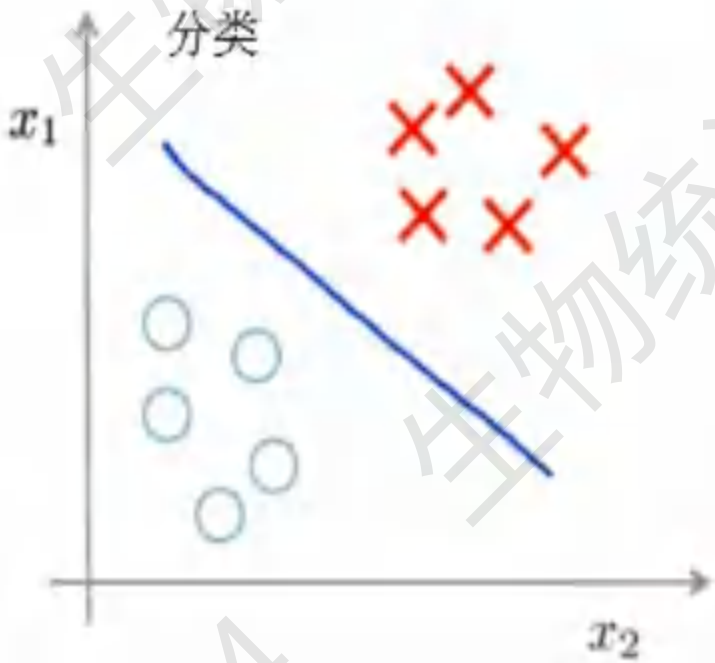


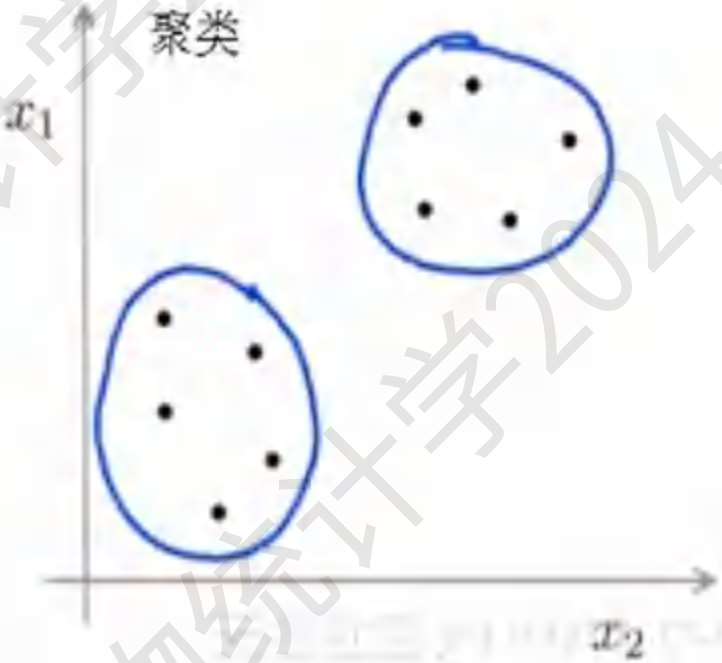
SVM 可应用于垃圾邮件识别、手写识别、文本分类、选股等。

生物统计学2024

7. K-means：计算质心，聚类无标签数据

- 在上面介绍的分类算法中，需要被分类的数据集已经有标记，例如数据集已经标记为○或者×，通过学习出假设函数对这两类数据进行划分。而对于没有标记的数据集，希望能有一种算法能够自动的将相同元素分为紧密关系的子集或簇，这就是聚类算法。





生物统计学2024

8. 关联分析：挖掘啤酒与尿布（频繁项集）的关联规则

- 算法中几个相关的概念：

➢ **频繁项集：**在数据库中大量频繁出现的数据集合。例如购物单数据中{'啤酒'}、{'尿布'}、{'啤酒', '尿布'}出现的次数都比较多。

➢ **关联规则：**由集合 A，可以在某置信度下推出集合 B。即如果 A 发生了，那么 B 也很有可能会发生。例如购买了{'尿布'}的人很可能会购买{'啤酒'}。

➢ **支持度：**指某频繁项集在整个数据集中的比例。假设数据集有 10 条记录，包含{'啤酒', '尿布'}的有 5 条记录，那么{'啤酒', '尿布'}的支持度就是 5/10 = 0.5。

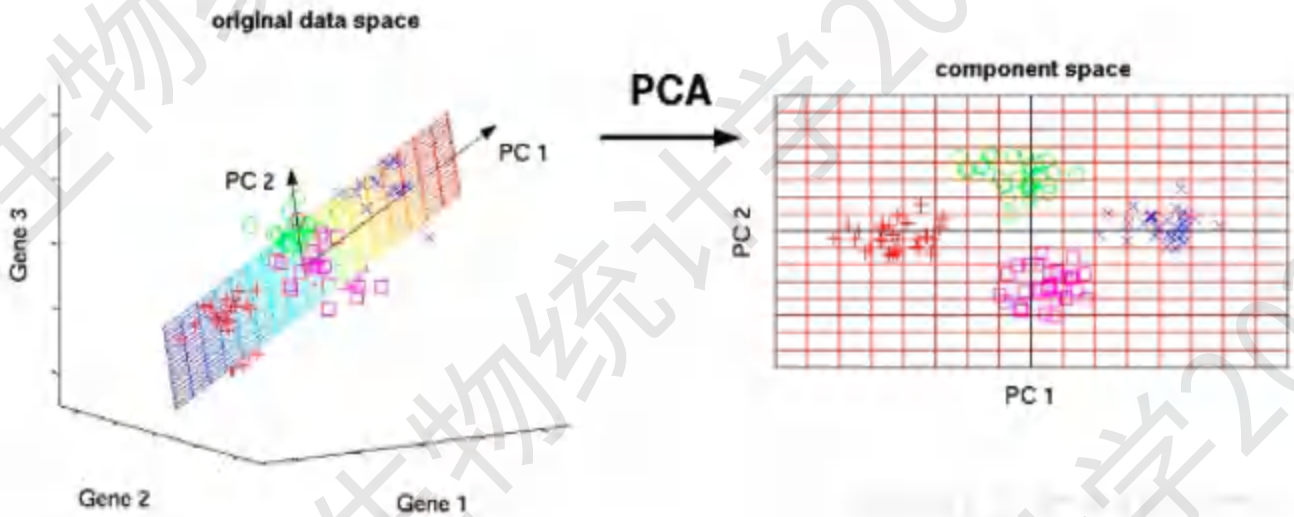
➢ **置信度：**有关联规则如{'尿布'} -> {'啤酒'}，它的置信度为 {'尿布'} -> {'啤酒'}

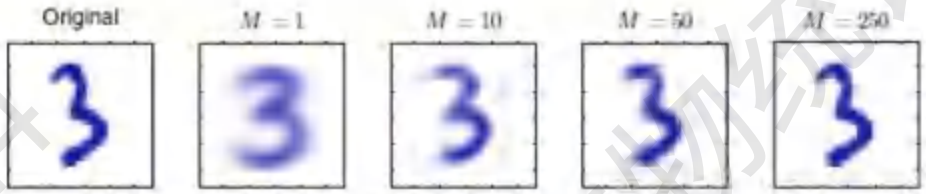
假设{'尿布', '啤酒'}的支持度为 0.45，{'尿布'}的支持度为 0.5，则{'尿布'} -> {'啤酒'}的置信度为 0.45 / 0.5 = 0.9。

生物统计学2024

9. PCA降维：减少数据维度，降低数据复杂度

- 降维是指将原高维空间中的数据点映射到低维度的空间中。因为高维特征的数目巨大，距离计算困难，分类器的性能会随着特征数的增加而下降；减少高维的冗余信息所造成的误差,可以提高识别的精度。

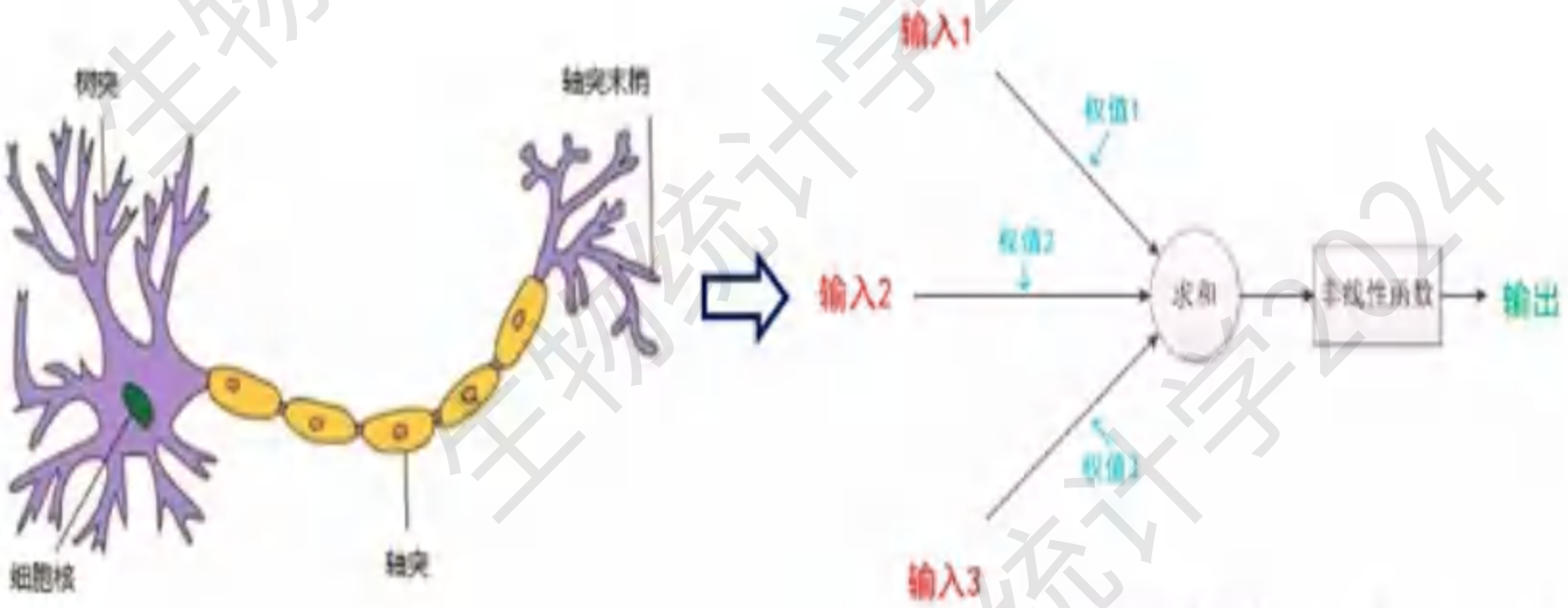




生物统计学2024

10. 人工神经网络：逐层抽象，逼近任意函数

- 前面介绍了九种传统的机器学习算法，现在介绍一下深度学习的基础：人工神经网络。 它是模拟人脑神经网络而设计的模型，由多个节点（人工神经元）相互联结而成，可以用来对数据之间的复杂关系进行建模。

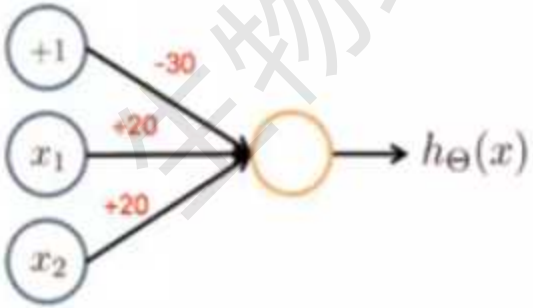


生物统计学2024

- 例如利用单层神经网络实现逻辑与门和同或门

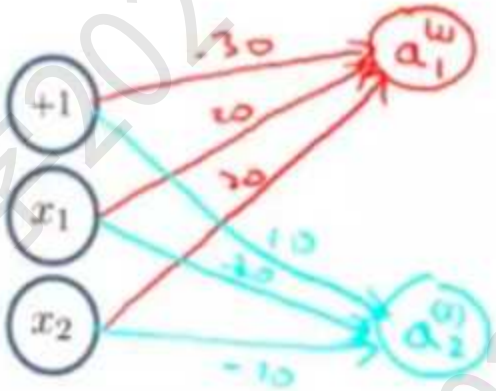
与门AND

$$h_\Theta(x) = g(-30 + 20x_1 + 20x_2)$$



x1	x2	hΘ(x)
0	0	g(−30)≈0
0	1	g(−10)≈0
1	0	g(−10)≈0
1	1	g(10)≈1

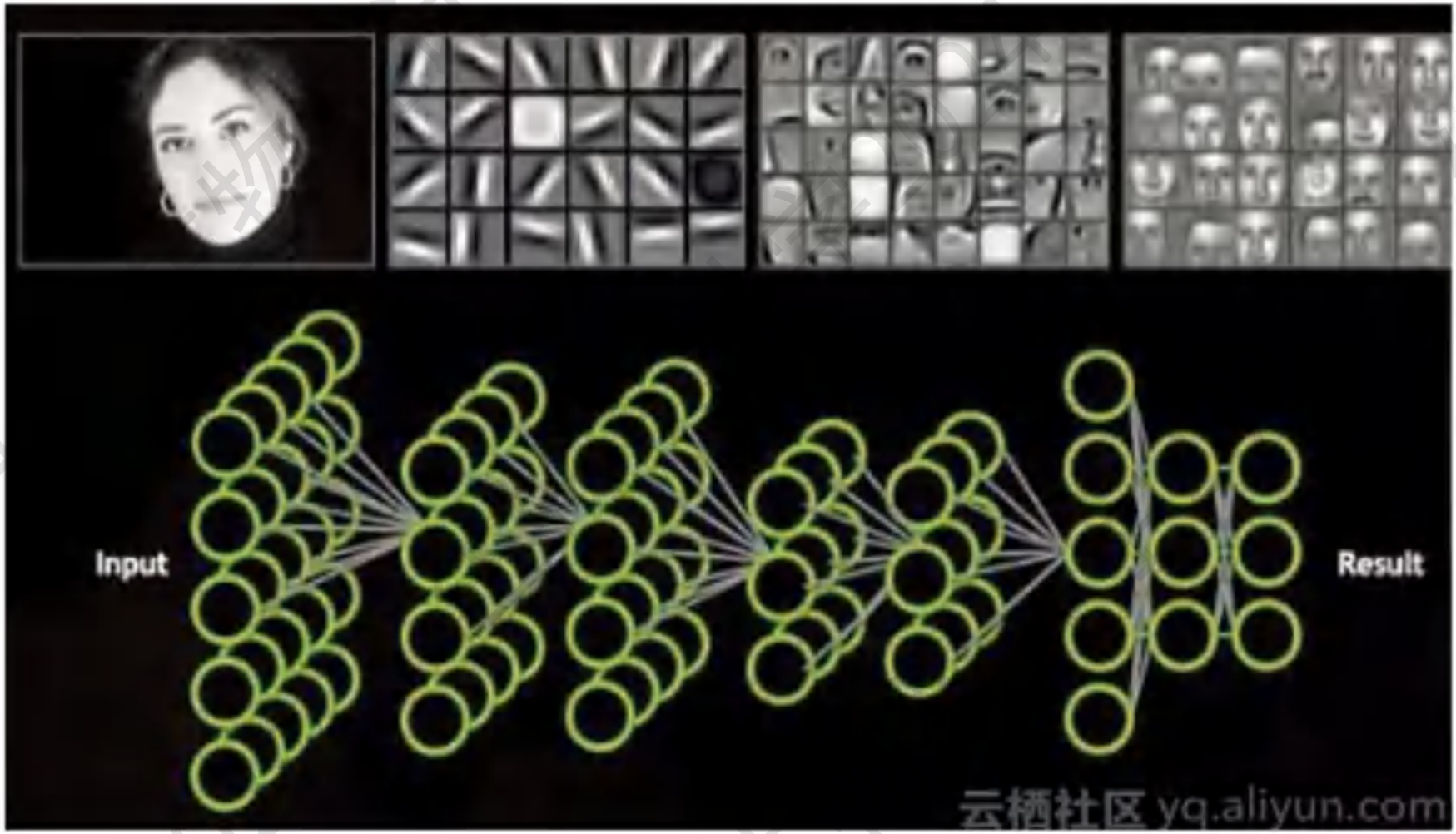
同或门XNOR



x1	x2	a1(2)	a2(2)	hΘ(x)
0	0	0	1	**1**
0	1	0	0	**0**
1	0	0	0	**0**
1	1	1	0	**1**

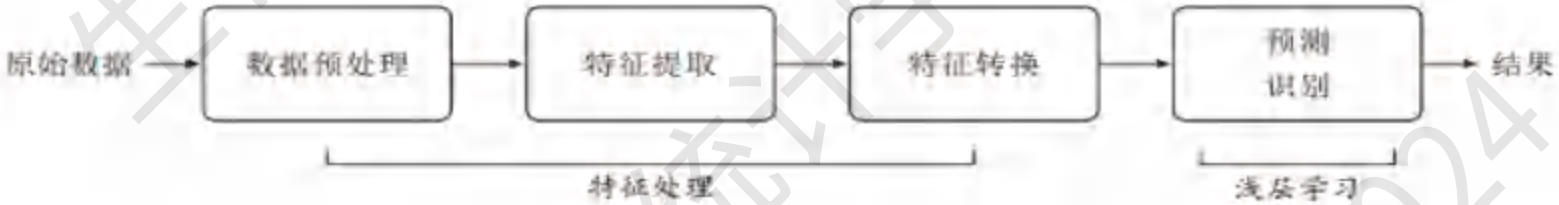
生物统计学2024

- 多层神经网络的每一层神经元学习到的是前一层神经元值的更抽象的表示，通过抽取更抽象的特征来对事物进行区分，从而获得更好的区分与分类能力。



11. 深度学习：赋予人工智能以璀璨的未来

- 深度学习就是一种基于对数据进行表征学习的方法，使用多层网络，能够学习抽象概念，同时融入自我学习，逐步从大量的样本中逐层抽象出相关的概念，然后做出理解，最终做出判断和决策。通过构建具有一定“深度”的模型，可以让模型来自动学习好的特征表示（从底层特征，到中层特征，再到高层特征），从而最终提升预测或识别的准确性。



传统机器学习的数据处理流程。

深度学习的数据处理流程。

生物统计学2024

深度学习的历史变迁：

深度学习经历了三次浪潮：

- 20世纪40年代～60年年代，深度学习的雏形出现在控制论中；
- 20世界80年代～90年代，深度学习表现为联结主义；
- 2006年以后，正式以深度学习之名复兴。

- **第一次浪潮：以感知机和线性模型为代表**

  不能解决与或问题

- **第二次浪潮：以多层感知机和BP模型为代表**

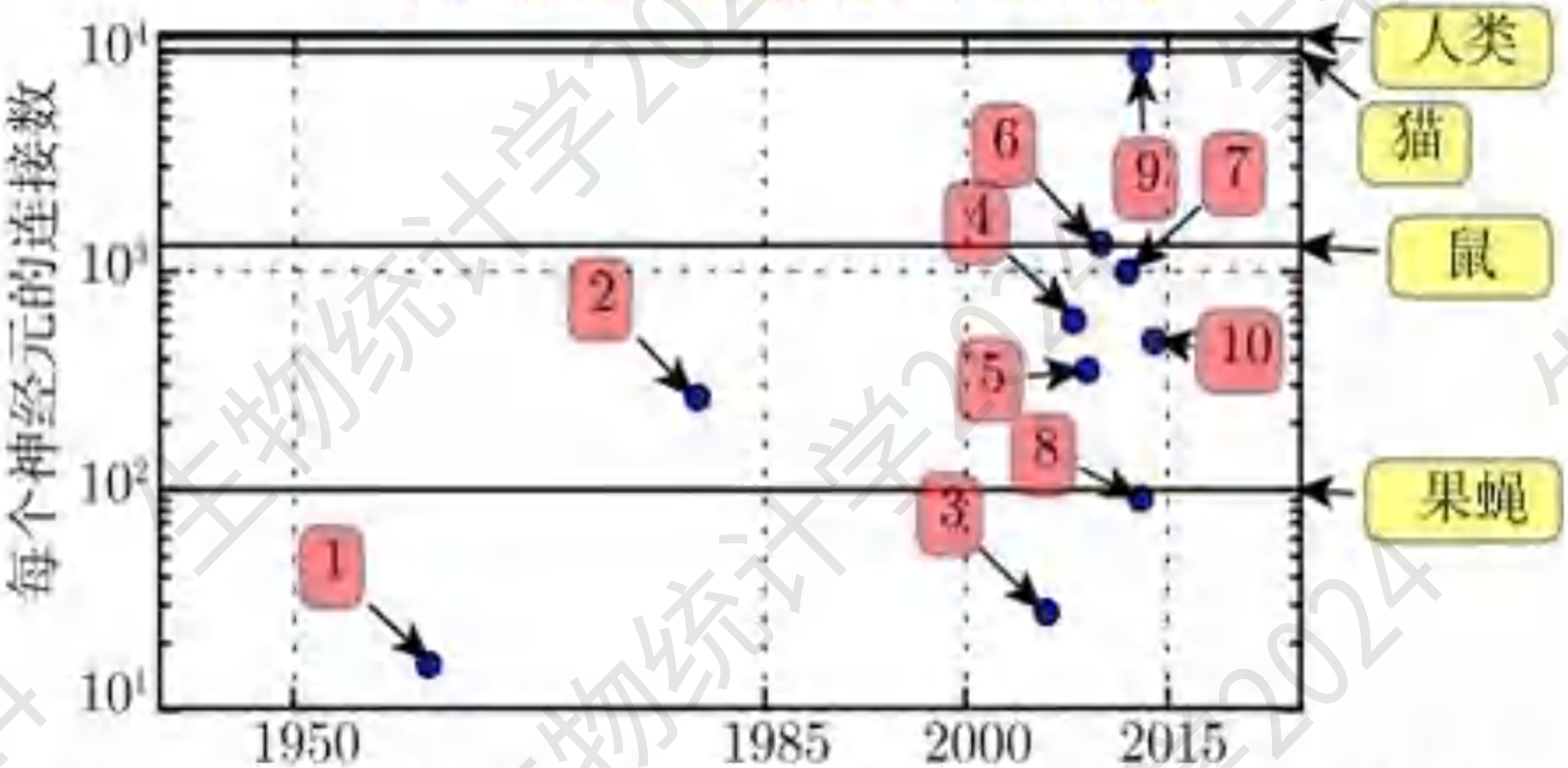
  以统计学为基础，应用核函数和图模型的支持向量机算法（SVM算法）等各种浅层有监督的机器学习模型广泛应用，且深度神经网络不可训练

- **第三次浪潮：以无监督学习为代表。**

  解决了深层神经网络的计算能力问题；解决了深度神经网络后向误差反馈梯度消失的问题。

生物统计学2024

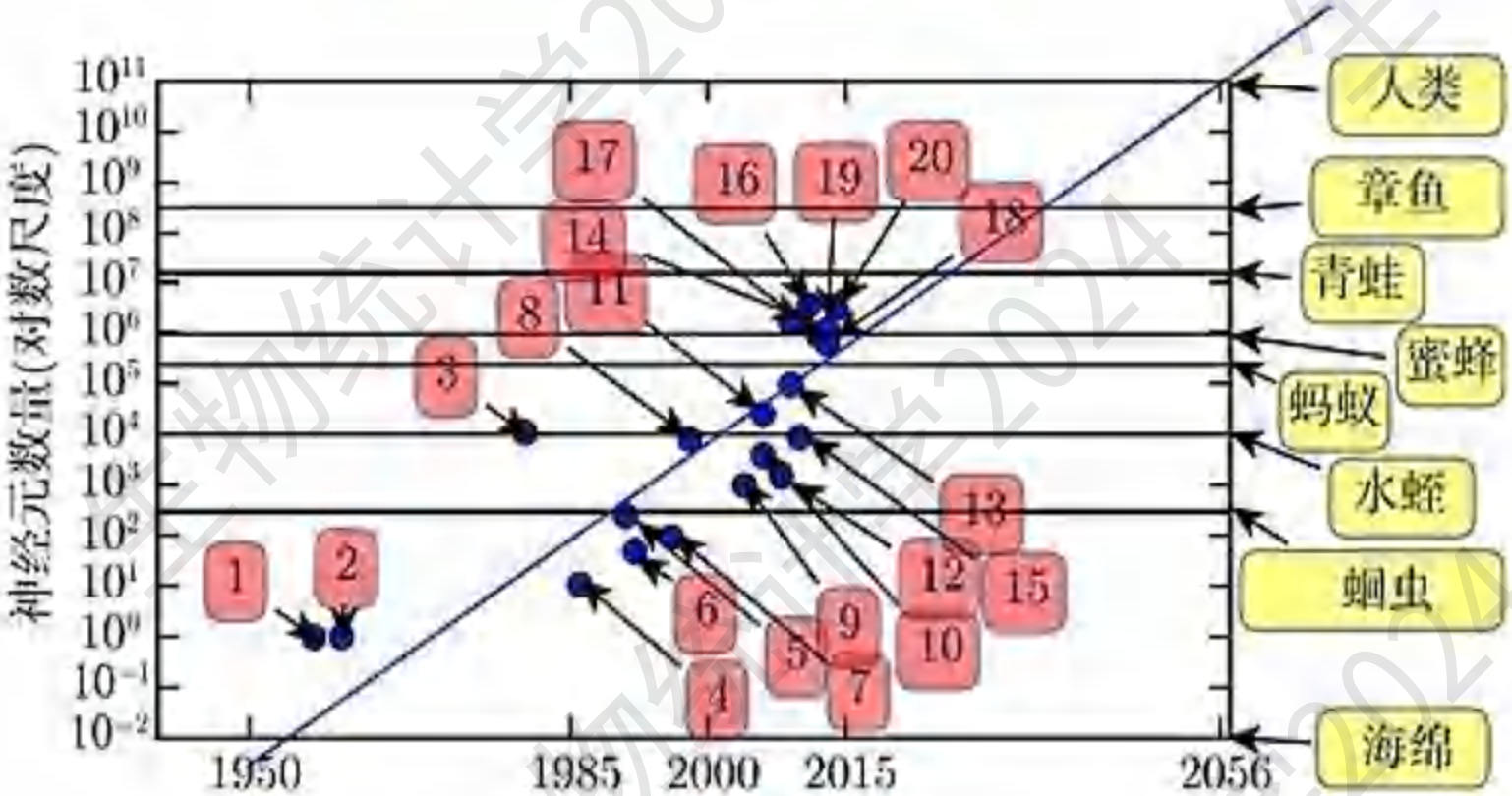
与日俱增的每个神经元的连接数



- 最初，人工神经网络中神经元之间的连接数受限于硬件能力。而现在，神经元之间的连接数大多是出于设计考虑。一些人工神经网络中每个神经元的连接数与猫一样多，并且对于其他神经网络来说，每个神经元的连接数与较小哺乳动物（如小鼠）一样多，这种情况是非常普遍的。甚至人类大脑每个神经元的连接数也没有过高的数量。
- 1. 自适应线性单元（Widrow and Hoff, 1960）；2. 神经认知机（Fukushima, 1980）；3. GPU- 加速卷积网络（Chellapilla et al., 2006）；4. 深度玻尔兹曼机（Salakhutdinov and Hinton, 2009a）；5. 无监督卷积网络（Jarrett et al., 2009b）；6. GPU- 加速多层感知机（Ciresan et al., 2010）；7. 分布式自编码器（Le et al., 2012）；8. Multi-GPU 卷积网络（Krizhevsky et al., 2012a）；9. COTS HPC 无监督卷积网络（Coates et al., 2013）；10. GoogLeNet（Szegedy et al., 2014a

生物统计学2024

与日俱增的神经网络规模



- 自从引入隐藏单元，人工神经网络的规模大约每 2.4 年翻一倍。
- 1. 感知机 （Rosenblatt, 1958, 1962）；2. 自适应线性单元 （Widrow and Hoff,1960）；3. 神经认知机 （Fukushima, 1980）；4. 早期后向传播网络 （Rumelhart et al., 1986b）；5. 用于语音识别的循环神经网络 （Robinson and Fallside, 1991）；6. 用于语音识别的多层感知机 （Bengio et al., 1991）；7. 均匀场 sigmoid 信念网络 （Saul et al., 1996）；8. LeNet5 （LeCun et al., 1998c）；9. 回声状态网络 （Jaeger and Haas, 2004）；10. 深度信念网络 （Hinton et al., 2006a）；11. GPU- 加速卷积网络 （Chellapilla et al., 2006）；12. 深度玻尔兹曼机 （Salakhutdinov and Hinton, 2009a）；13. GPU加速深度信念网络 （Raina et al., 2009a）；14. 无监督卷积网络 （Jarrett et al., 2009b）；15. GPU- 加速多层感知机 （Ciresan et al., 2010）；16. OMP-1 网络 （Coates and Ng, 2011）；17. 分布式自编码器 （Le et al., 2012）；18. MultiGPU 卷积网络 （Krizhevsky et al., 2012a）；19. COTS HPC 无监督卷积网络 （Coates et al., 2013）；20. GoogLeNet （Szegedy et al., 2014a）

- 目前深度学习的应用十分广泛，例如图像识别、语音识别、机器翻译、自动驾驶、金融风控、智能机器人等。

Finance

Radiology

生物统计学2024
云栖社区 yq.aliyun.com

Machine Learning in biology

Supervised

$x \longrightarrow y$

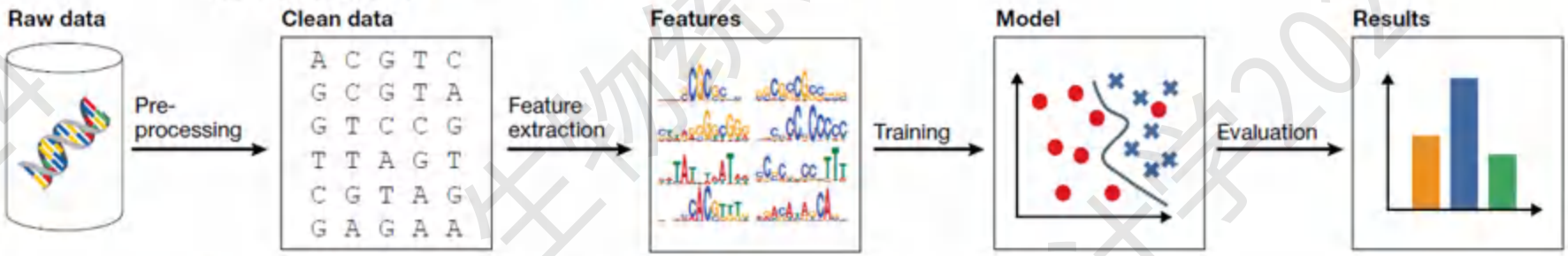
- Linear regression
- Logistic regression
- Random Forest
- SVM
- ...

Unsupervised

x

- PCA
- Factor analysis
- Clustering
- Outlier detection
- ...

Machine learning algorithm: supervised and unsupervised



An exemplification of Machine learning in biology : classification model

Angermueller C, et al, 2016, Deep learning for computational biol

Deep learning

Deep learning is a part of machine learning.

Deep:

Complex Model : Multi-layer characters, Many parameters (Curse of dimensionality)

Training data & Testing data : a big volume of data

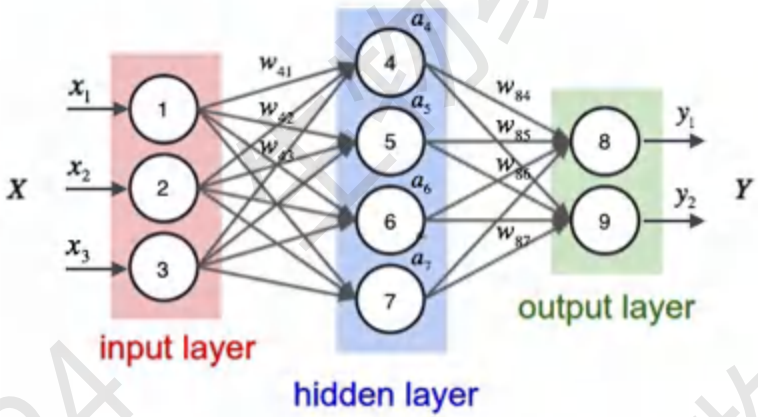
Adjustment: A single model construction could cost one week

几乎所有深度学习算法都可以被描述为一个简单配方：
特定数据集、代价函数、优化过程和模型

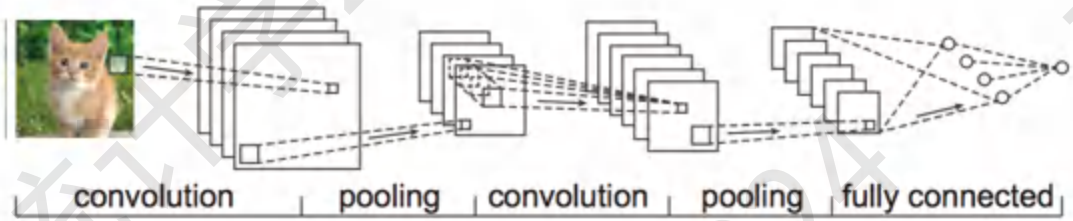
生物统计学2024

Deep learning

Now, neural network and convolution neural network models that work best



Neural Network



Convolution Neural Network

生物统计学2024

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

Pooling, SPM

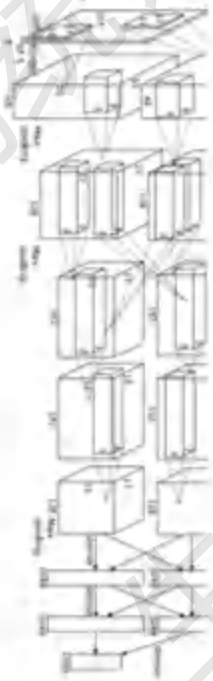
Linear SVM

[Lin CVPR 2011]

Lion image by Swissfrog is licensed under CC BY 3.0

Year 2012

SuperVision

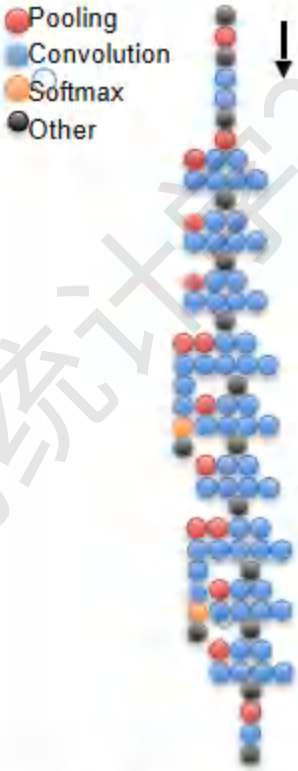


[Krizhevsky NIPS 2012]

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

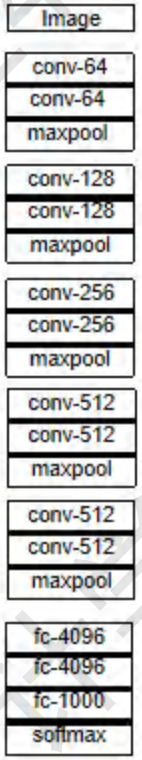
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

VGG



[Simonyan arxiv 2014]

Year 2015

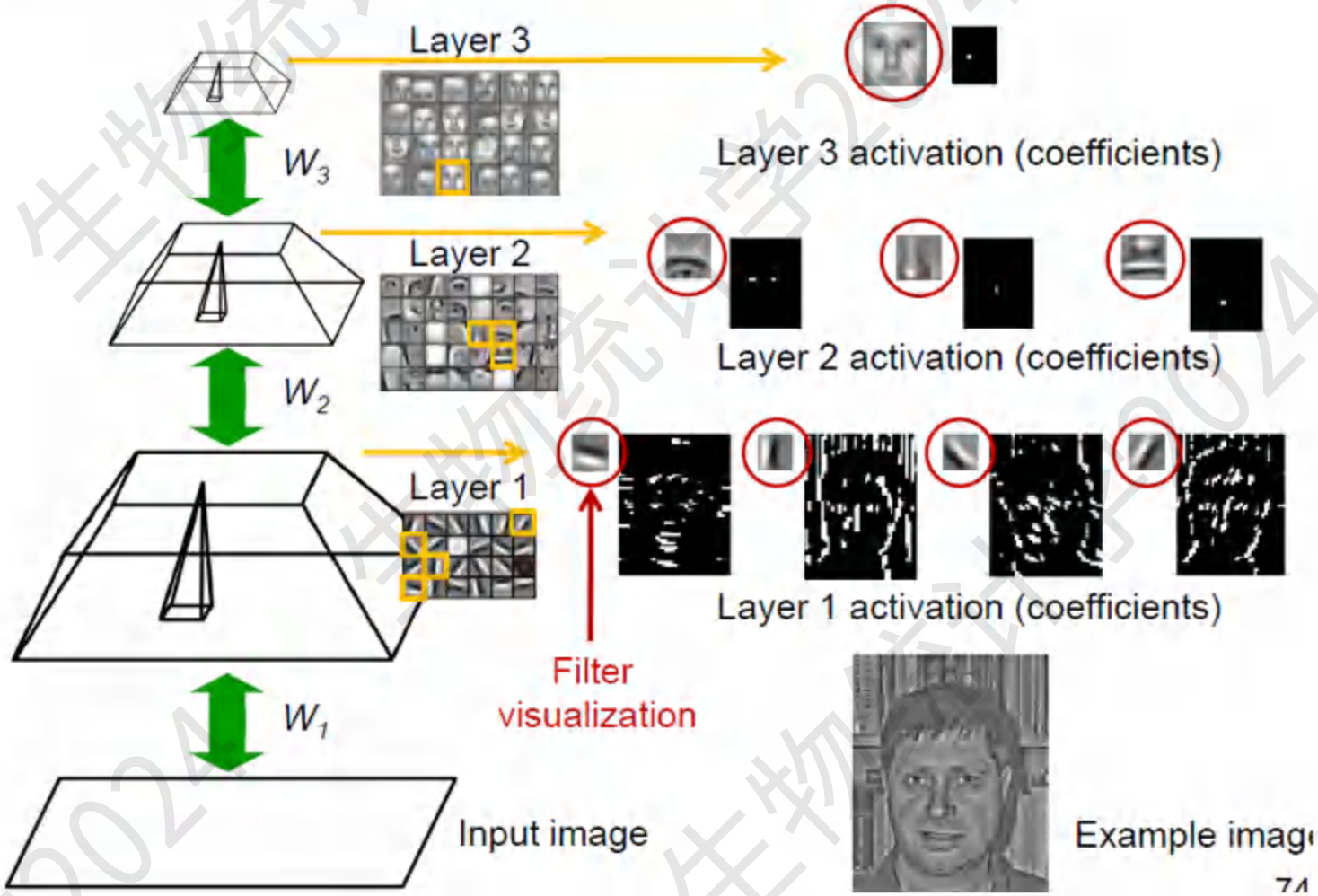
MSRA

[He ICCV 2015]

生物统计学2024

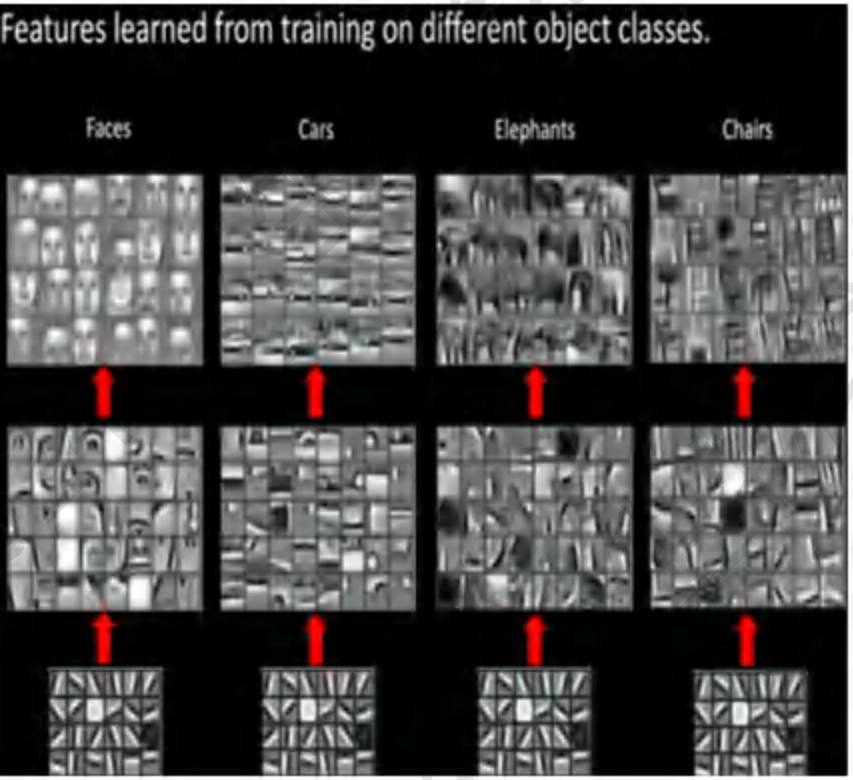
深度学习的应用

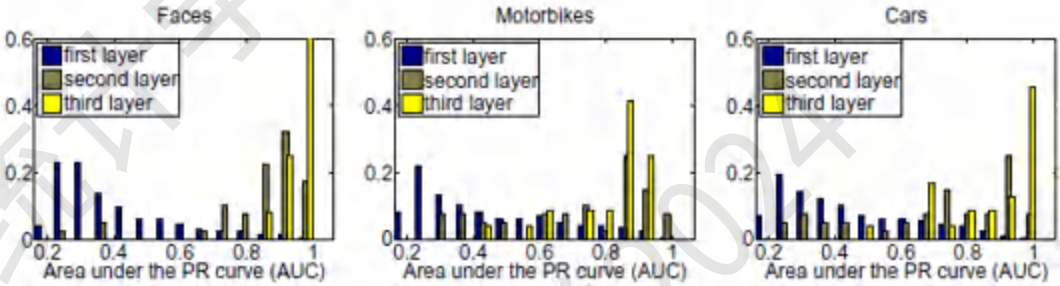
- 深度学习在图像识别上的应用



深度学习的应用

- 深度学习在图像识别上的应用





Features	Faces	Motorbikes	Cars
First layer	0.39±0.17	0.44±0.21	0.43±0.19
Second layer	0.86±0.13	0.69±0.22	0.72±0.23
Third layer	0.95±0.03	0.81±0.13	0.87±0.15

生物统计学2024

深度学习的应用

- 深度学习在图像识别上的应用

originals

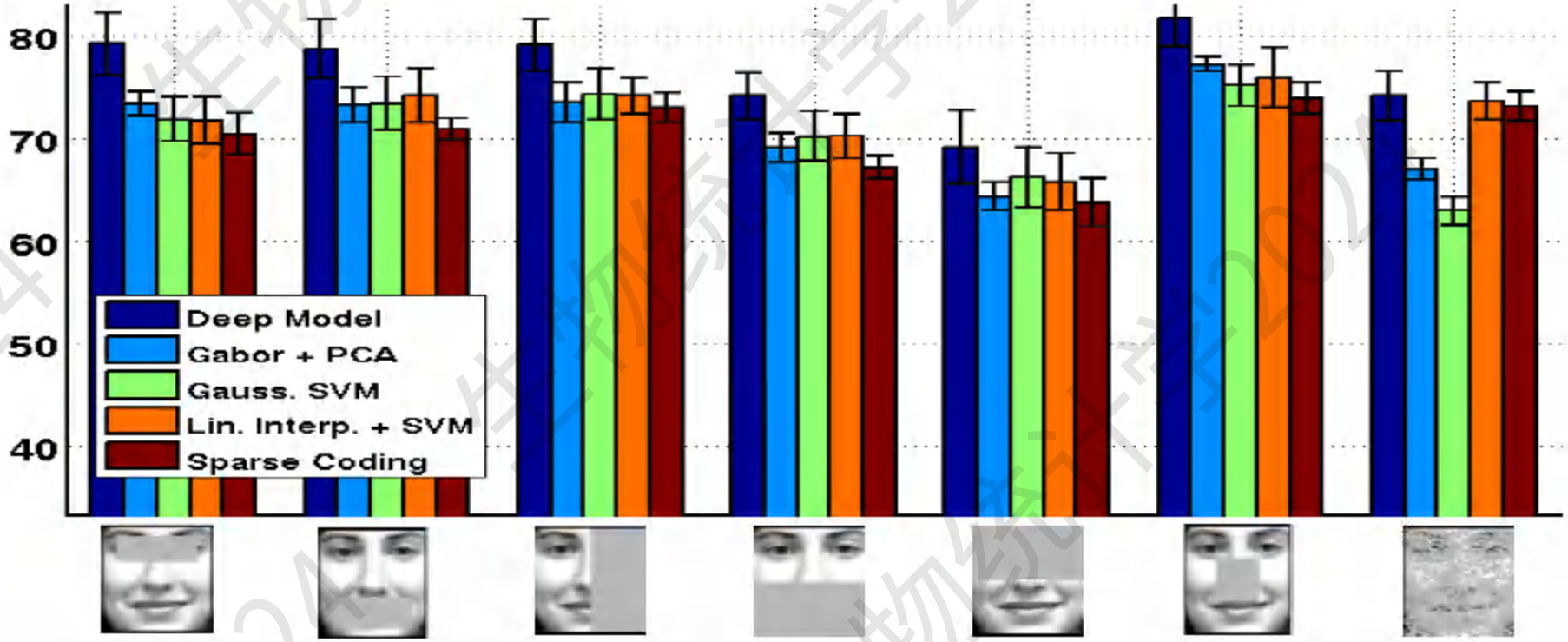
Type 1 occlusion: eyes

Restored images

生物统计学2024

深度学习的应用

- 深度学习在图像识别上的应用



生物统计学2024

深度学习的应用

- 深度学习在图像识别上的应用

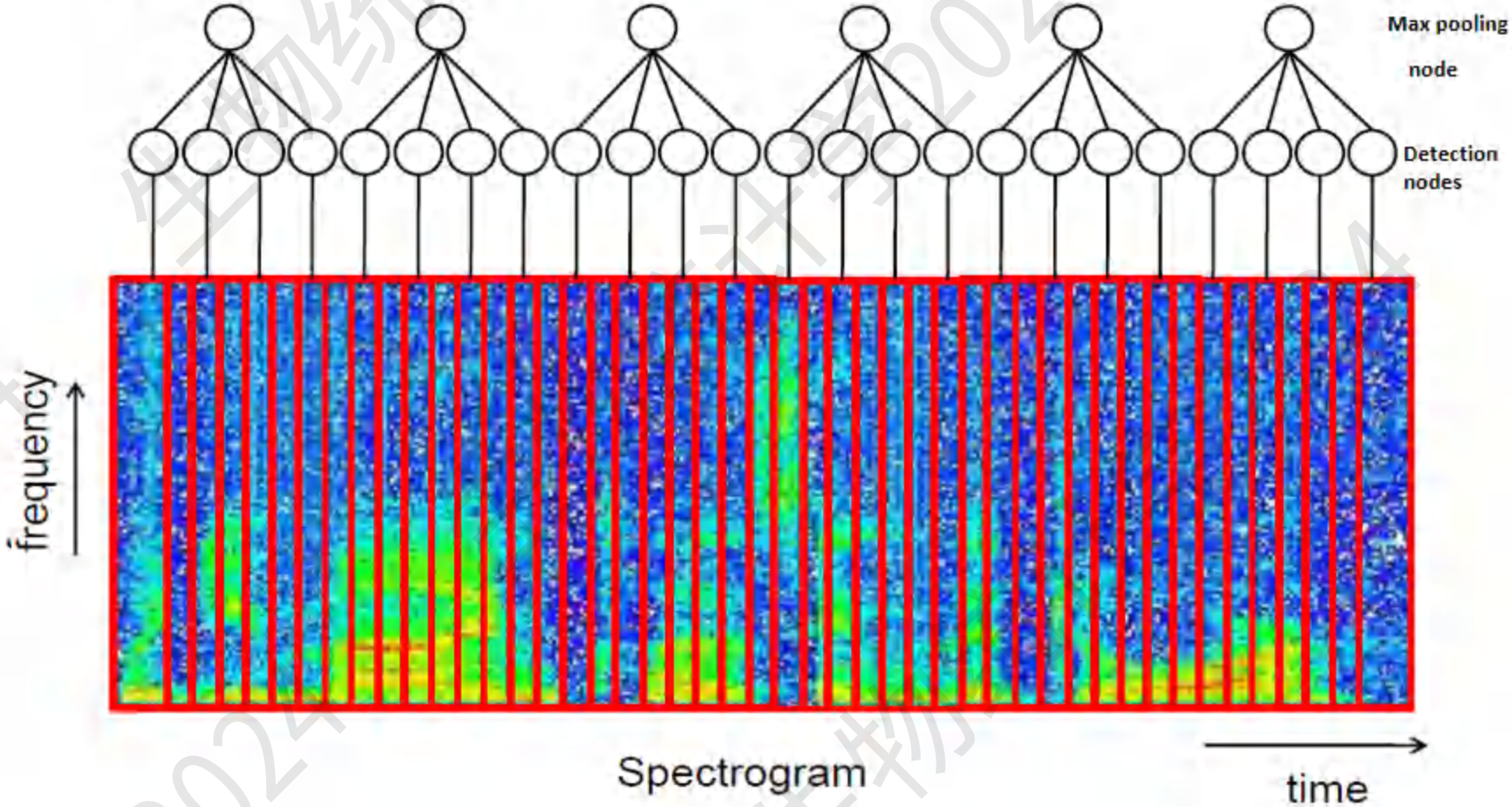
< Caltech 256 >

# of training images	30	60
Griffin et al. [2]	34.10	-
vanGemert et al., PAMI 2010	27.17	-
ScSPM [Yang et al., CVPR 2009]	34.02	40.14
LLC [Wang et al., CVPR 2010]	41.19	47.68
Sparse CRBM [Sohn et al., ICCV 2011]	42.05	47.94

生物统计学2024

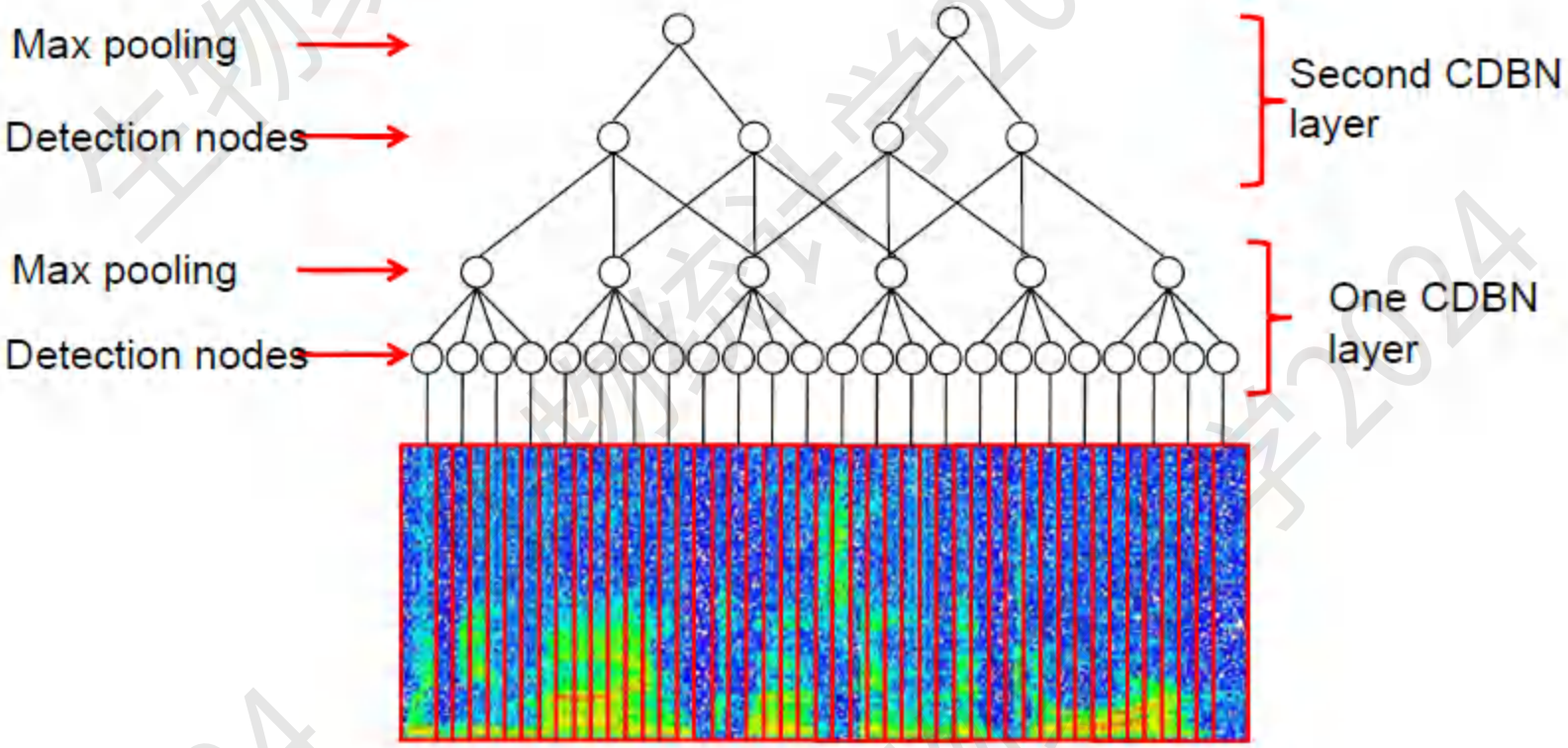
深度学习的应用

- 深度学习在音频识别上的应用



深度学习的应用

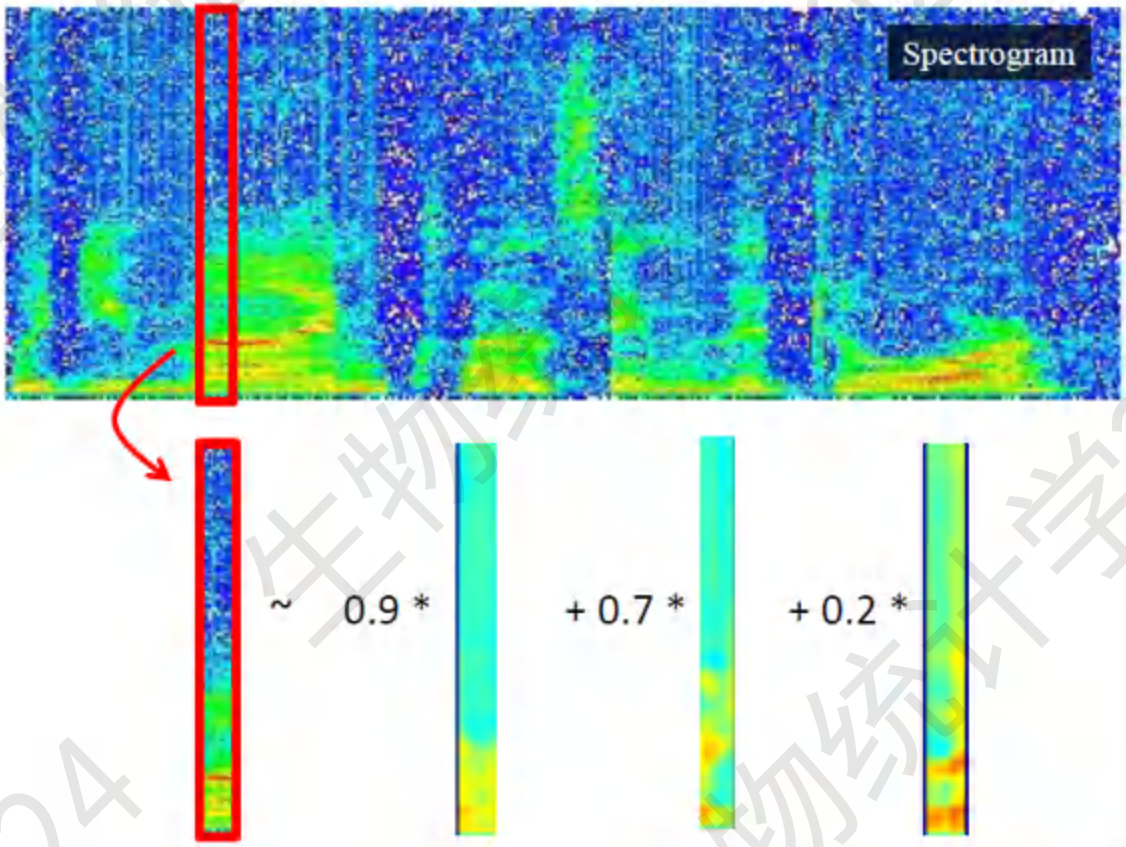
- 深度学习在音频识别上的应用



生物统计学2024

深度学习的应用

- 深度学习在音频识别上的应用



[Lee, Largman, Pham, Ng, NIPS 2009]

生物统计学2024

深度学习的应用

- 深度学习在音频识别上的应用
- Speaker identification

TIMIT Speaker identification	Accuracy
Prior art (Reynolds, 1995)	99.7%
Convolutional DBN	100.0%

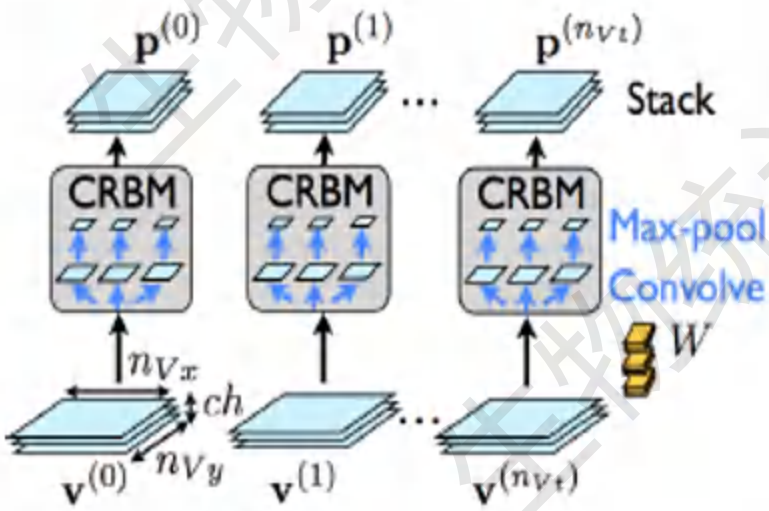
- Phone classification

TIMIT Phone classification	Accuracy
Clarkson et al. (1999)	77.6%
Petrov et al. (2007)	78.6%
Sha & Saul (2006)	78.9%
Yu et al. (2009)	79.2%
Convolutional DBN	80.3%
Transformation-invariant RBM (Sohn et al., ICML 2012)	81.5%

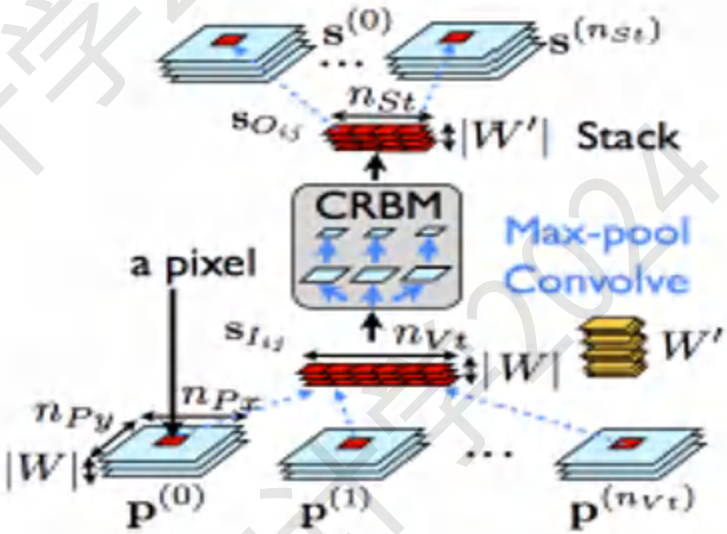
生物统计学2024

深度学习的应用

- 深度学习在视频识别上的应用



Spatial pooling layer



Temporal pooling layer

生物统计学2024

深度学习的应用

- 深度学习在视频识别上的应用

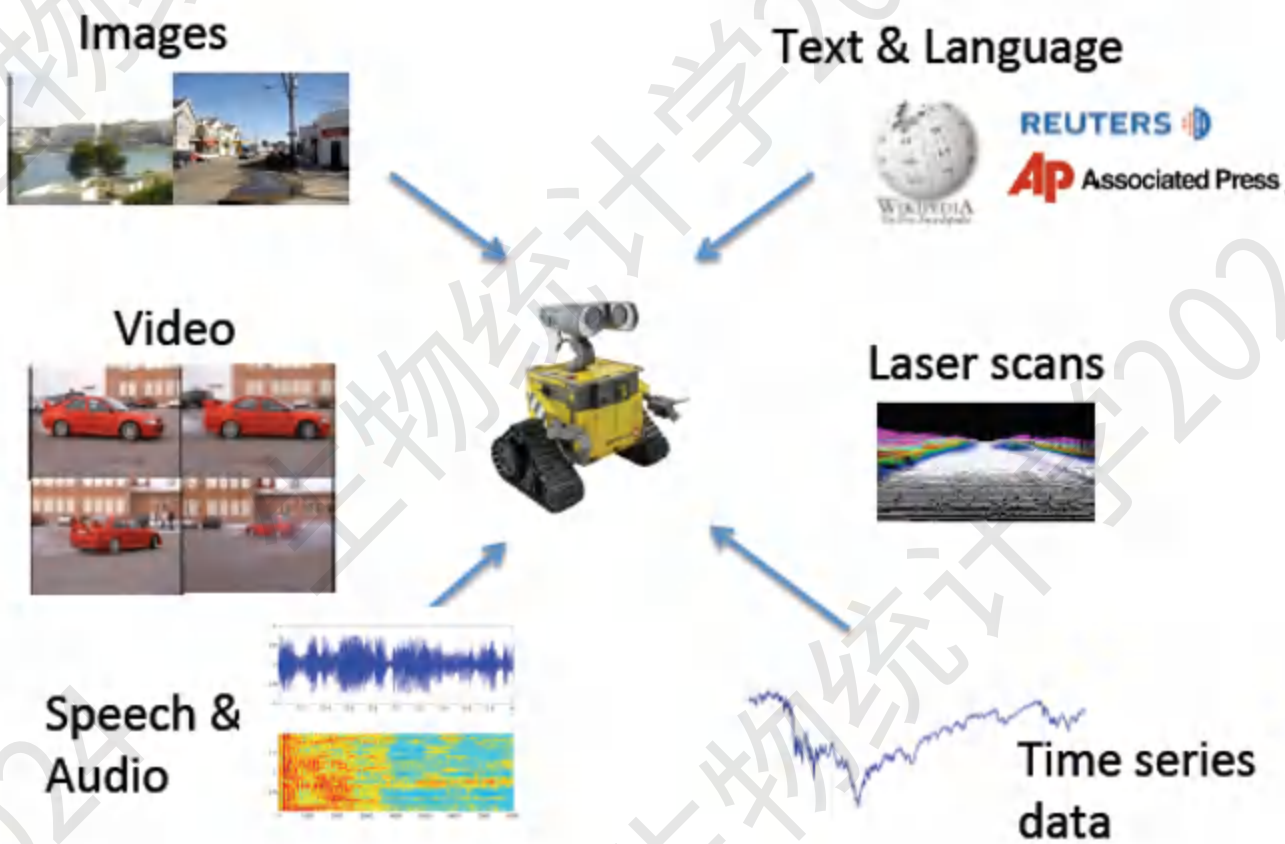
Video Activity recognition (Hollywood 2 benchmark)

Method	Accuracy
Hessian + ESURF [Williems et al 2008]	38%
Harris3D + HOG/HOF [Laptev et al 2003, 2004]	45%
Cuboids + HOG/HOF [Dollar et al 2005, Laptev 2004]	46%
Hessian + HOG/HOF [Laptev 2004, Williems et al 2008]	46%
Dense + HOG / HOF [Laptev 2004]	47%
Cuboids + HOG3D [Klaser 2008, Dollar et al 2005]	46%
Unsupervised feature learning (our method)	**52%**

Unsupervised feature learning significantly improves on the previous state-of-the-art.

深度学习的应用

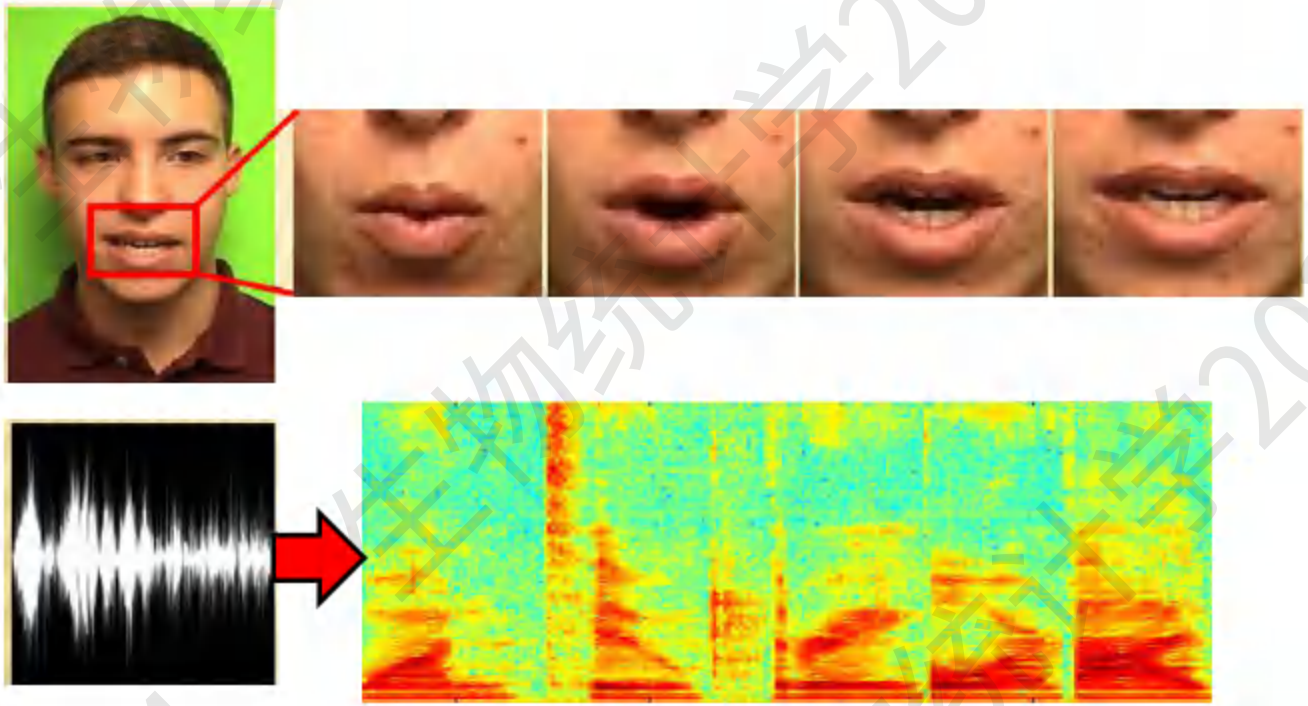
- 深度学习在多模态学习中的应用



生物统计学2024

深度学习的应用

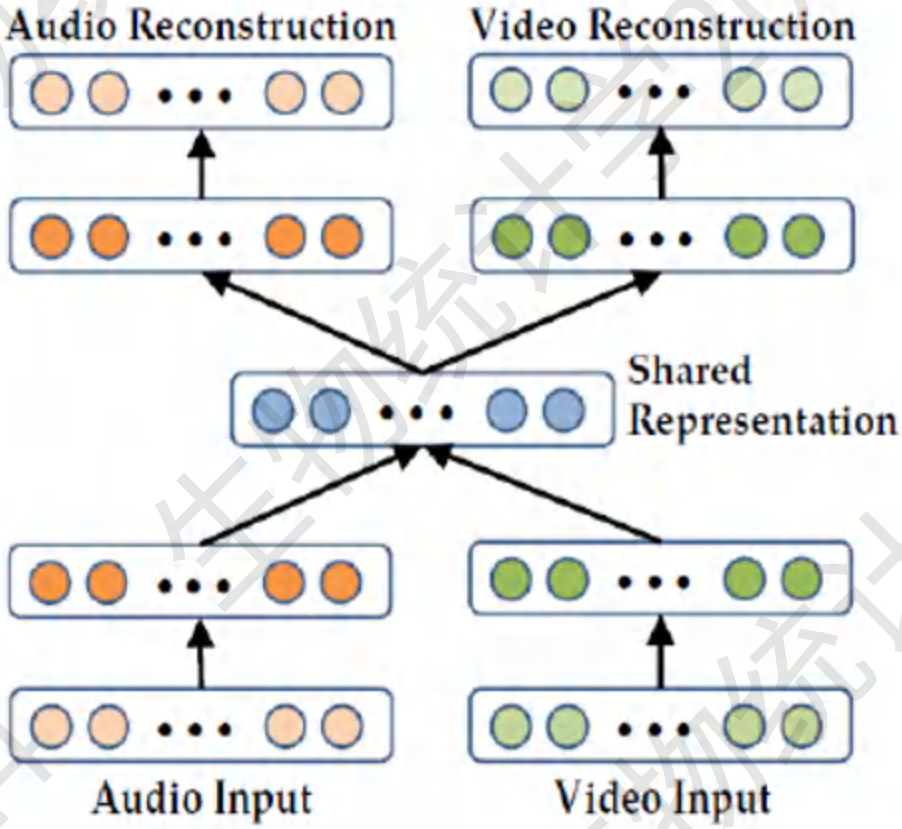
- 深度学习在多模态学习中的应用



生物统计学2024

深度学习的应用

- 深度学习在多模态学习中的应用



生物统计学2024

深度学习的应用

- 深度学习在多模态学习中的应用
  - Visualization of learned filters

Audio(spectrogram) and Video features learned over 100ms windows

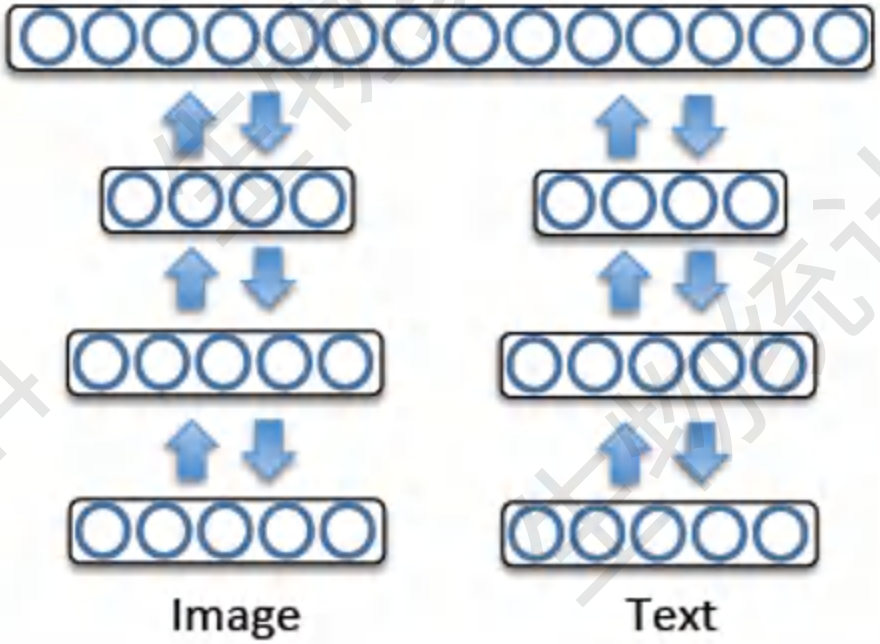
  - Results: AVLetters Lip reading dataset

Method	Accuracy
Prior art (Zhao et al., 2009)	58.9%
Multimodal deep autoencoder (Ngiam et al., 2011)	**65.8%**

生物统计学2024

深度学习的应用

- 深度学习在多模态学习中的应用



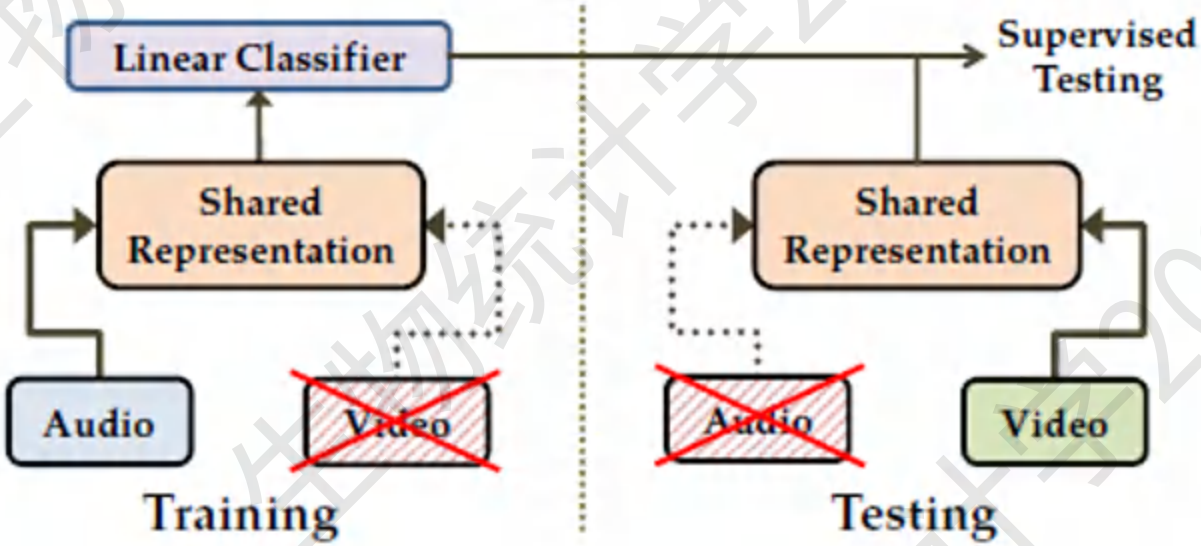
- Multimodal Inputs (images + text), 38 classes.

Learning Algorithm	Mean Average Precision
Image-text SVM	0.475
Image-text LDA	0.492
Multimodal DBN	0.566

生物统计学2024

深度学习的应用

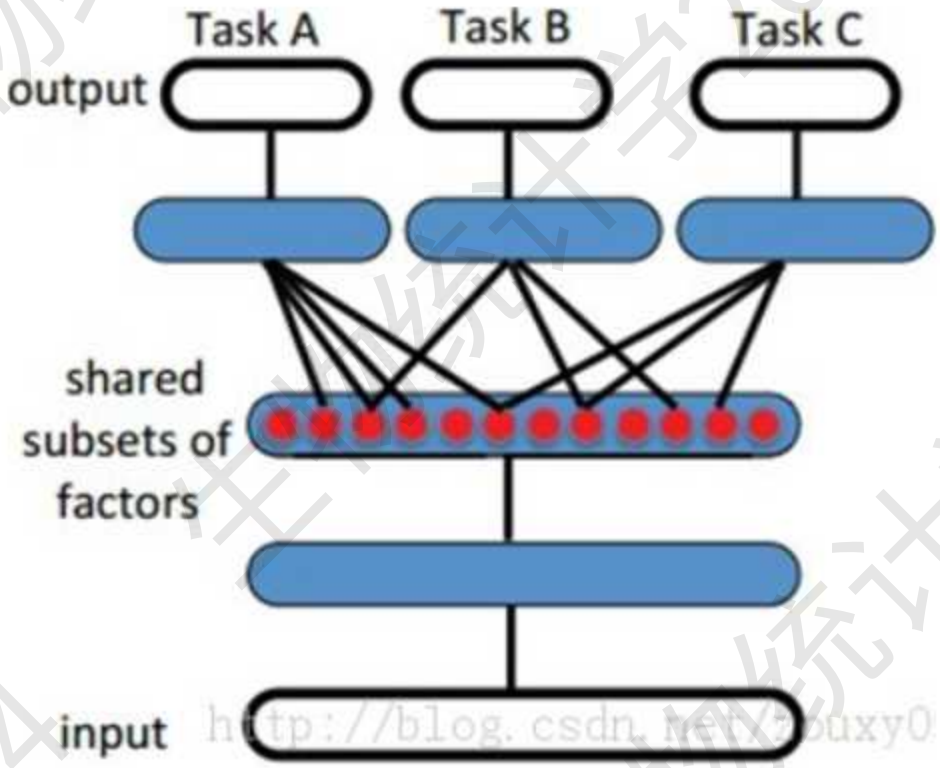
- 深度学习在多模态学习中的应用



生物统计学2024

深度学习的应用

- 深度学习在多任务学习中的应用



深度学习的应用

- 深度学习在多任务学习中的应用

✓ 在深度学习模型中，对于相关任务的联合学习，往往会取得较好的特征表达；

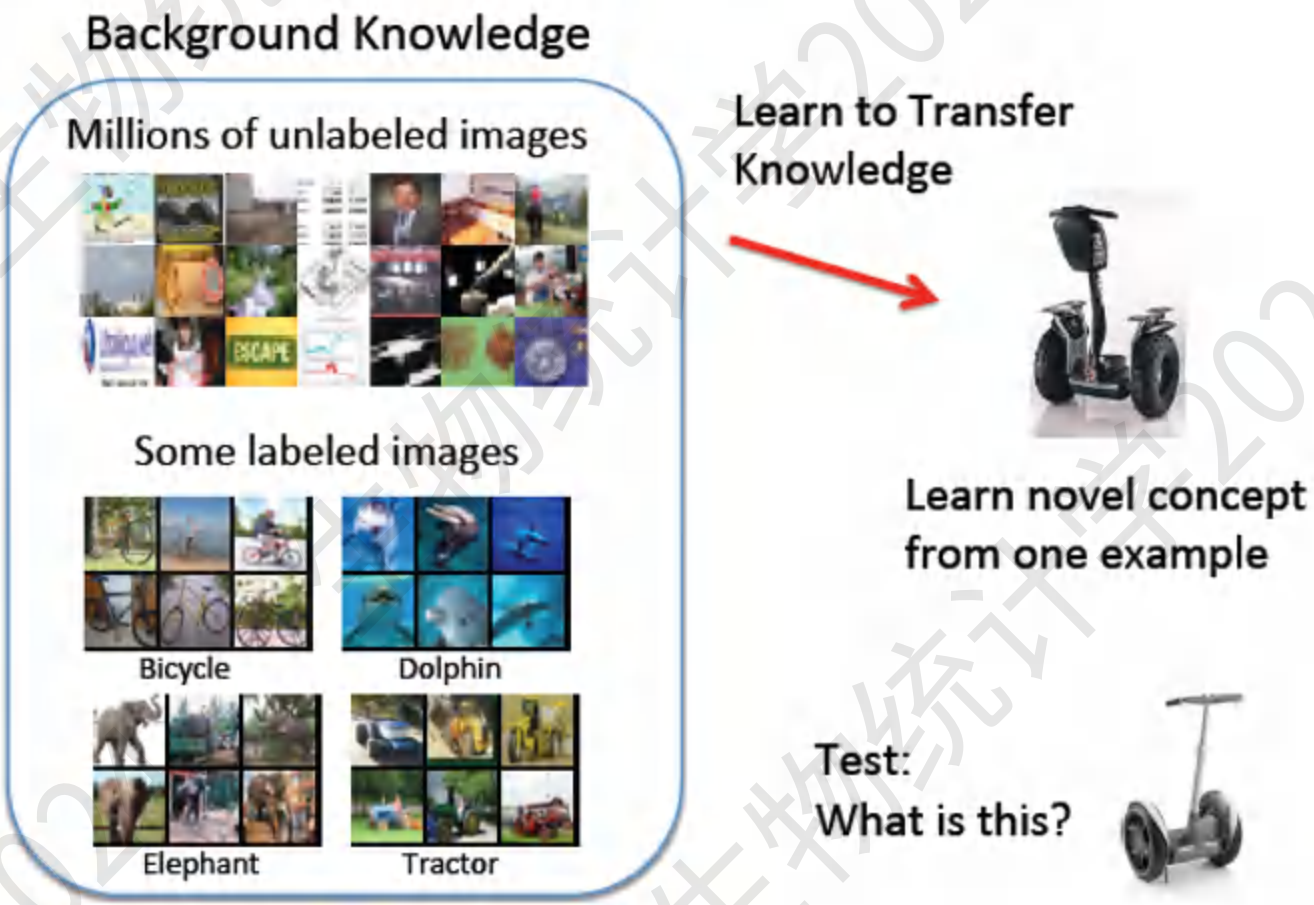
✓ 多任务联合学习，能够增强损失函数的作用效能；

比如：单独进行人脸检测会比较难（光照、遮挡等因素），但是当人脸检测与人脸识别这两个相关的任务联合学习时，人脸检测的难度反而降低了。

生物统计学2024

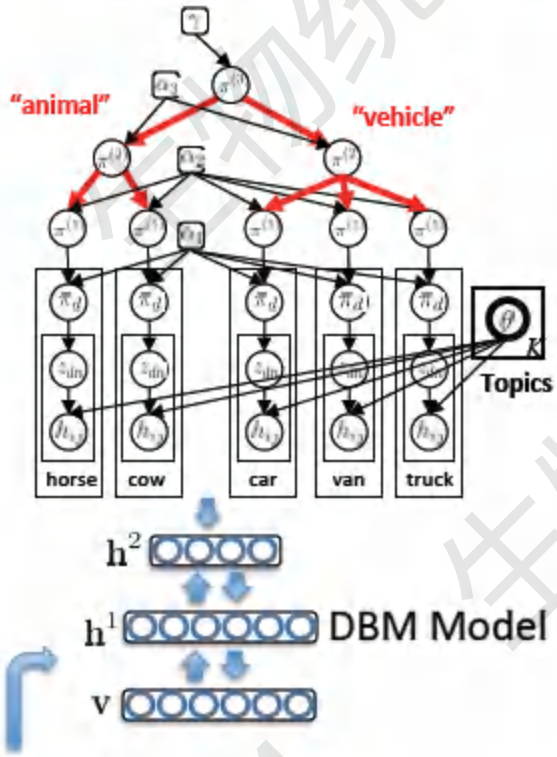
深度学习的应用

- 基于深度学习的迁移学习应用

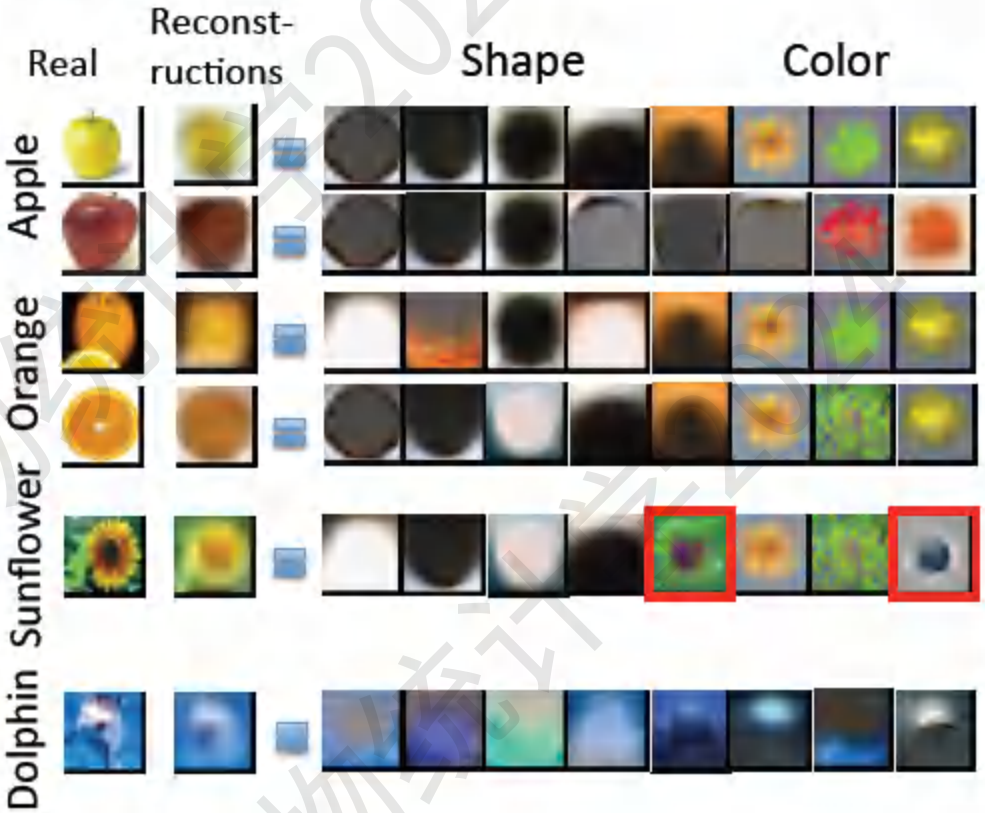


深度学习的应用

- 基于深度学习的迁移学习应用

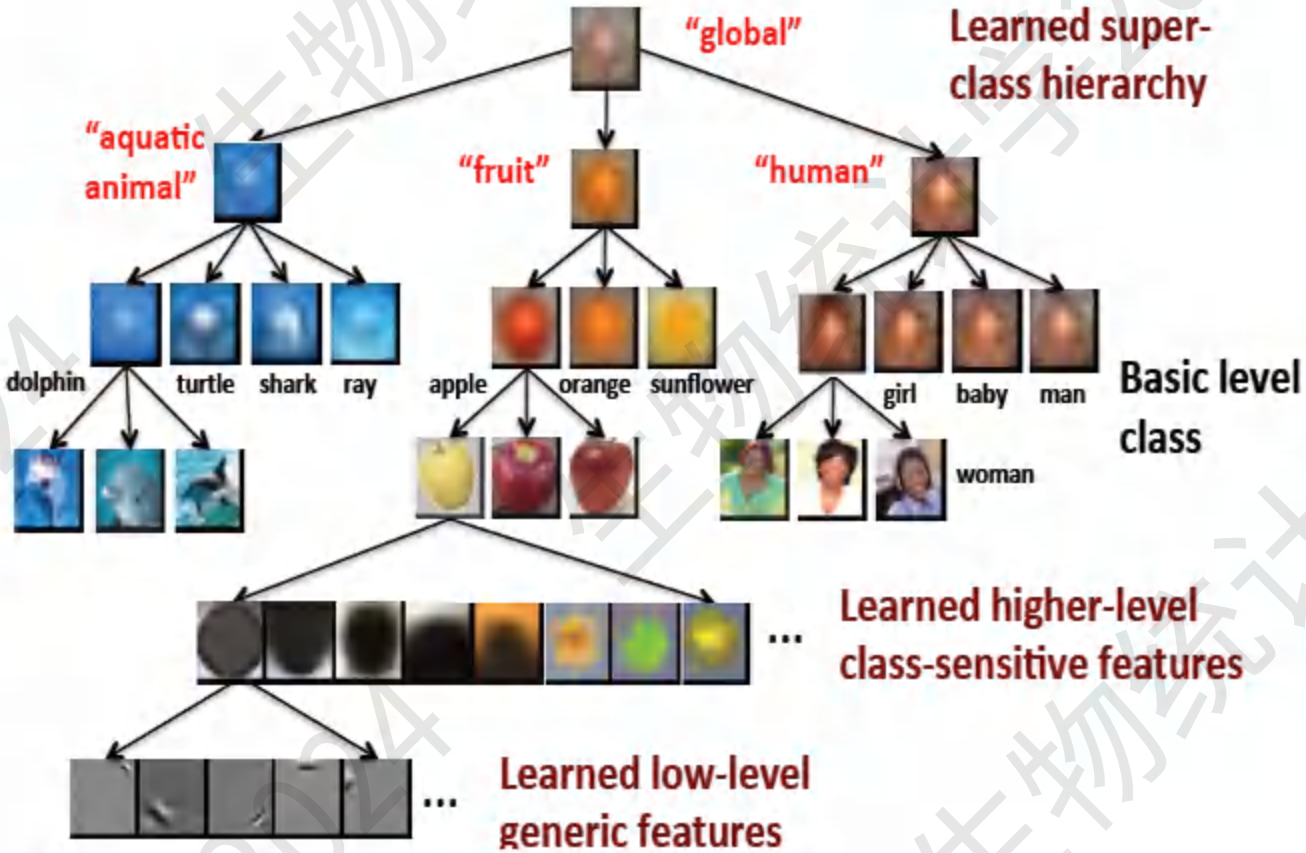


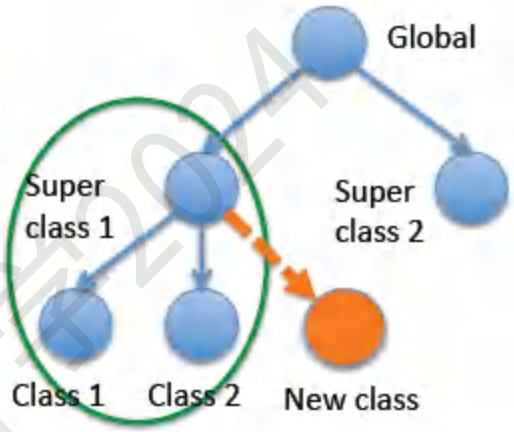
Low-level features:
replace GIST, SIFT



深度学习的应用

- 基于深度学习的迁移学习应用





生物统计学2024

深度学习的应用

- 深度学习在大尺度数据集上的应用

□大尺度数据集：

✓样本总数>100M,

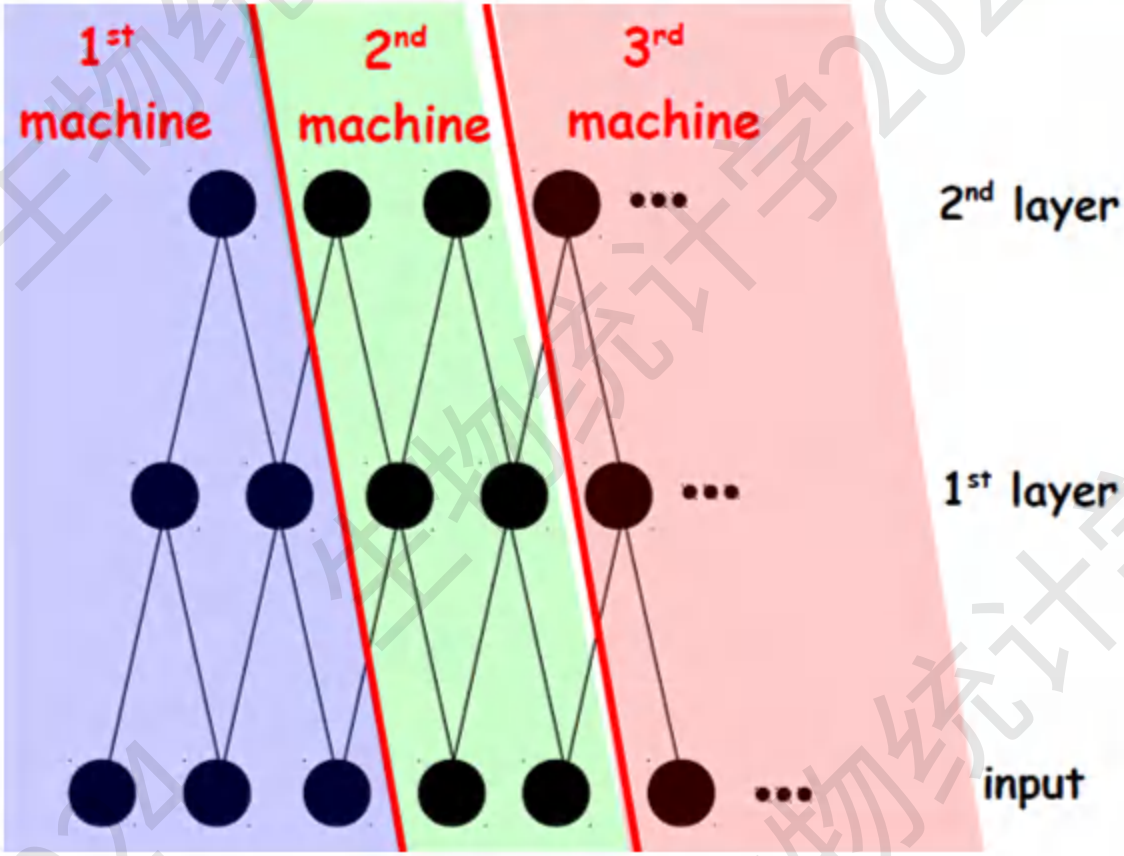
✓类别总数>10K,

✓特征维度>10K

生物统计学2024

深度学习的应用

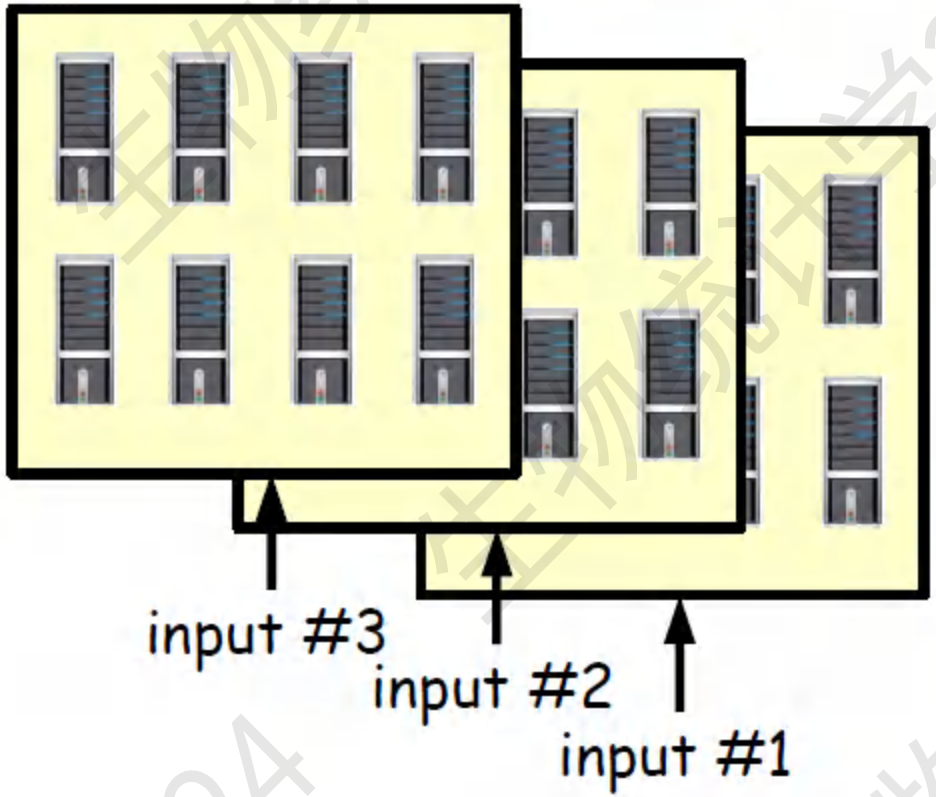
- 深度学习在大尺度数据集上的应用



生物统计学2024

深度学习的应用

- 深度学习在大尺度数据集上的应用



MODEL PARALLELISM

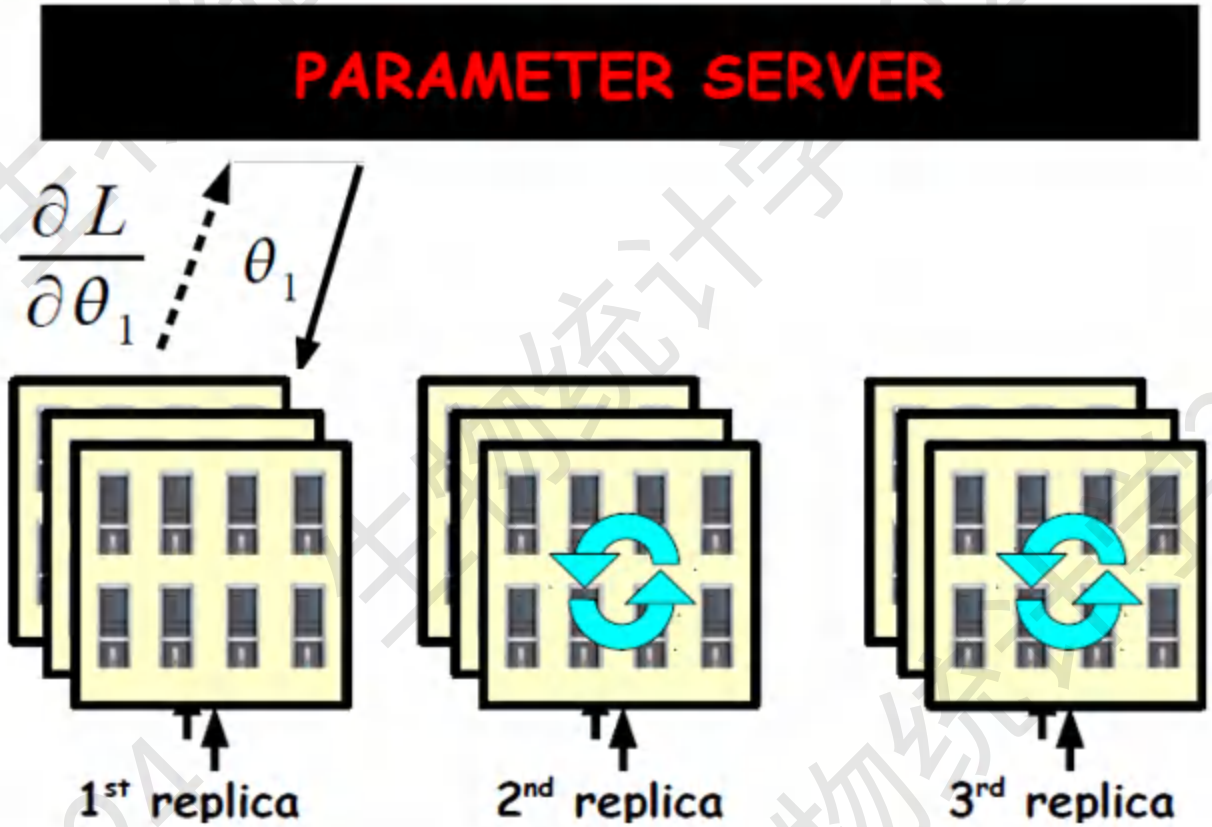
+

DATA PARALLELISM

生物统计学2024

深度学习的应用

- 深度学习在大尺度数据集上的应用



生物统计学2024

深度学习的应用

- 深度学习在大尺度数据集上的应用

IMAGENET v.2011 (16M images, 20K categories)

METHOD	ACCURACY %
Weston & Bengio 2011	9.3
Linear Classifier on deep features	13.1
Deep Net (from random)	13.6
Deep Net (from unsup.)	15.8

生物统计学2024

深度学习的应用

- 深度学习的**State-of-the-art**

Images

CIFAR Object classification	Accuracy
Prior art (Ciresan et al., 2011)	80.5%
Stanford Feature learning	**82.0%**

NORB Object classification	Accuracy
Prior art (Scherer et al., 2010)	94.4%
Stanford Feature learning	**95.0%**

Video

Hollywood2 Classification	Accuracy
Prior art (Laptev et al., 2004)	48%
Stanford Feature learning	**53%**

YouTube	Accuracy
Prior art (Liu et al., 2009)	71.2%
Stanford Feature learning	**75.8%**

KTH	Accuracy
Prior art (Wang et al., 2010)	92.1%
Stanford Feature learning	**93.9%**

UCF	Accuracy
Prior art (Wang et al., 2010)	85.6%
Stanford Feature learning	**86.5%**

Text/NLP

Paraphrase detection	Accuracy
Prior art (Das & Smith, 2009)	76.1%
Stanford Feature learning	**76.4%**

Sentiment (MR/MPQA data)	Accuracy
Prior art (Nakagawa et al., 2010)	77.3%
Stanford Feature learning	**77.7%**

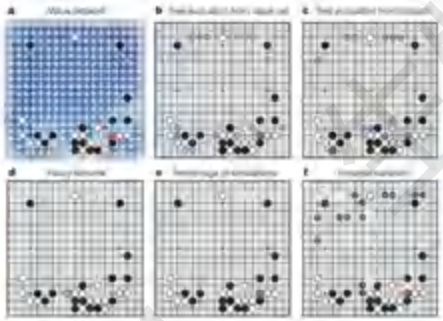
Multimodal (audio/video)

AVLetters Lip reading	Accuracy
Prior art (Zhao et al., 2009)	58.9%
Stanford Feature learning	**65.8%**

Other unsupervised feature learning records:
Pedestrian detection (Yann LeCun)
Speech recognition (Geoff Hinton)
PASCAL VOC object classification (Kai Yu)

Andrew

AlphaGo



The input to the policy network is a 19 × 19 × 48 image stack consisting of 48 feature planes. The first hidden layer zero pads the input into a 23 × 23 image, then convolves *k* filters of kernel size 5 × 5 with stride 1 with the input image and applies a rectifier nonlinearity. Each of the subsequent hidden layers 2 to 12 zero pads the respective previous hidden layer into a 21 × 21 image, then convolves *k* filters of kernel size 3 × 3 with stride 1, again followed by a rectifier nonlinearity. The final layer convolves 1 filter of kernel size 1 × 1 with stride 1, with a different bias for each position, and applies a softmax function. The match version of AlphaGo used *k* = 192 filters; Fig. 2b and Extended Data Table 3 additionally show the results of training with *k* = 128, 256 and 384 filters.

policy network:
[19x19x48] Input
CONV1: 192 5x5 filters , stride 1, pad 2 => [19x19x192]
CONV2..12: 192 3x3 filters, stride 1, pad 1 => [19x19x192]
CONV: 1 1x1 filter, stride 1, pad 0 => [19x19] *(probability map of promising moves)*

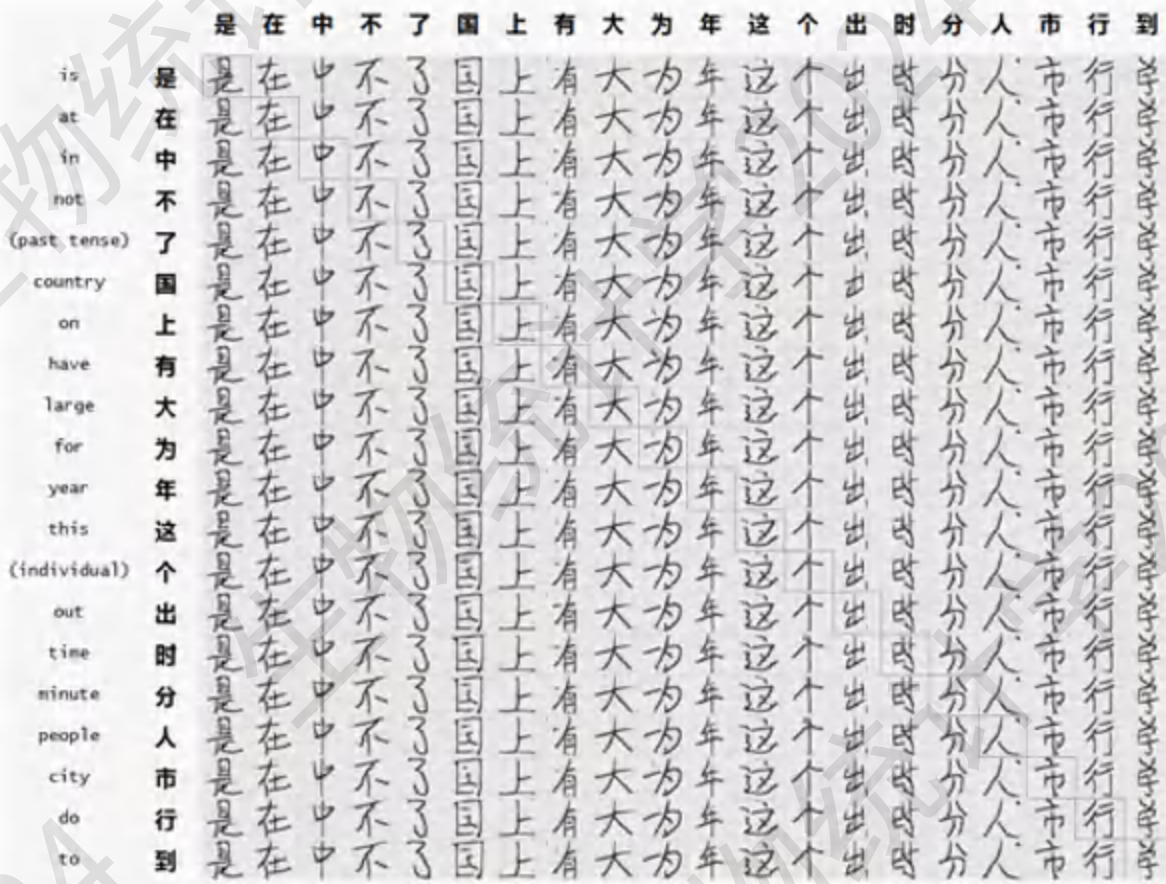
- 单机版：48 个CPU 和8 块GPU
- 走子速度：3 毫秒-2 微秒

生物统计学2024

Mask RCNN

Figure 4. More results of **Mask R-CNN** on COCO test images, using ResNet-101-FPN and running at 5 fps, with 35.7 mask AP (Table 1).

图像生成



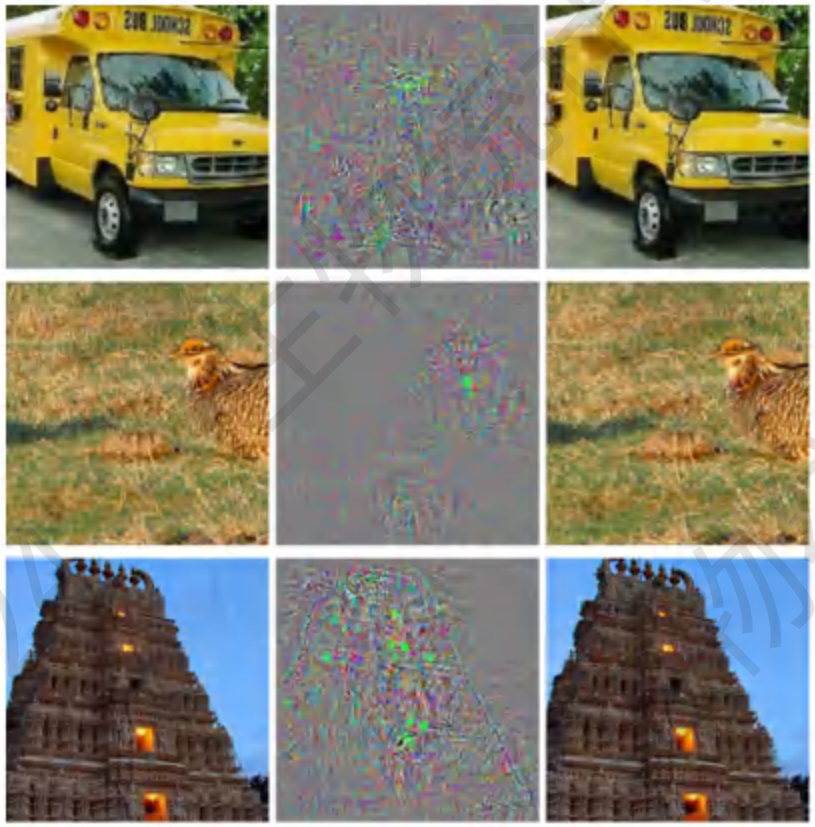
生物统计学2024

Deep Dream

生物统计学2024

画风迁移

对抗样本



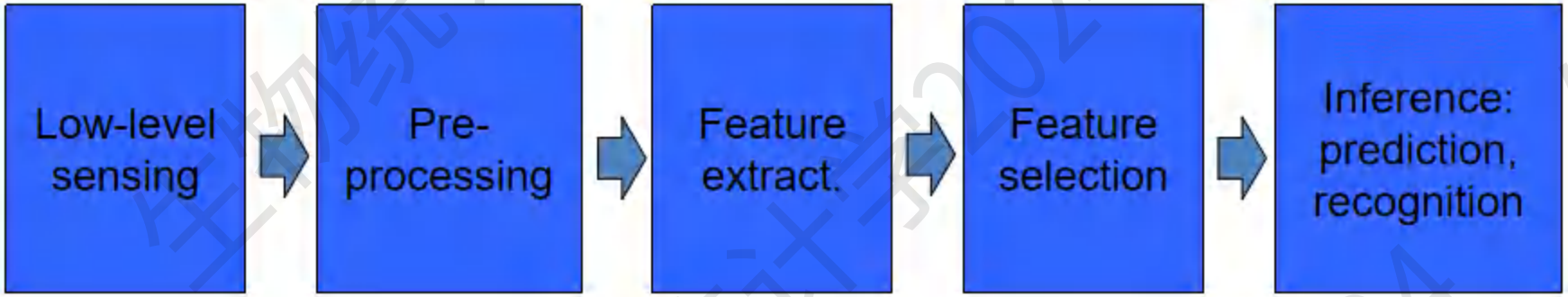
(a)

(b)

生物统计学2024

动 机

传统的模式识别方法：



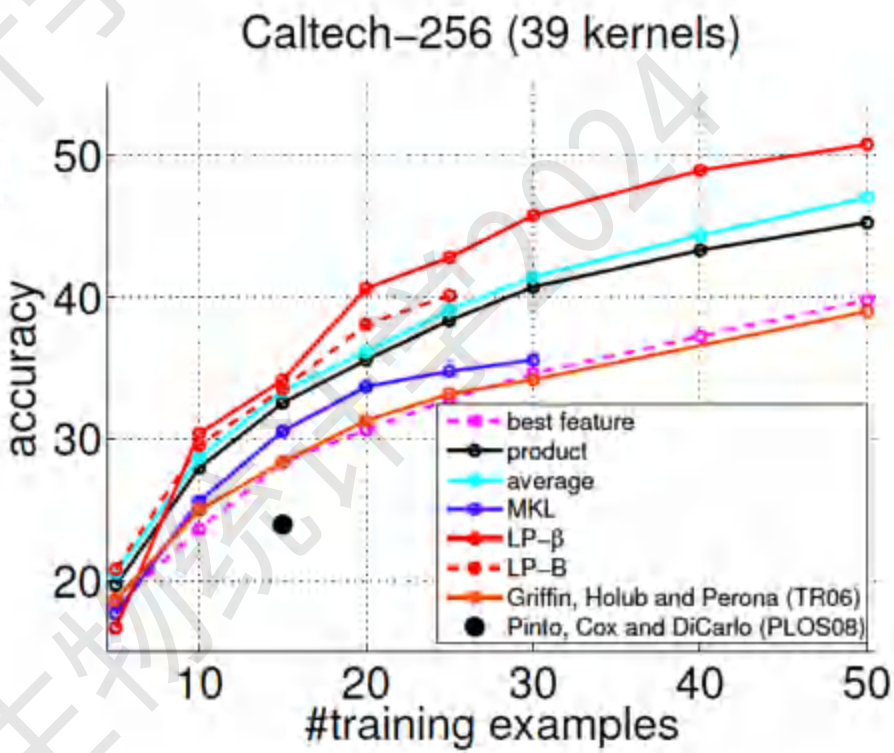
- 良好的特征表达，对最终算法的准确性起了非常关键的作用；
- 识别系统主要的计算和测试工作耗时主要集中在特征提取部分；
- 特征的样式目前一般都是人工设计的，靠人工提取特征。

生物统计学2024

动 机——为什么要自动学习特征

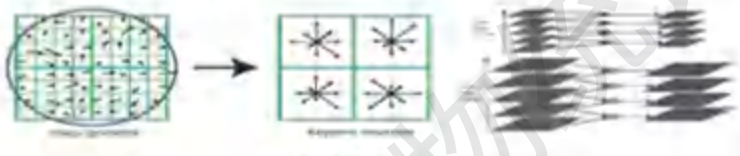
- 实验：LP-β Multiple Kernel Learning
  - Gehler and Nowozin, On Feature Combination for Multiclass Object Classification, ICCV'09
- 采用39 个不同的特征
  - PHOG, SIFT, V1S+, Region Cov. Etc.
- 在普通特征上MKL表现有限

结论：特征更重要



动 机——为什么要自动学习特征

- 机器学习中，获得好的特征是识别成功的关键
- 目前存在大量人工设计的特征，不同研究对象特征不同，特征具有多样性，如：SIFT, HOG, LBP等
- 手工选取特征费时费力，需要启发式专业知识，很大程度上靠经验和运气
- 是否能自动地学习特征？

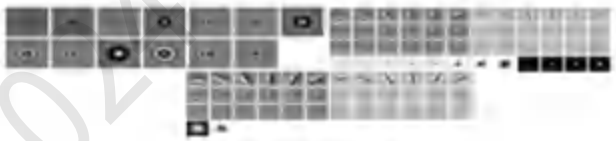


SIFT

Spin image

HoG

RIFT



Textons

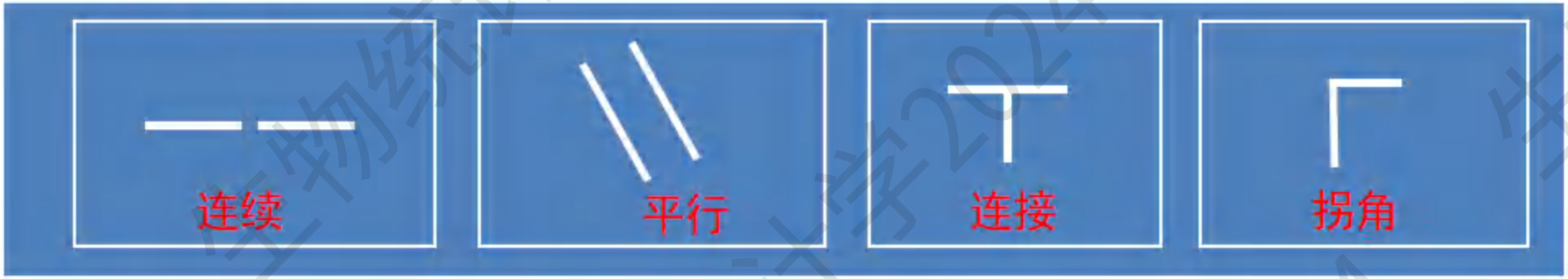
GLOH

生物统计学2024

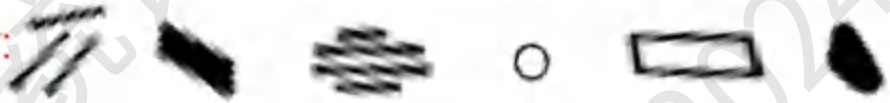
动 机——为什么要自动学习特征

- 中层特征

✓ 中层信号：



“Tokens” from Vision by D.Marr:

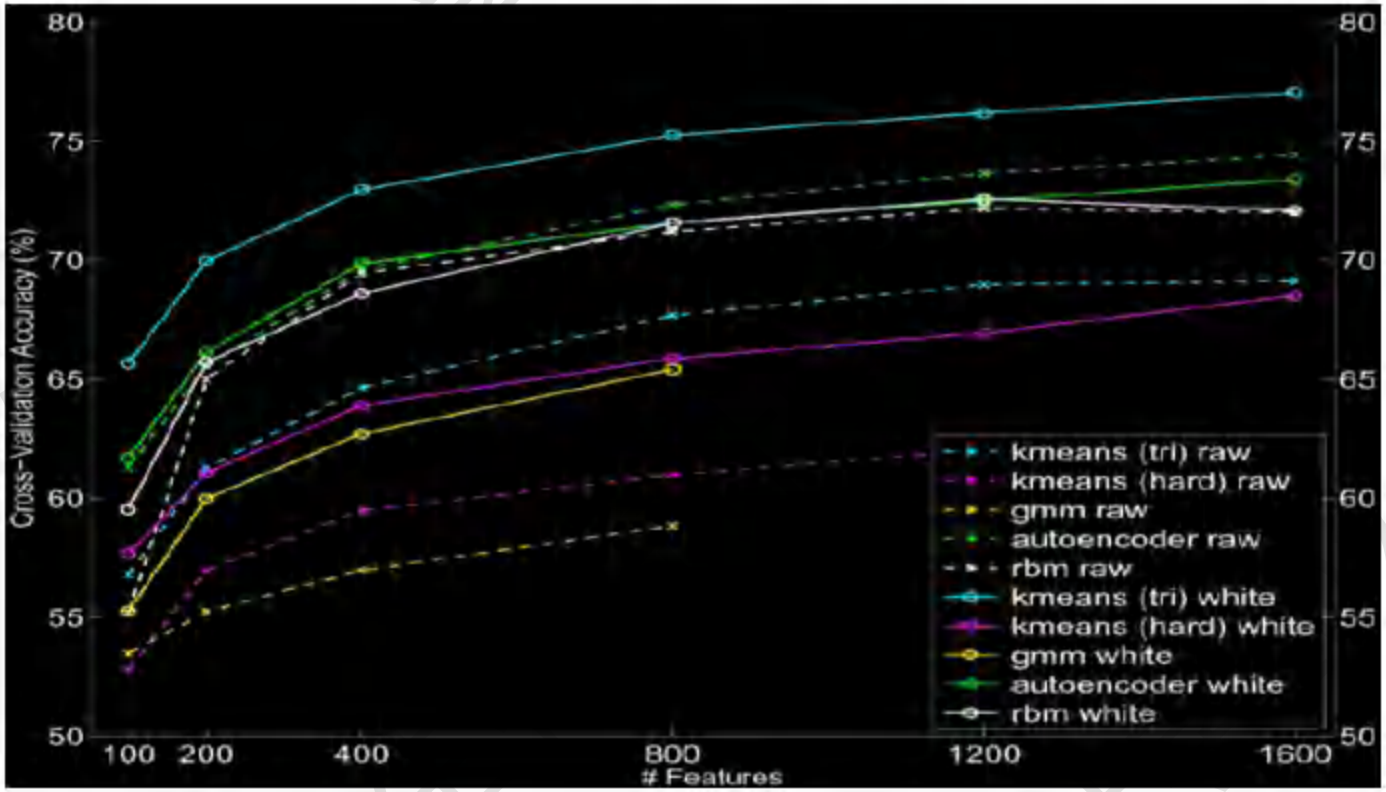


✓ 物体部件:

- 他们对于人工而言是十分困难的，那么如何学习呢？

生物统计学2024

动 机——为什么要自动学习特征



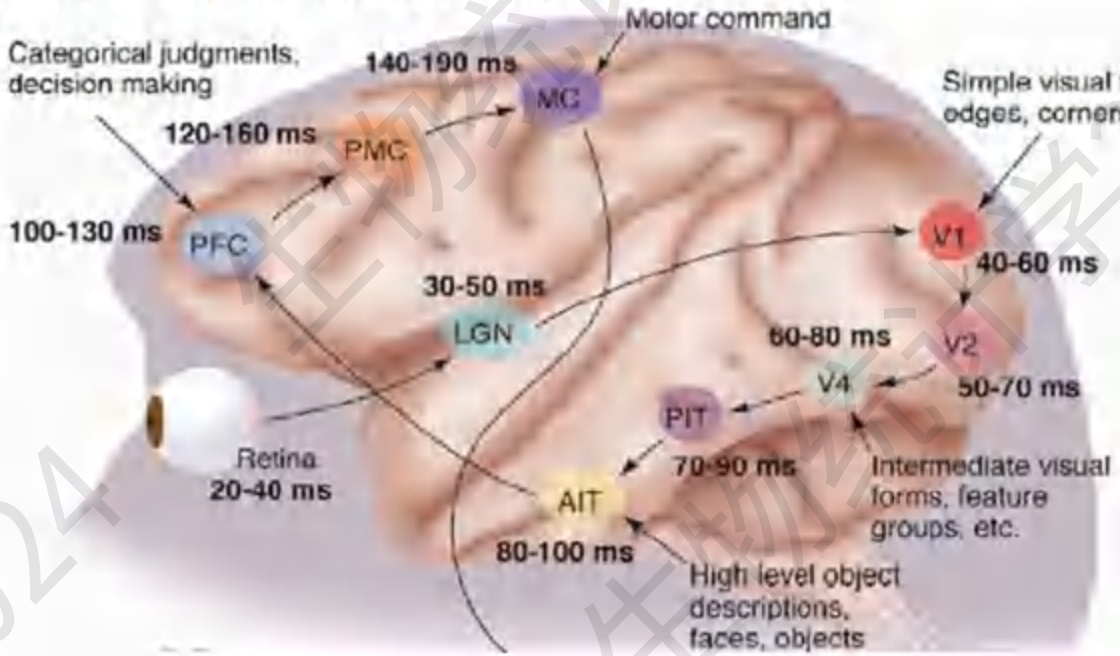
- 一般而言，特征越多，给出信息就越多，识别准确性会得到提升；
- 但特征多，计算复杂度增加，探索的空间大，可以用来训练的数据在每个特征上就会稀疏。
- 结论：不一定特征越多越好！需要有多少个特征，需要学习确定。

动 机——为什么采用层次网络结构

- **人脑视觉机理**

✓ 1981年的诺贝尔医学奖获得者 David Hubel和TorstenWiesel发现了视觉系统的信息处理机制

✓ 发现了一种被称为“方向选择性细胞的神经元细胞，当瞳孔发现了眼前的物体的边缘，而且这个边缘指向某个方向时，这种神经元细胞就会活跃



动 机——为什么采用层次网络结构

- **人脑视觉机理**

✓ 人的视觉系统的信息处理是分级的

✓ 高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义或者意图

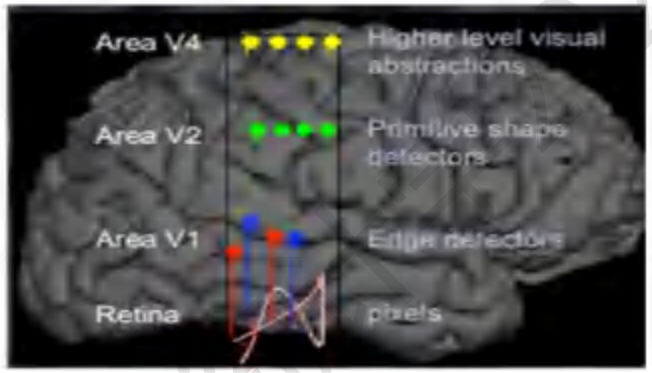
✓ 抽象层面越高，存在的可能猜测就越少，就越利于分类

object models

object parts (combination of edges)

edges

pixels



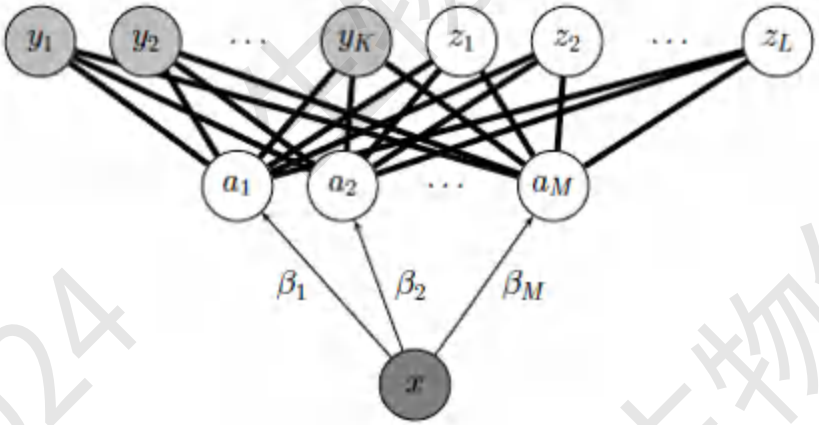
生物统计学2024

动 机——为什么采用层次网络结构

- 视觉的层次性

✓ 属性学习，类别作为属性的一种组合映射

Lampert et al. CVPR'09



otter

black: yes
white: no
brown: yes
stripes: no
water: yes
eats fish: yes

polar bear

black: no
white: yes
brown: no
stripes: no
water: yes
eats fish: yes

zebra

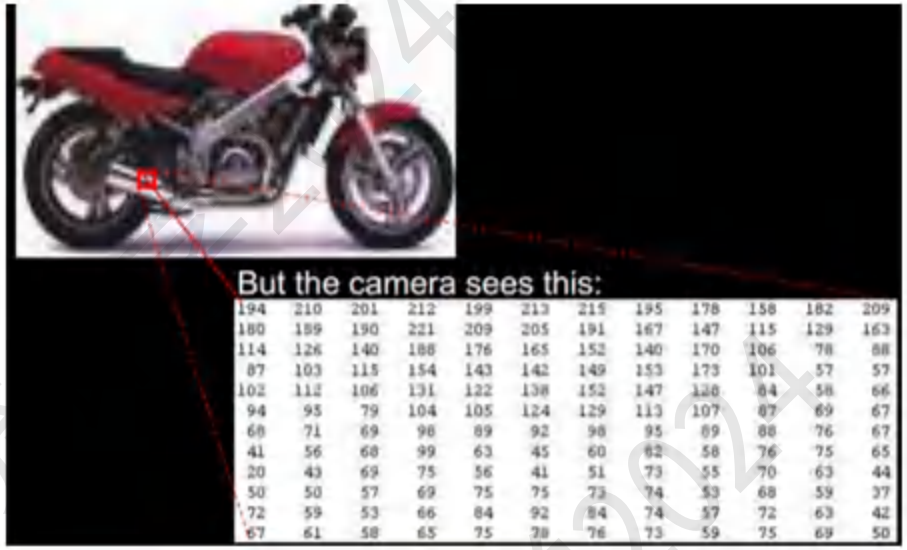
black: yes
white: yes
brown: no
stripes: yes
water: no
eats fish: no

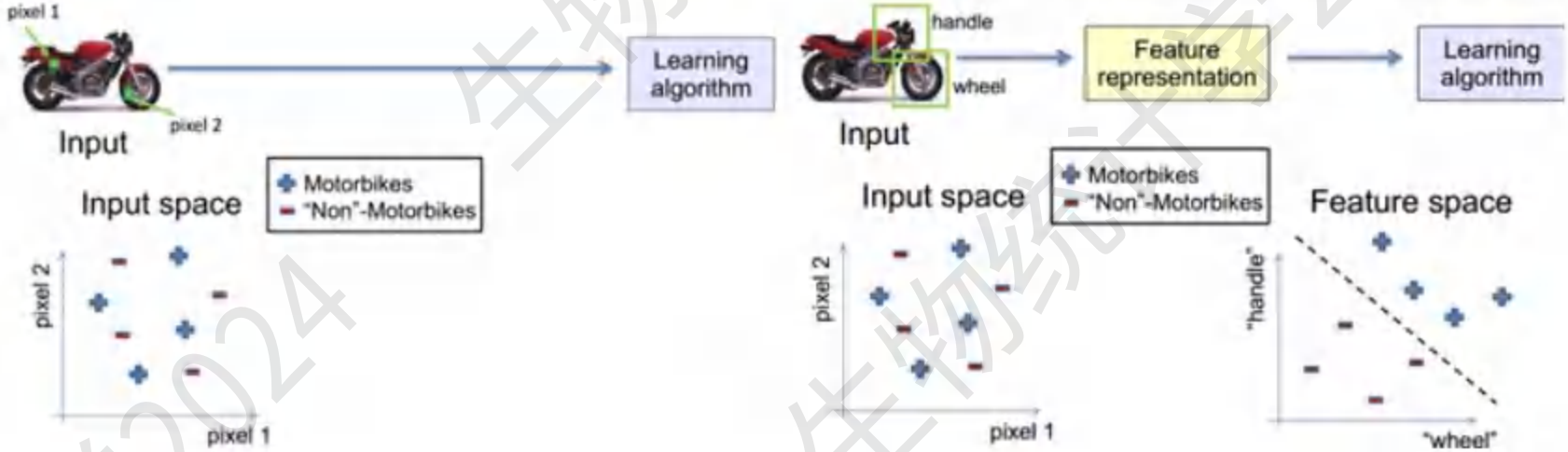
生物统计学2024

动 机——为什么采用层次网络结构

- 特征表示的粒度

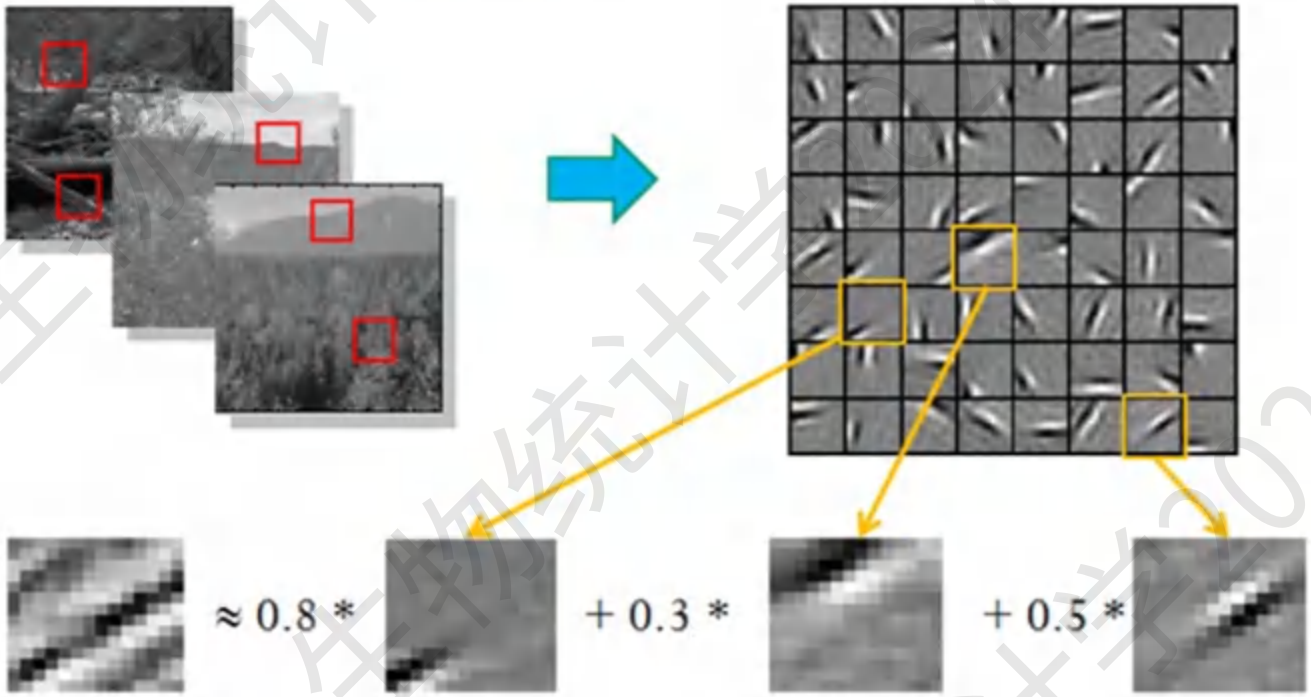
✓ 具有结构性（或者语义）的高层特征对于分类更有意义





动 机——为什么采用层次网络结构

- 初级（浅层）特征表示

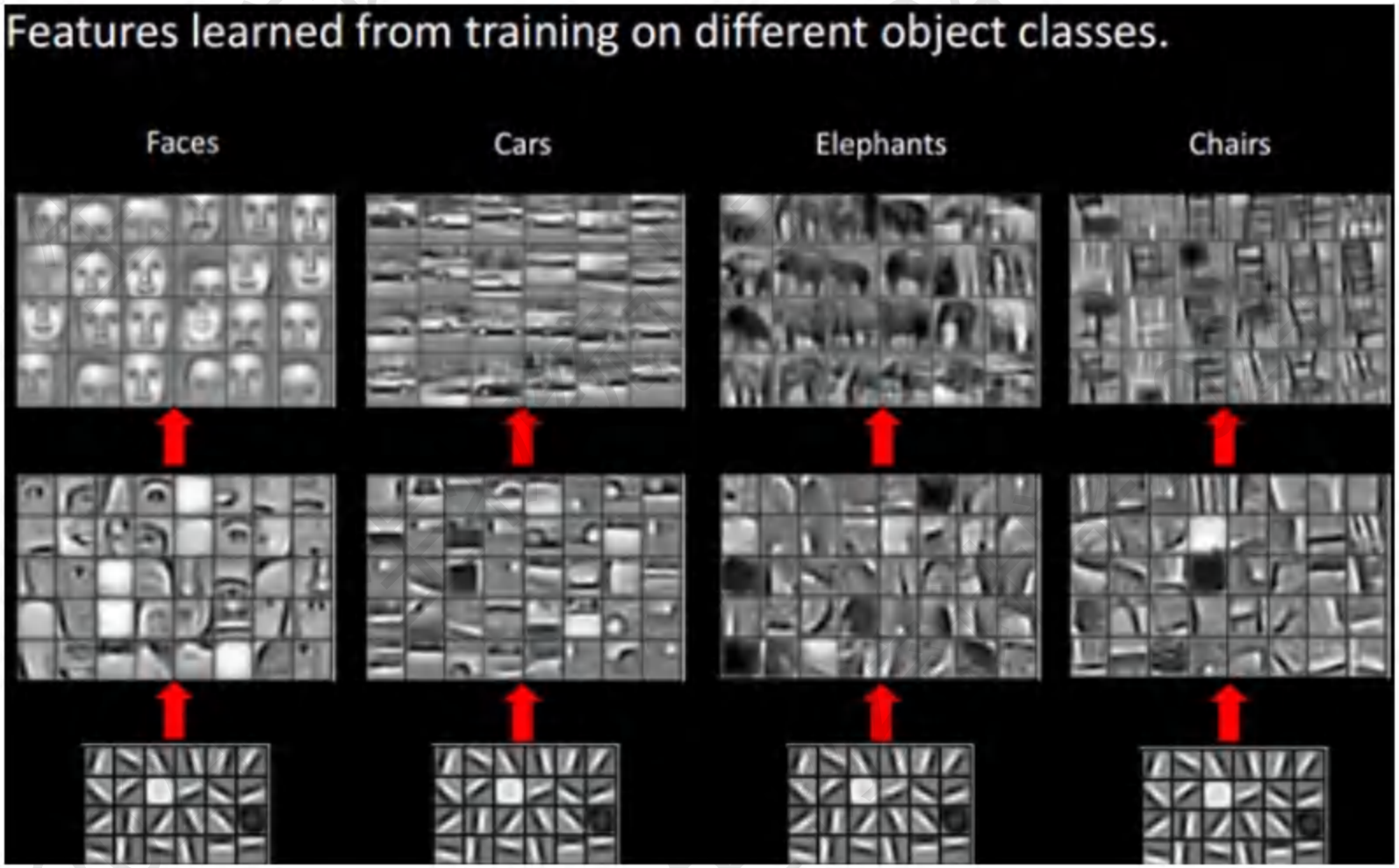


$[a_1, \ldots, a_{64}] = [0, 0, \ldots, 0, \mathbf{0.8}, 0, \ldots, 0, \mathbf{0.3}, 0, \ldots, 0, \mathbf{0.5}, 0]$
(feature representation)

生物统计学2024

动 机——为什么采用层次网络结构

- 结构性特征表示



动 机——为什么采用层次网络结构

- 浅层学习的局限

✓ 人工神经网络（BP算法）

—虽被称作多层感知机，但实际是种只含有一层隐层节点的浅层模型

✓ SVM、Boosting、最大熵方法（如LR，Logistic Regression）

—带有一层隐层节点（如SVM、Boosting），或没有隐层节点（如LR）的浅层模型

局限性：有限样本和计算单元情况下对复杂函数的表示能力有限，针对复杂分类问题其泛化能力受限。

生物统计学2024

深度学习

- **2006年，加拿大多伦多大学教授、机器学习领域的泰斗Geoffrey Hinton在《科学》上发表论文提出深度学习主要观点：**

  1）多隐层的人工神经网络具有优异的特征学习能力，学习得到的特征对数据有更本质的刻画，从而有利于可视化或分类；

  2）深度神经网络在训练上的难度，可以通过“逐层初始化”（layer-wise pre-training）来有效克服，逐层初始化可通过无监督学习实现的。

生物统计学2024

深度学习

- 本质：通过构建多隐层的模型和海量训练数据（可为无标签数据），来学习更有用的特征，从而最终提升分类或预测的准确性。“深度模型”是手段，“特征学习”是目的。
- 与浅层学习区别：

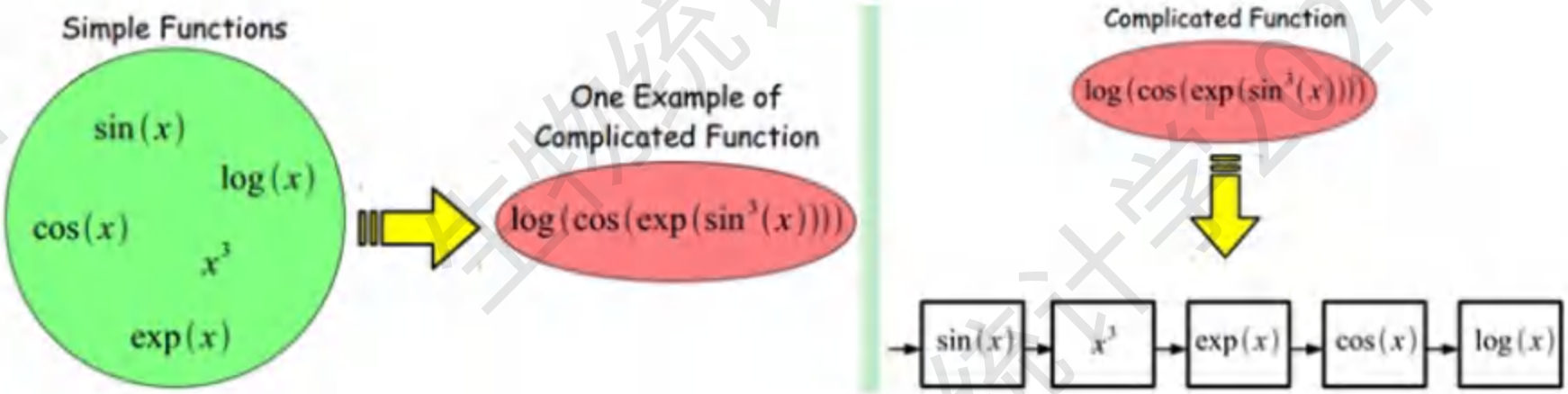
1）强调了模型结构的深度，通常有5-10多层的隐层节点；

2）明确突出了特征学习的重要性，通过逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间，从而使分类或预测更加容易。与人工规则构造特征的方法相比，利用大数据来学习特征，更能够刻画数据的丰富内在信息。

生物统计学2024

深度学习

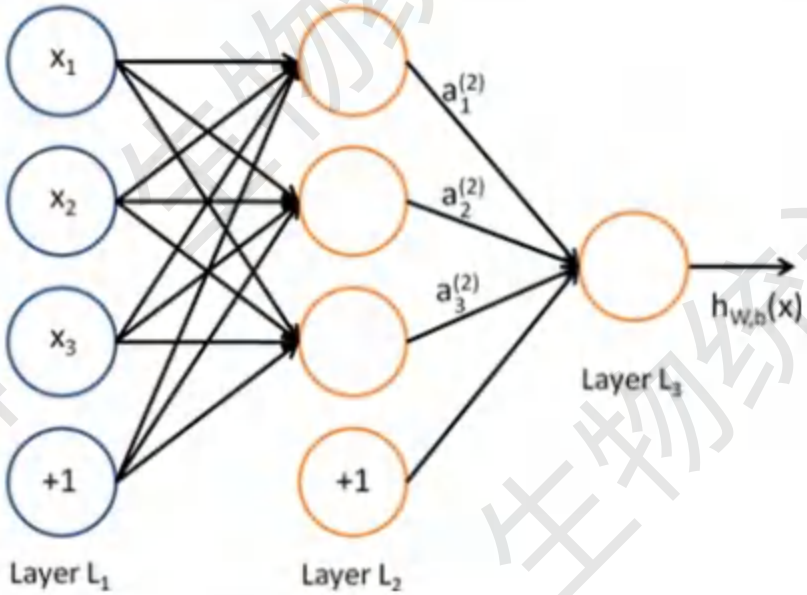
- 好处：可通过学习一种深层非线性网络结构，实现复杂函数逼近，表征输入数据分布式表示。



生物统计学2024

深度学习 vs. 神经网络

神经网络：



深度学习：

含多个隐层的深度学习模型

生物统计学2024

深度学习 vs. 神经网络

相同点： 二者均采用分层结构，系统包括输入层、隐层（多层）、输出层组成的多层网络，只有相邻层节点之间有连接，同一层以及跨层节点之间相互无连接，每一层可以看作是一个logistic 回归模型。

不同点：

神经网络：采用BP算法调整参数，即采用迭代算法来训练整个网络。随机设定初值，计算当前网络的输出，然后根据当前输出和样本真实标签之间的差去改变前面各层的参数，直到收敛；

深度学习：采用逐层训练机制。采用该机制的原因在于如果采用BP机制，对于一个deep network（7层以上），残差传播到最前面的层将变得很小，出现所谓的gradient diffusion（梯度扩散）。

深度学习 vs. 神经网络

- 神经网络的局限性：

1）比较容易过拟合，参数比较难调整，而且需要不少技巧；

2）训练速度比较慢，在层次比较少（小于等于3）的情况下效果并不比其它方法更优；

生物统计学2024

Deep learning

Yoshua Bengio:

Science is NOT a battle, it is a collaboration. We all build on each other's ideas. Science is an act of love, not war. Love for the beauty in the world that surrounds us and love to share and build something together. That makes science a highly satisfying activity, emotionally speaking!

科学不是战争而是合作，任何学科的发展从来都不是一条路走到黑，而是同行之间互相学习、互相借鉴、博采众长、相得益彰，站在巨人的肩膀上不断前行。机器学习的研究也是一样，你死我活那是邪教，开放包容才是正道。

生物统计学2024

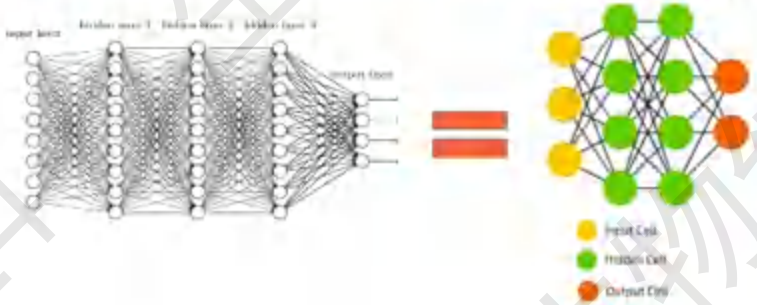
Part II

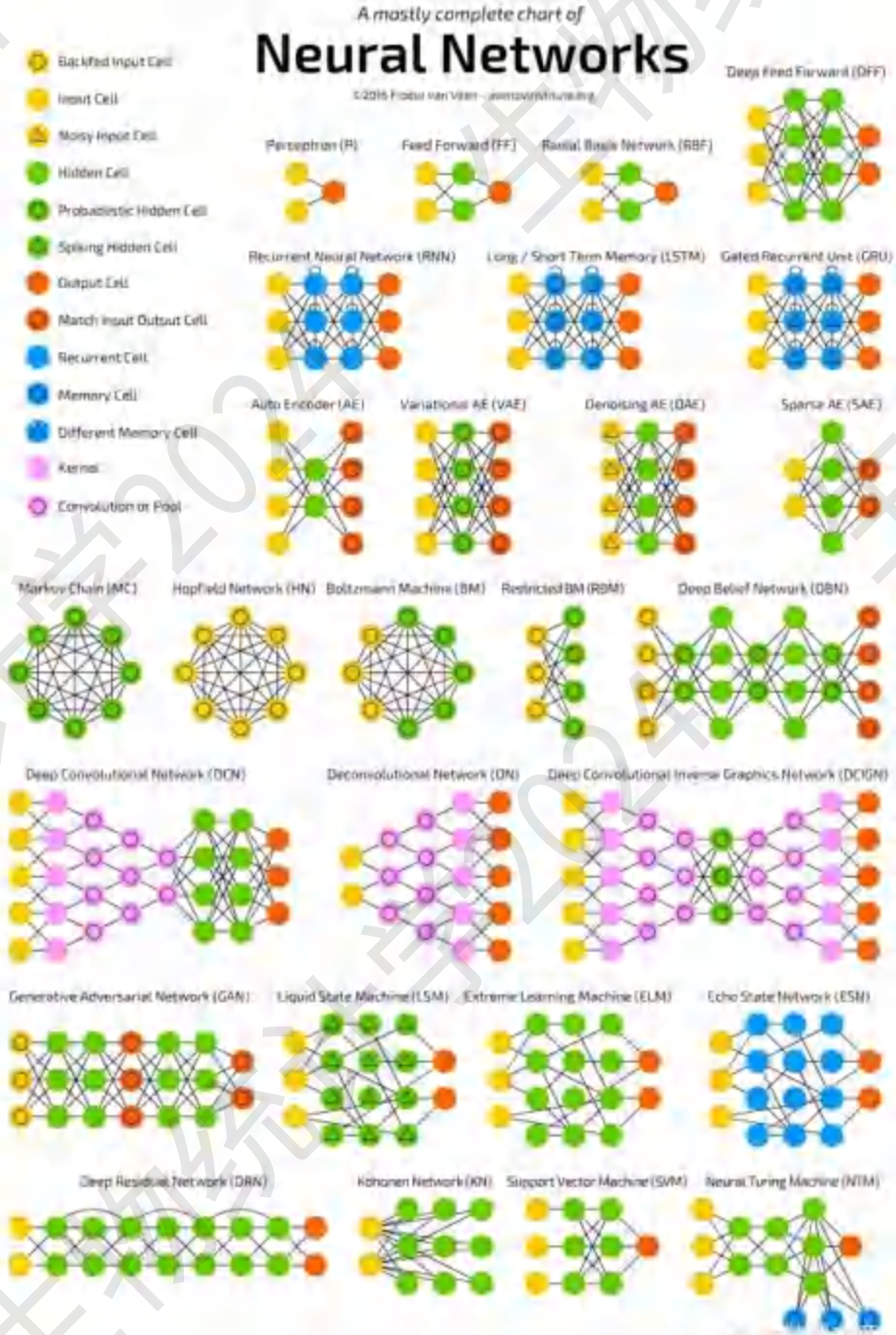
深度学习的基本方法
（以及神经网络算法）

生物统计学2024

Neural network

Deep Neural Network (DNN)

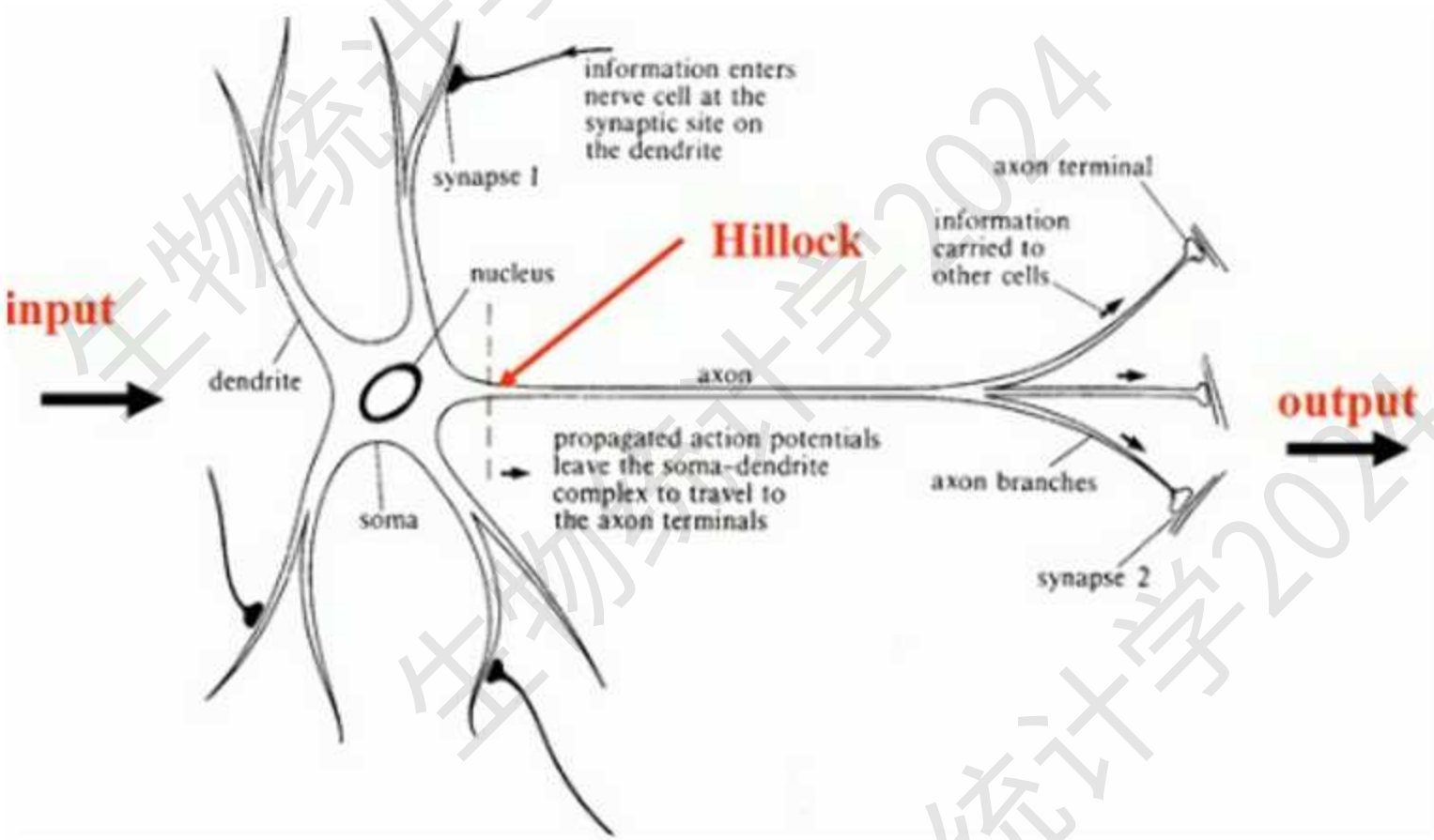




Adapted from: http://www.asimovinstitute.org/neural-network-zoo/

生物统计学2024

Neural network



生物统计学2024

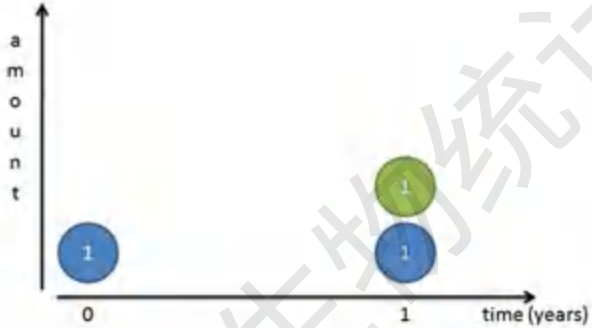
Neural network

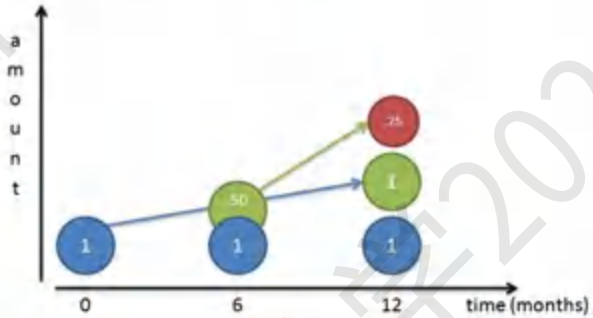
- 自然对数e
- 逻辑回归
- 神经网络
- 反向传播算法(BP算法)
- 示例

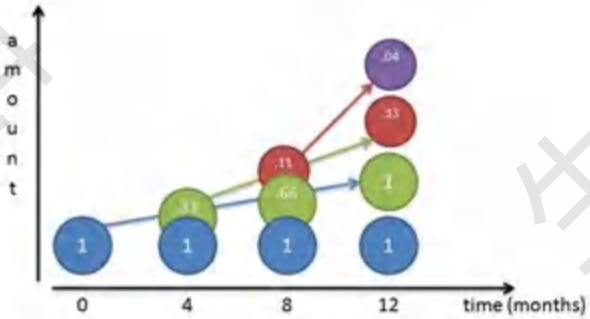
生物统计学2024

Neural network

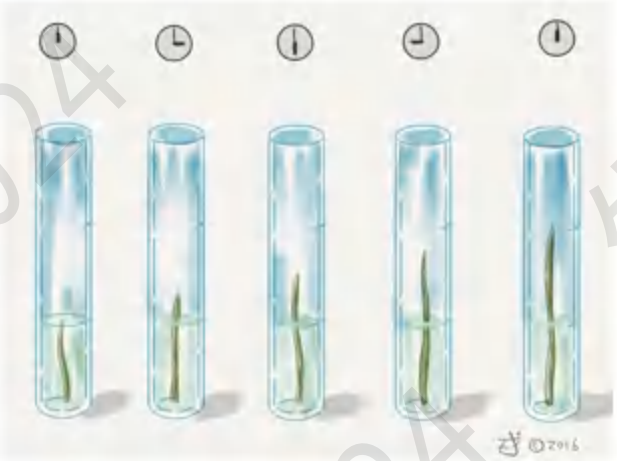
自然对数e





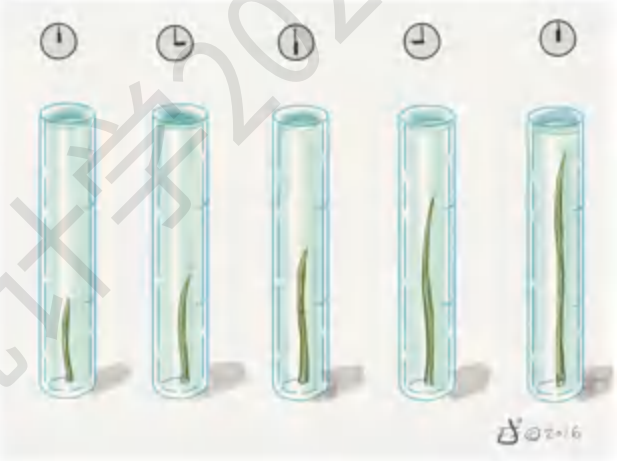


1元存1年，在年利率100%下，无论怎么利滚利，其余额总有一个上限，这个天花板就是e



有一种草齐头泡水半天变2倍

中间加一次水能长到2.25倍



加足水最多能长到2.718...倍

生物统计学2024

Neural network

自然对数e

$$e = 1 + 1/1! + 1/2! + 1/3! + 1/4! + 1/5! + 1/6! + 1/7! + \ldots = 1 + 1 + 1/2 + 1/6 + 1/24 + 1/120 + \ldots \approx 2.71828$$

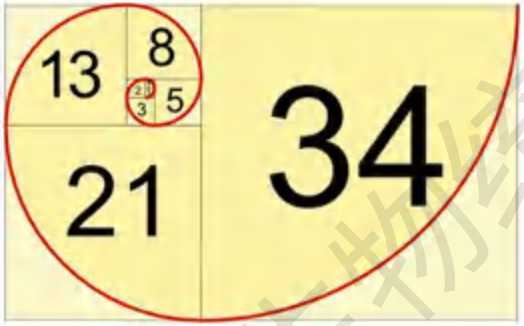
$$e=(1+1/x)^x \approx 2.71828$$

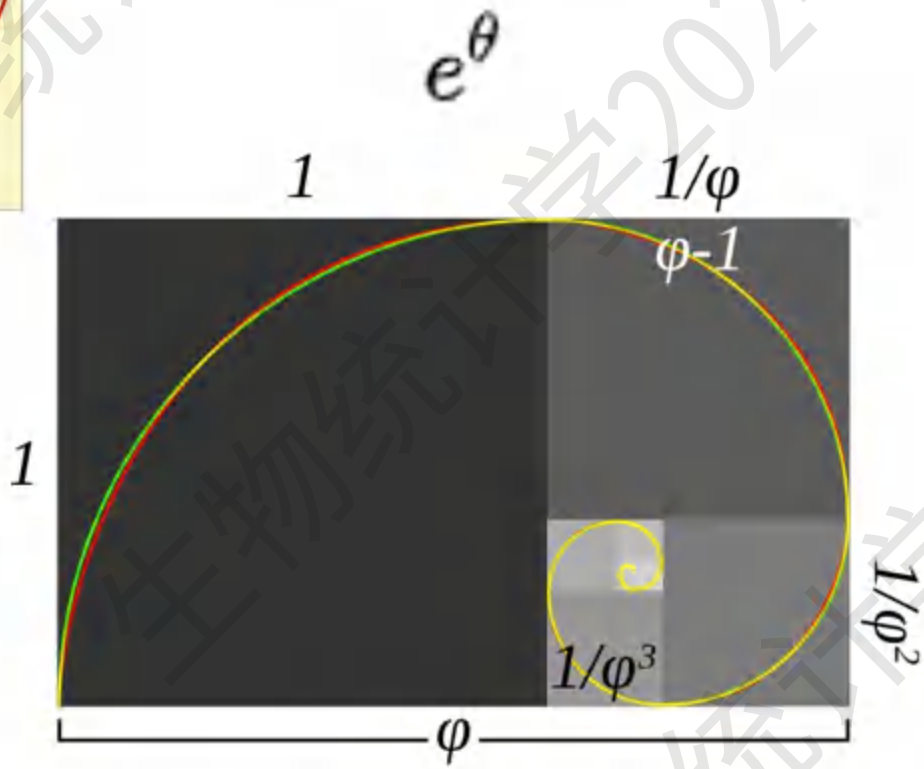
Reference:
https://www.zhihu.com/question/20296247/answer/20370489

生物统计学2024

Neural network

自然对数e vs. 斐波拉切数列

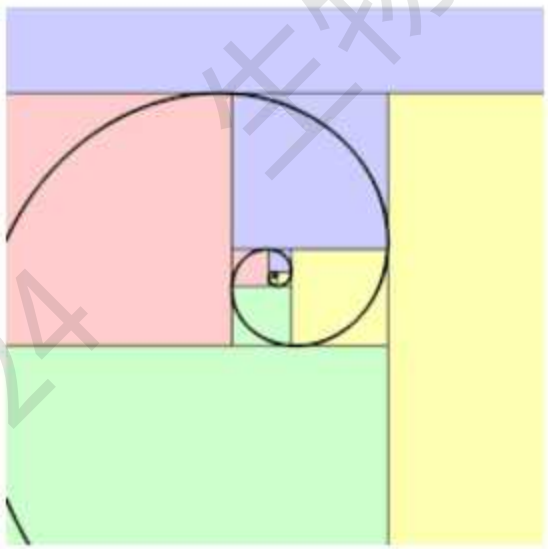




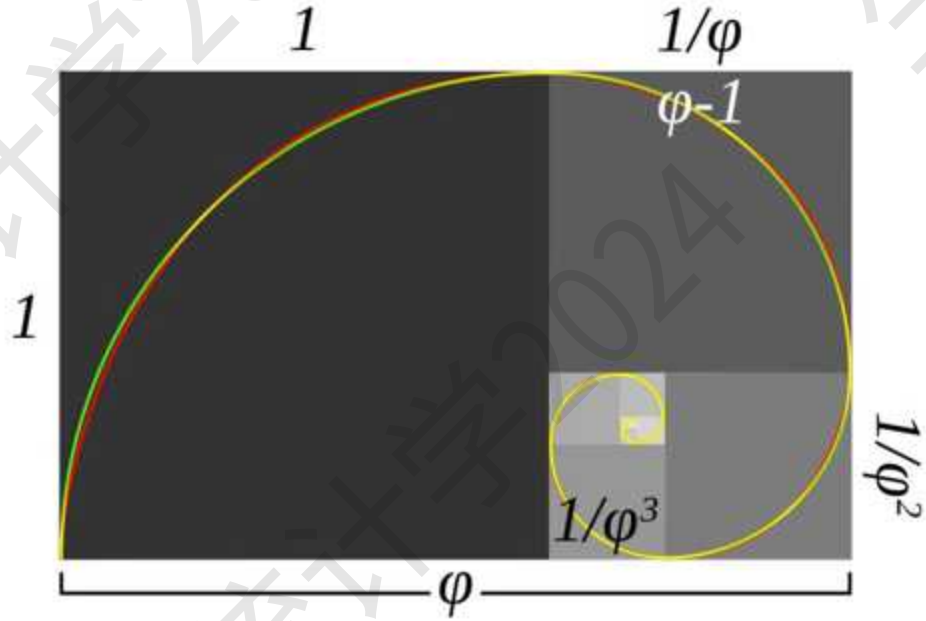
斐波那契螺线仅仅是对一种叫黄金螺线（Golden spiral）的近似，黄金螺线是一种内涵黄金分割比例的对数螺线。
红色的才是黄金曲线，绿色的是“假黄金螺线”（斐波那契螺线），近似却不重合。

Neural network

自然对数e & 黄金分割



e^{θ}

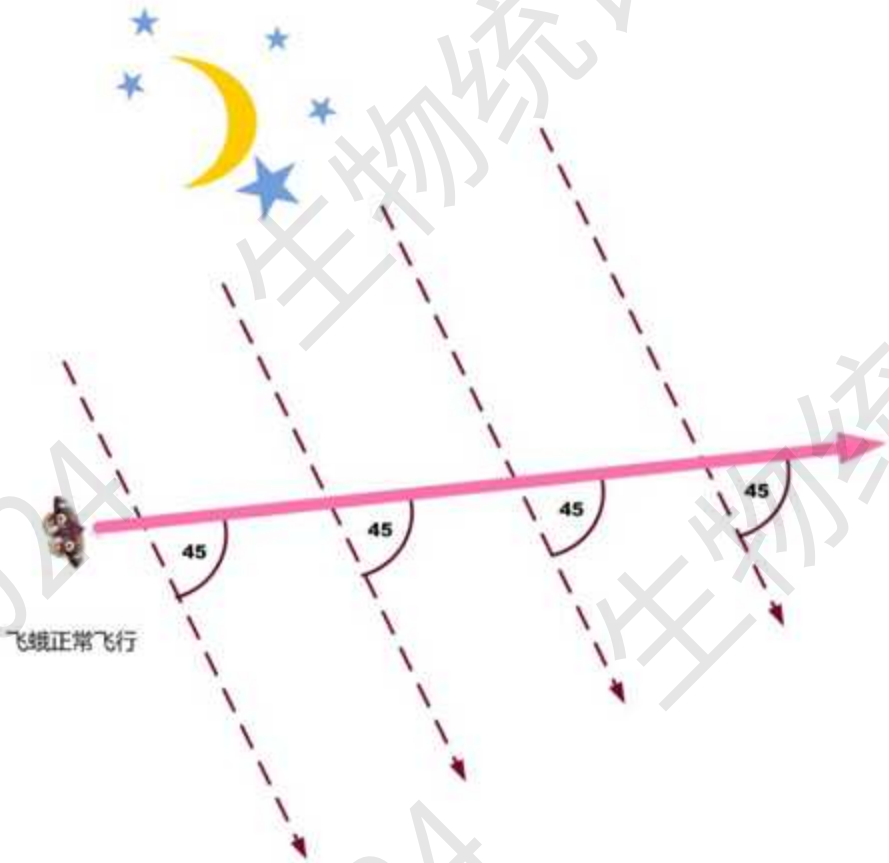


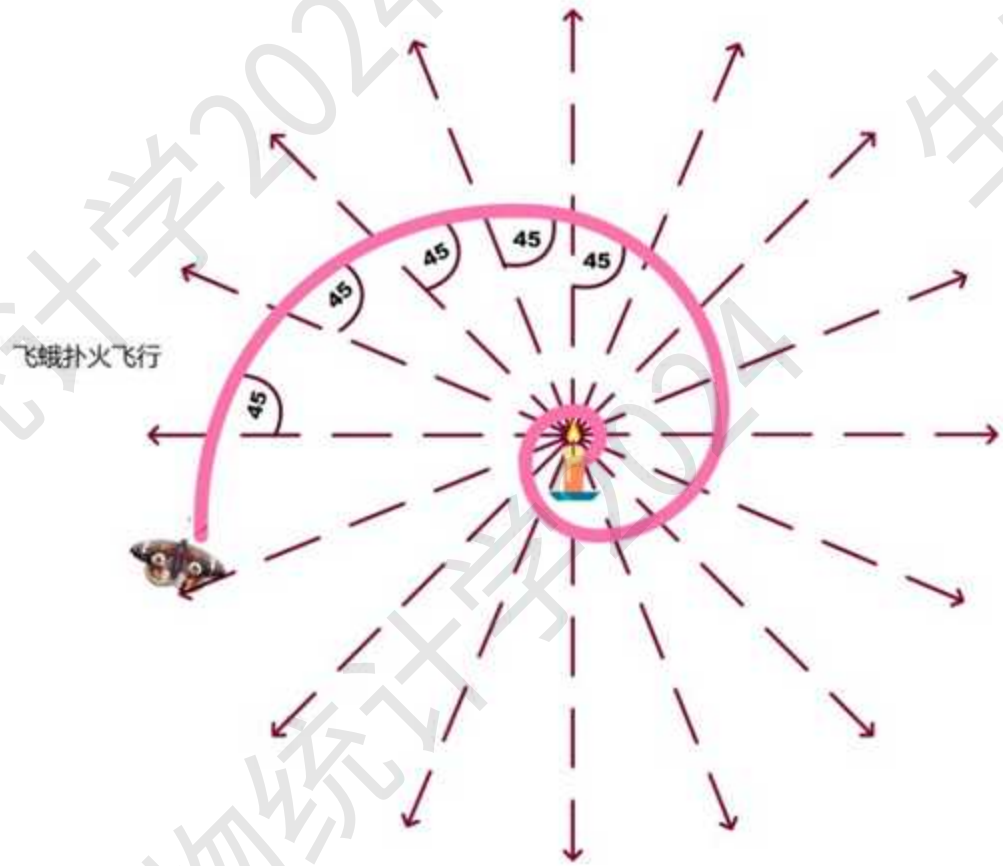
φ=1.6180339887498948482...

生物统计学2024

Neural network

自然对数e





生物统计学2024

Neural network

自然对数e

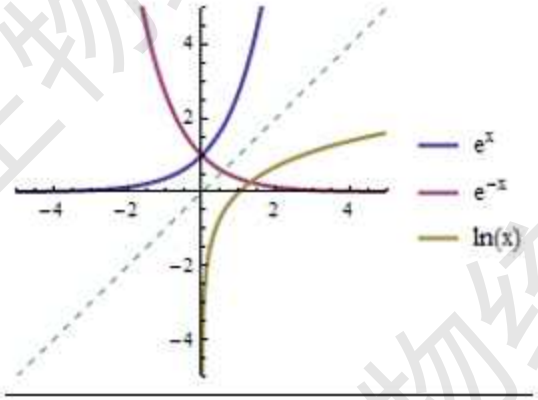
幂(power) → $y = e^{x}$ ← 指数(exponent)

底数(base number)

真数(正数)

自然对数 (natural logarithm)

对数(logarithm) → $x = \log_{e}(y) = \ln(y)$

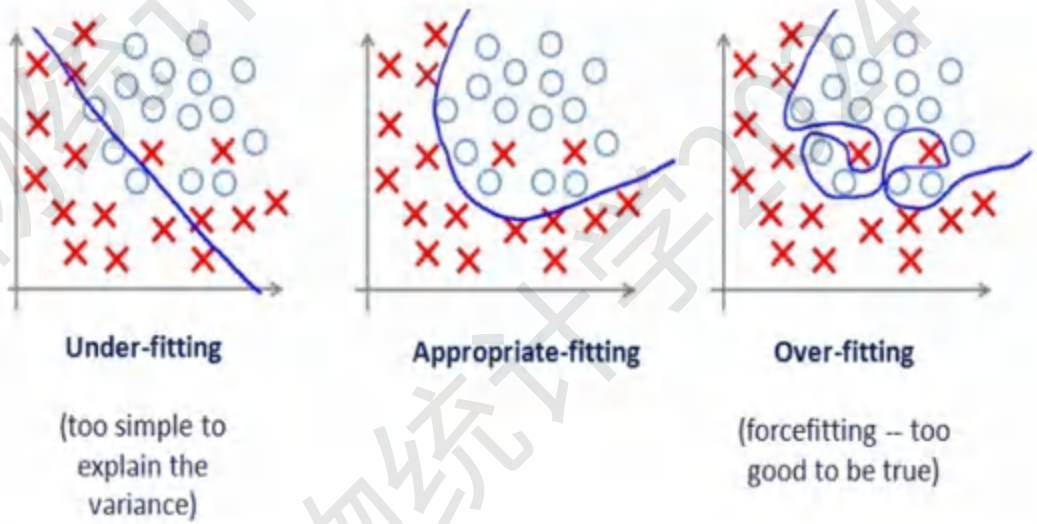


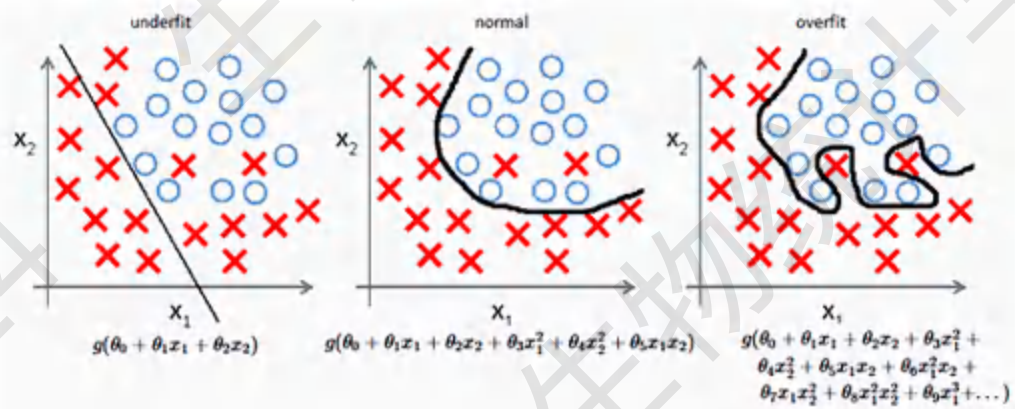
e^x和e^-x的图形是对称的；ln(x)是e^x的逆函数，它们呈45度对称。

生物统计学2024

Neural network

逻辑回归

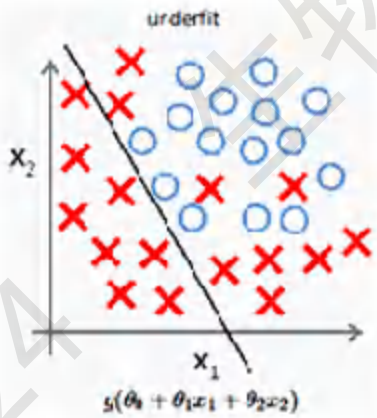


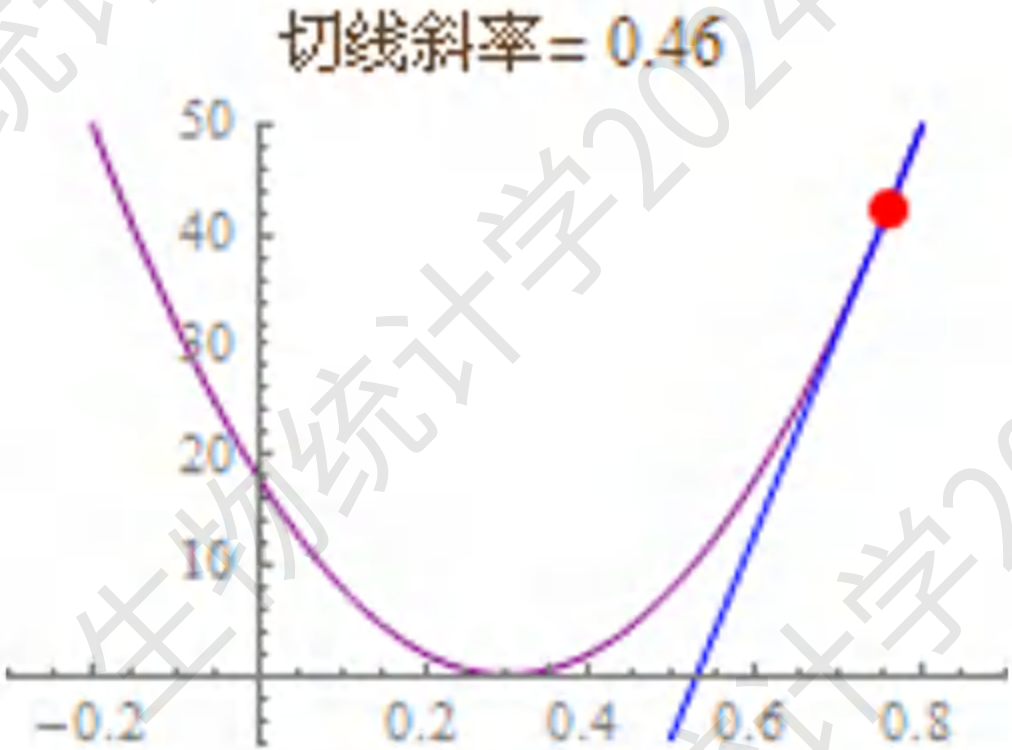


生物统计学2024

Neural network

逻辑回归



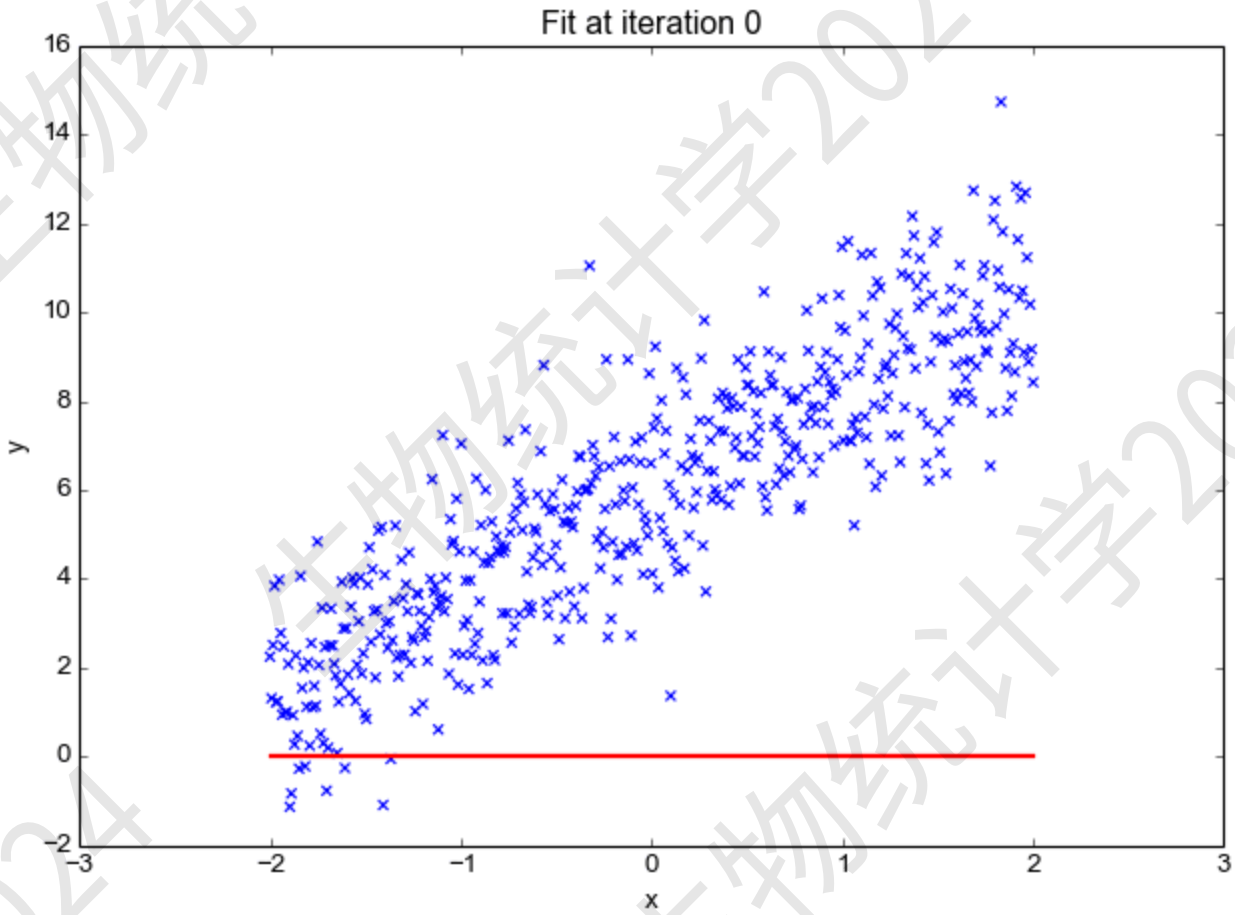


切线每次旋转的幅度叫做学习率(Learning Rate)，加大学习率会加快拟合速度，但是如果调得太大会导致切线旋转过度而无法收敛。[学习率其实是个预先设置好的参数，不会每次变化，不过可以影响每次变化的幅度。]

生物统计学2024

Neural network

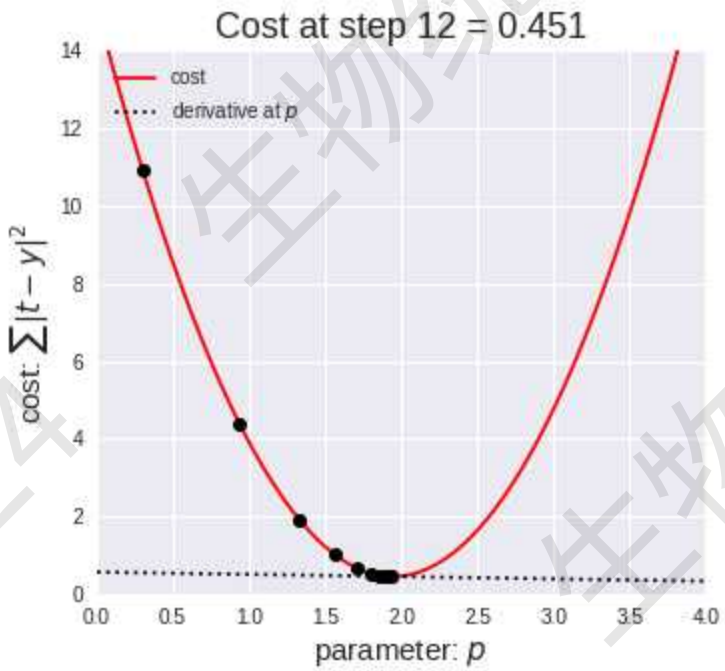
学习率(Learning Rate)

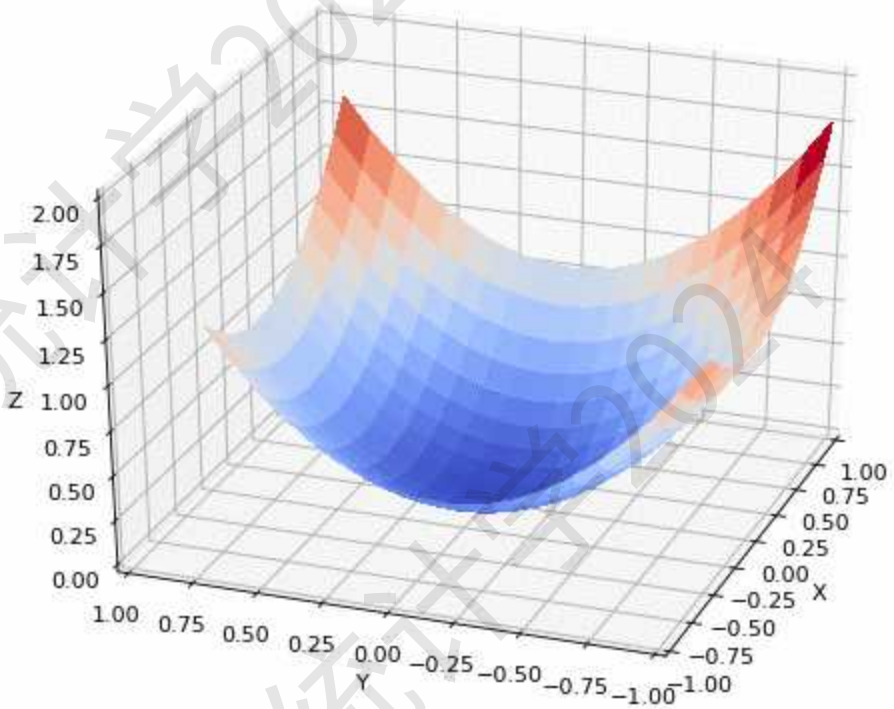


生物统计学2024

Neural network

学习率(Learning Rate)用于梯度下降

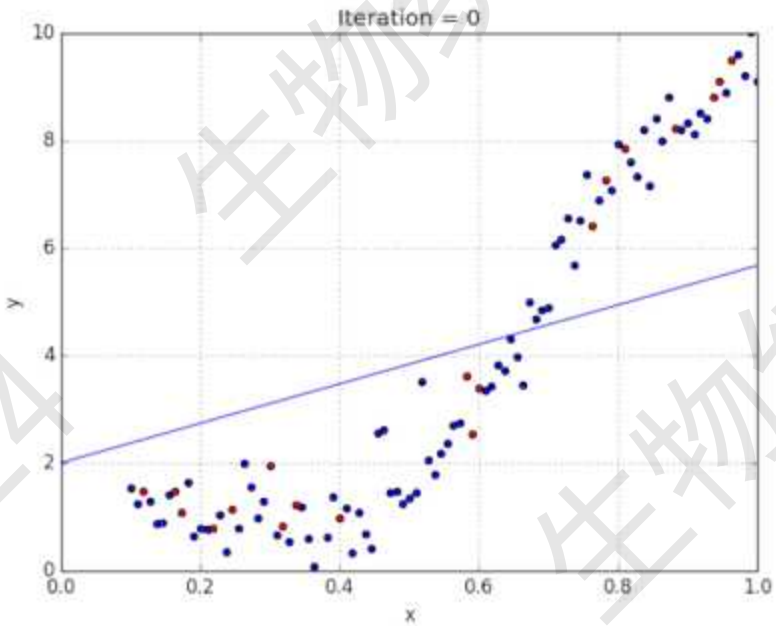


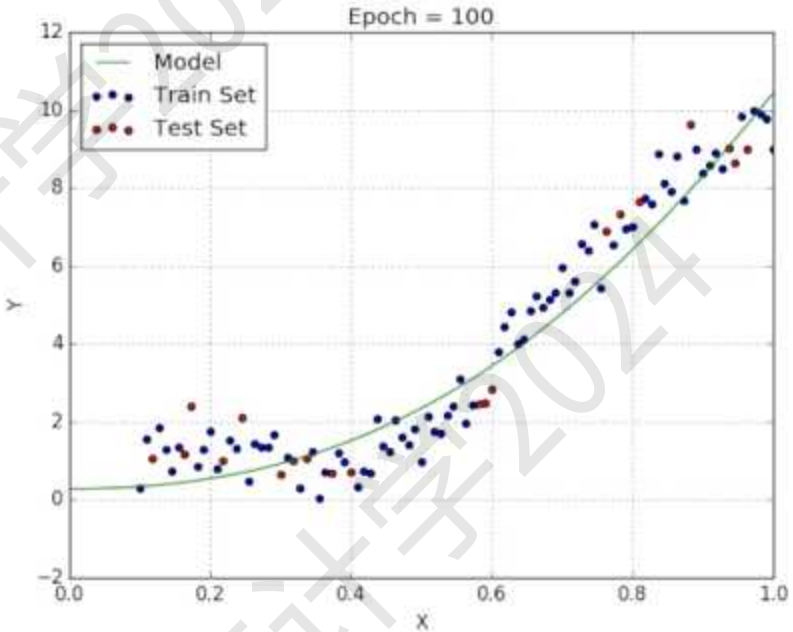


生物统计学2024

Neural network

学习率(Learning Rate)用于回归

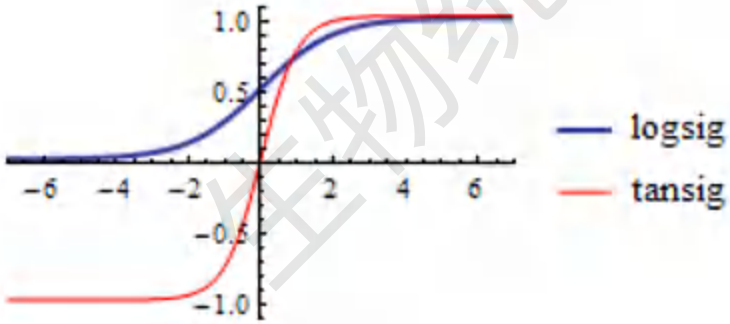


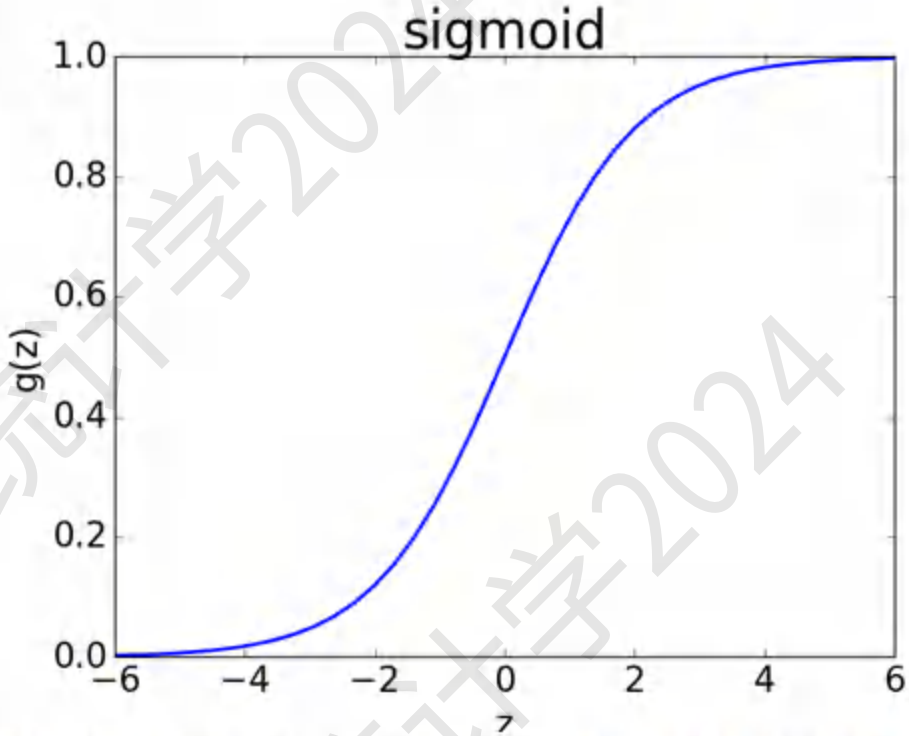


生物统计学2024

Neural network

逻辑回归(Logistic Regression，逻辑斯谛函数)





$$y = \text{logsig}(x) = \frac{1}{1+e^{-x}}$$

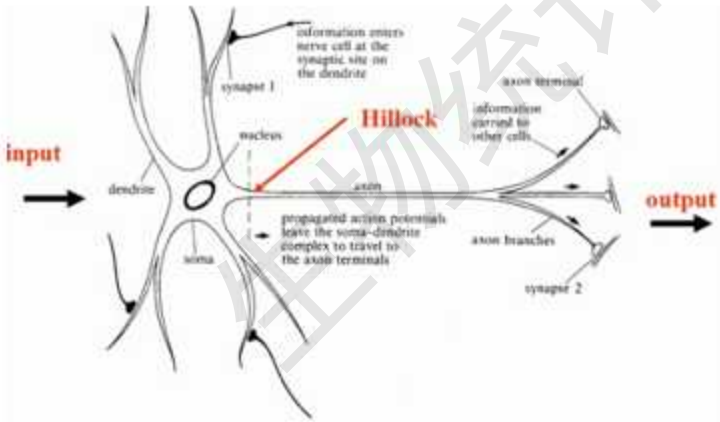
$$y = \text{tansig}(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

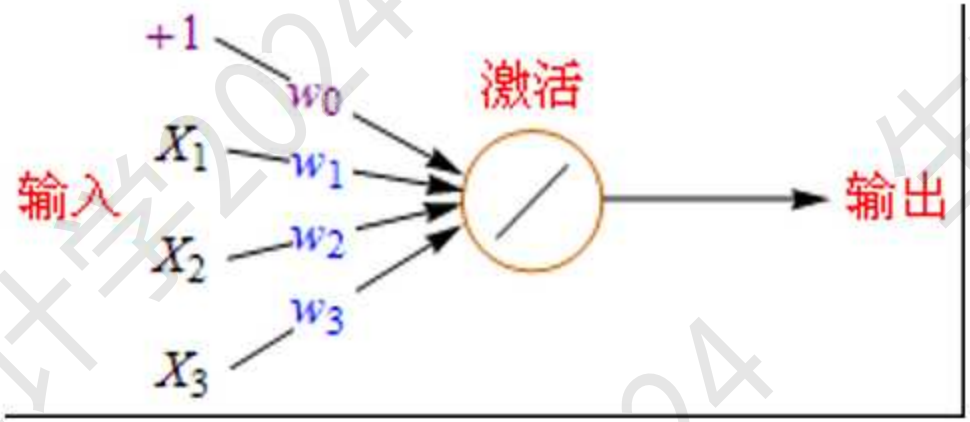
y的阈值处于（−∞，+∞），此时不能很好的给出属于某一类的概率，因为概率的范围是[0,1],我们需要一个更好的映射函数，能够将分类的结果很好的映射成为[0,1]之间的概率，并且这个函数能够具有很好的可微分性。

生物统计学2024

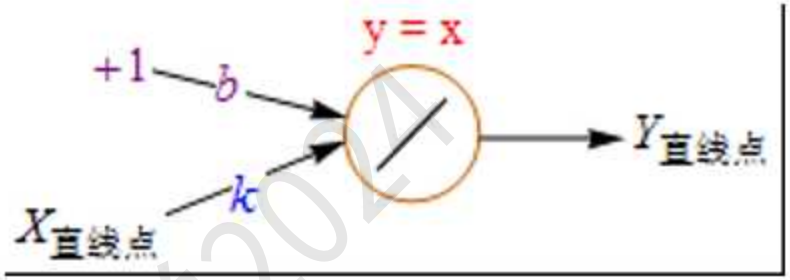
Neural network

逻辑回归

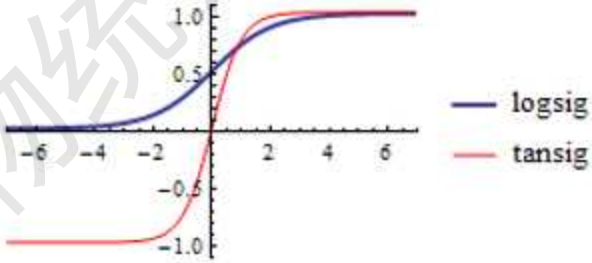




Purelin激活函数



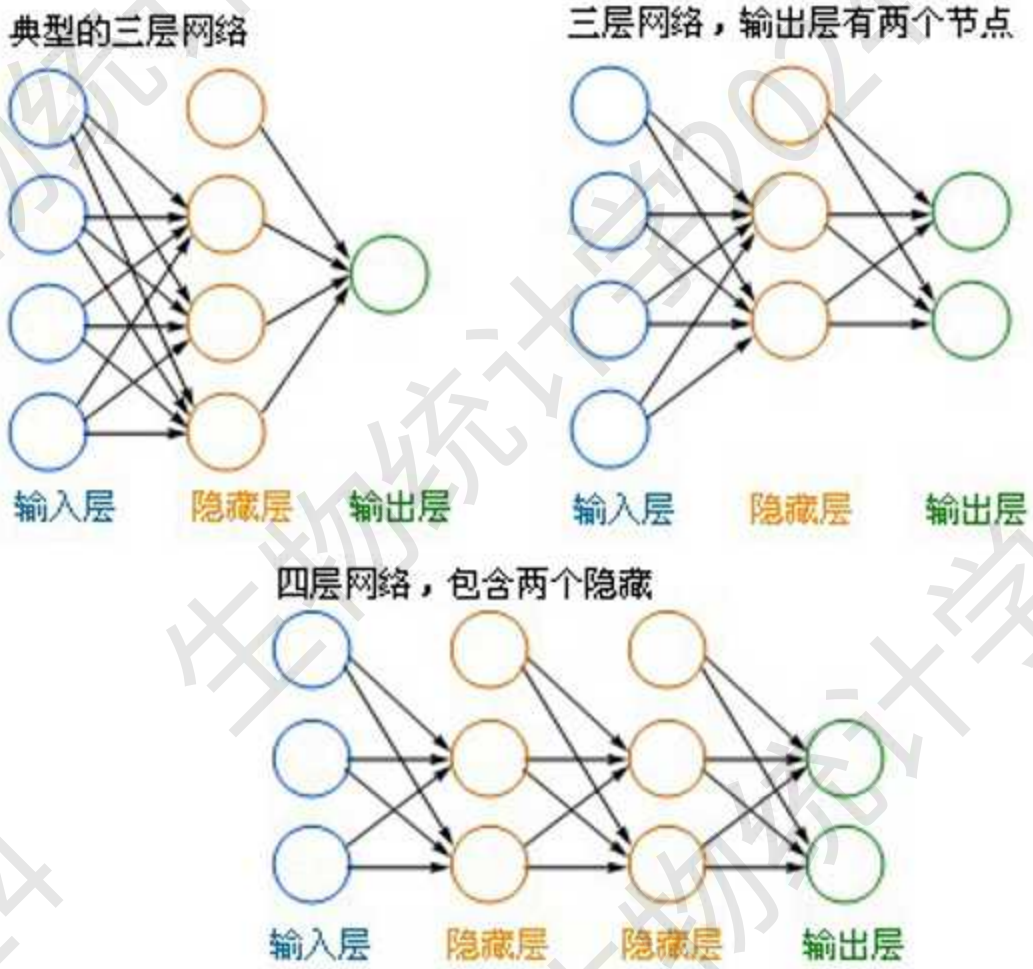
Sigmoid激活函数



生物统计学2024

Neural network

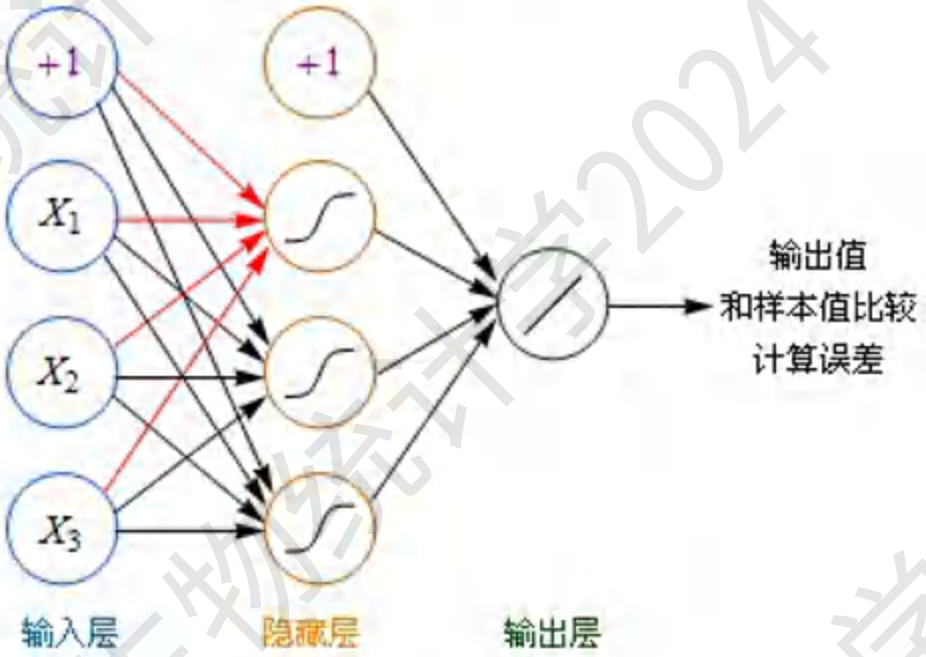
神经网络



生物统计学2024

Neural network

神经网络

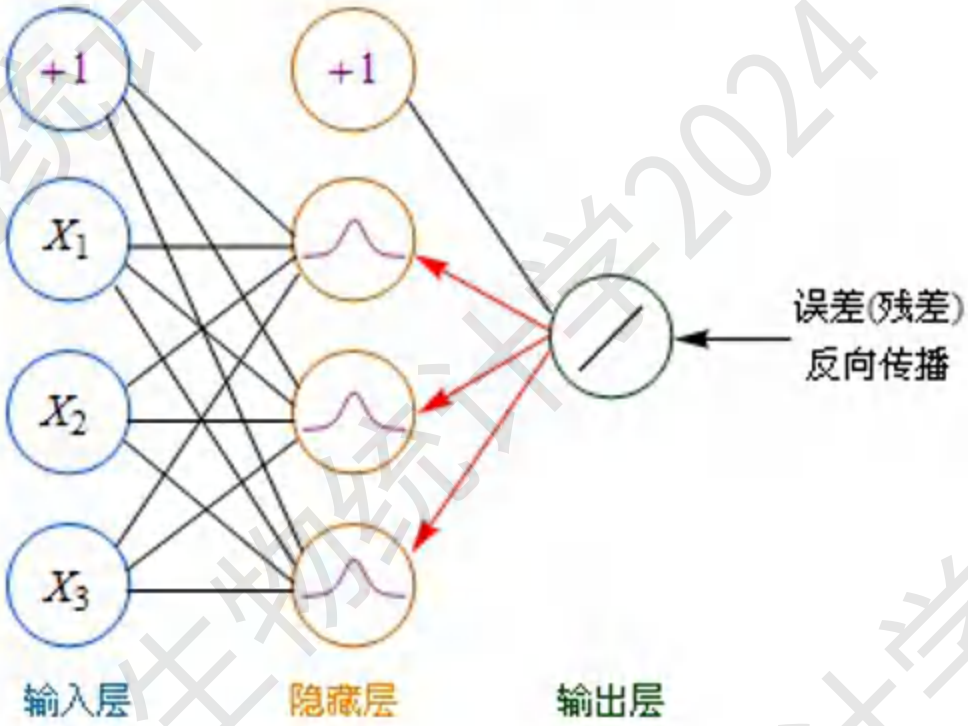


- 隐藏层用都是用Sigmoid作激活函数，而输出层用的是Purelin。这是因为Purelin可以保持之前任意范围的数值缩放，便于和样本值作比较，而Sigmoid的数值范围只能在0~1之间。
- 起初输入层的数值通过网络计算分别传播到隐藏层，再以相同的方式传播到输出层，最终的输出值和样本值作比较，计算出误差，这个过程叫前向传播(Forward Propagation)。

生物统计学2024

Neural network

反向传播算法(BP算法)



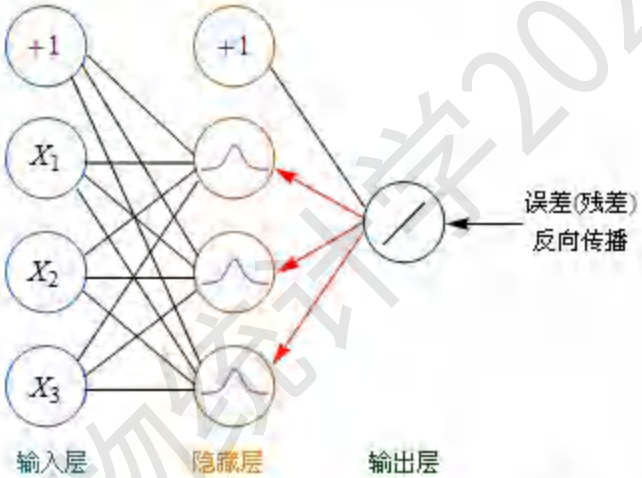
利用前向传播最后输出的结果来计算误差的偏导数，再用这个偏导数和前面的隐藏层进行加权求和，如此一层一层的向后传下去，直到输入层(不计算输入层)，最后利用每个节点求出的偏导数来更新权重

生物统计学2024

Neural network

反向传播算法(BP算法)

前馈神经网络(FeedForward Neural Network)，也叫BP神经网络(Back Propagation Neural Network)。



如果输出层用Purelin作激活函数，Purelin的导数是1，输出层→隐藏层：残差 = -(输出值-样本值)

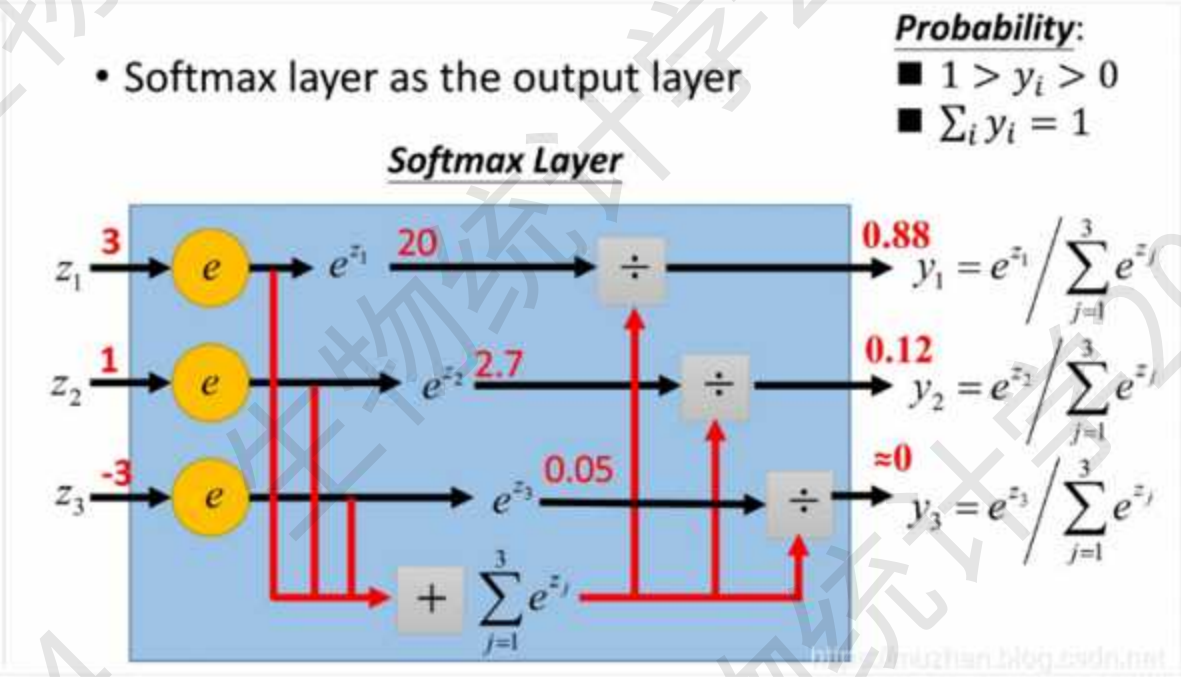
如果用Sigmoid(logsig)作激活函数，那么：Sigmoid导数 = Sigmoid*(1-Sigmoid)
输出层→隐藏层：残差 = -(Sigmoid输出值-样本值) * Sigmoid*(1-Sigmoid) = -(输出值-样本值)输出值(1-输出值)
隐藏层→隐藏层：残差 = (右层每个节点的残差加权求和)* 当前节点的Sigmoid*(1-当前节点的Sigmoid)

生物统计学2024

Neural network

softmax分类器

$$softmax(x_0) = \frac{e^{x_0}}{e^{x_0} + e^{x_1} + e^{x_2}}$$

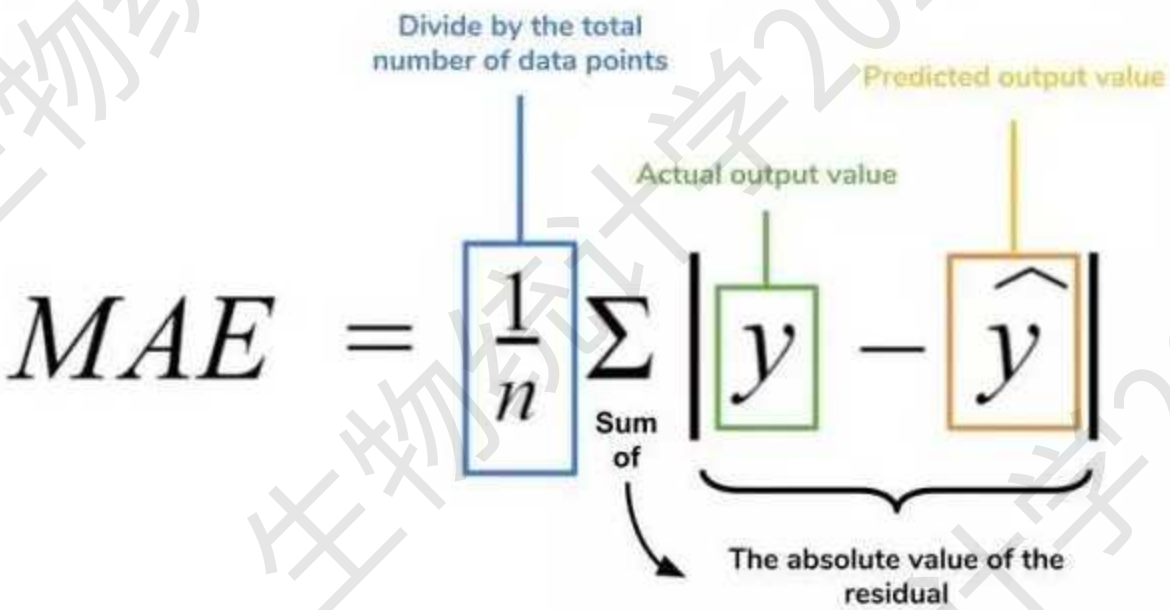


softmax对神经元的输出信号进行加工，输出为分类的概率值。

生物统计学2024

Neural network

平均绝对误差（MAE）



生物统计学2024

Neural network

均方误差（MSE）

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \hat{y}\right)^2}_{\text{The square of the difference between actual and predicted}}$$

生物统计学2024

Neural network

示例

图例：

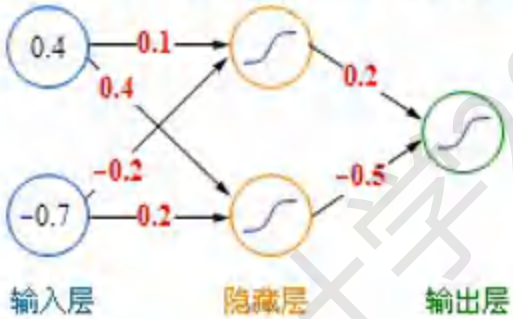
- logsig激活函数
- 残差(误差偏导数)
- 加权求和

训练集的数据，首先对第一行进行处理

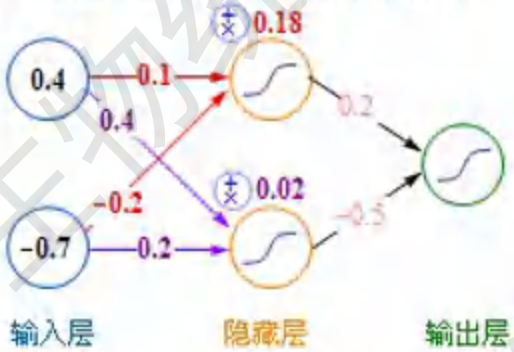
X1	X2	样本值
0.4	-0.7	0.1
0.3	-0.5	0.05
0.6	0.1	0.3
0.2	0.4	0.25
0.1	-0.2	0.12

为了适应输出层的logsig变换，X1,X2的数值范围规定在至0～1之间

1.每个节点之间的初始权重一般是随机生成的



2.对输入层节点进行加权求和计算



$0.4\times0.1+(-0.7)\times(-0.2)=0.18$

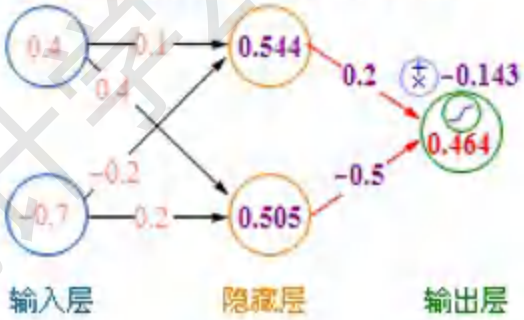
$0.4\times0.4+(-0.7)\times(0.2)=0.02$

3.执行Sigmoid激活

$$\text{logsig}(0.18)=\frac{1}{1+e^{-0.18}}=0.544$$

$$\text{logsig}(0.02)=\frac{1}{1+e^{-0.02}}=0.505$$

4.用相同的方法计算出输出层的值



$0.544\times0.2+0.505\times(-0.5)=-0.143$

$$\text{logsig}(-0.143)=\frac{1}{1+e^{0.143}}=0.464$$

生物统计学2024

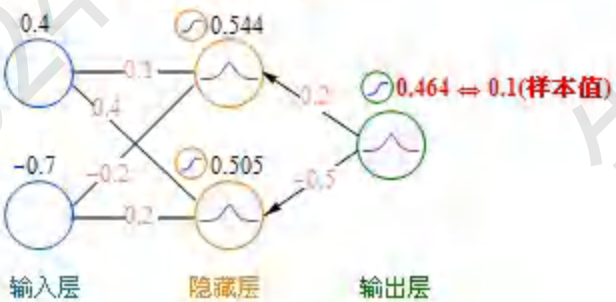
Neural network

示例

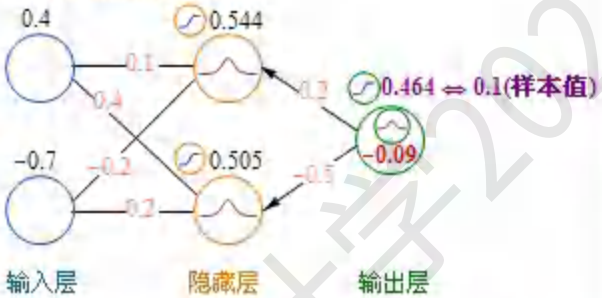
5.计算误差，误差接近0时收敛

误差=(0.464 - 0.1)^2=0.132 (误差<0.0001,可以收敛)

6.开始准备误差的反向传播



7.计算输出层的残差



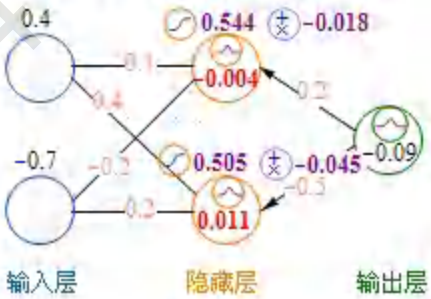
残差=-(0.464 - 0.1)×0.464×(1 - 0.464)=-0.09

8.输出层节点的残差加权求和

-0.09×0.2=-0.018 (输出层只有一个节点,不用求和)

-0.09×0.5=-0.045

9.继续求隐藏层的残差



残差1=(-0.018)×(0.544)×(1-0.544)=-0.004

残差2=0.045×0.505×(1-0.505)=0.011

10.开始准备更新第一层权重，设学习率为0.6

0.4×(-0.004)×0.6=-0.0016×0.6=-0.00096

0.4× 0.011 ×0.6= 0.0044×0.6= 0.00264

-0.7×(-0.004)×0.6= 0.0028×0.6= 0.00168

-0.7× 0.011 ×0.6=-0.0077×0.6=-0.00462

生物统计学2024

Neural network

示例

11.更新第一层权重

```
 0.1-0.00096= 0.09904
 0.4+0.00264= 0.40264
-0.2+0.00168=-0.19832
 0.2-0.00462= 0.19538
```

12.使用前面两步的方法，计算出后两层的权重

```
0.544×(-0.09)×0.6=-0.04896×0.6=-0.029376
0.505×(-0.09)×0.6=-0.04545×0.6=-0.02727

 0.2-0.029376= 0.170624
-0.5-0.02727 =-0.52727
```

利用更新之后的权重，对训练集的每一条数据反复进行前面1~12步的计算，直到最后收敛

- **Step 15 :**

  With the updated weights **[V]** and **[W]**, error is calculated again and next training set is taken and the error will then get adjusted.

- **Step 16 :**

  Iterations are carried out till we get the error less than the tolerance.

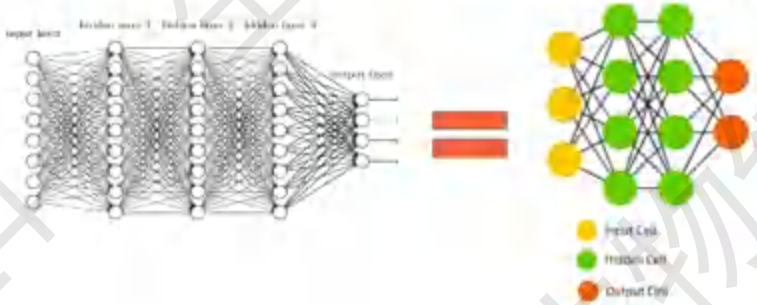
- **Step 17 :**

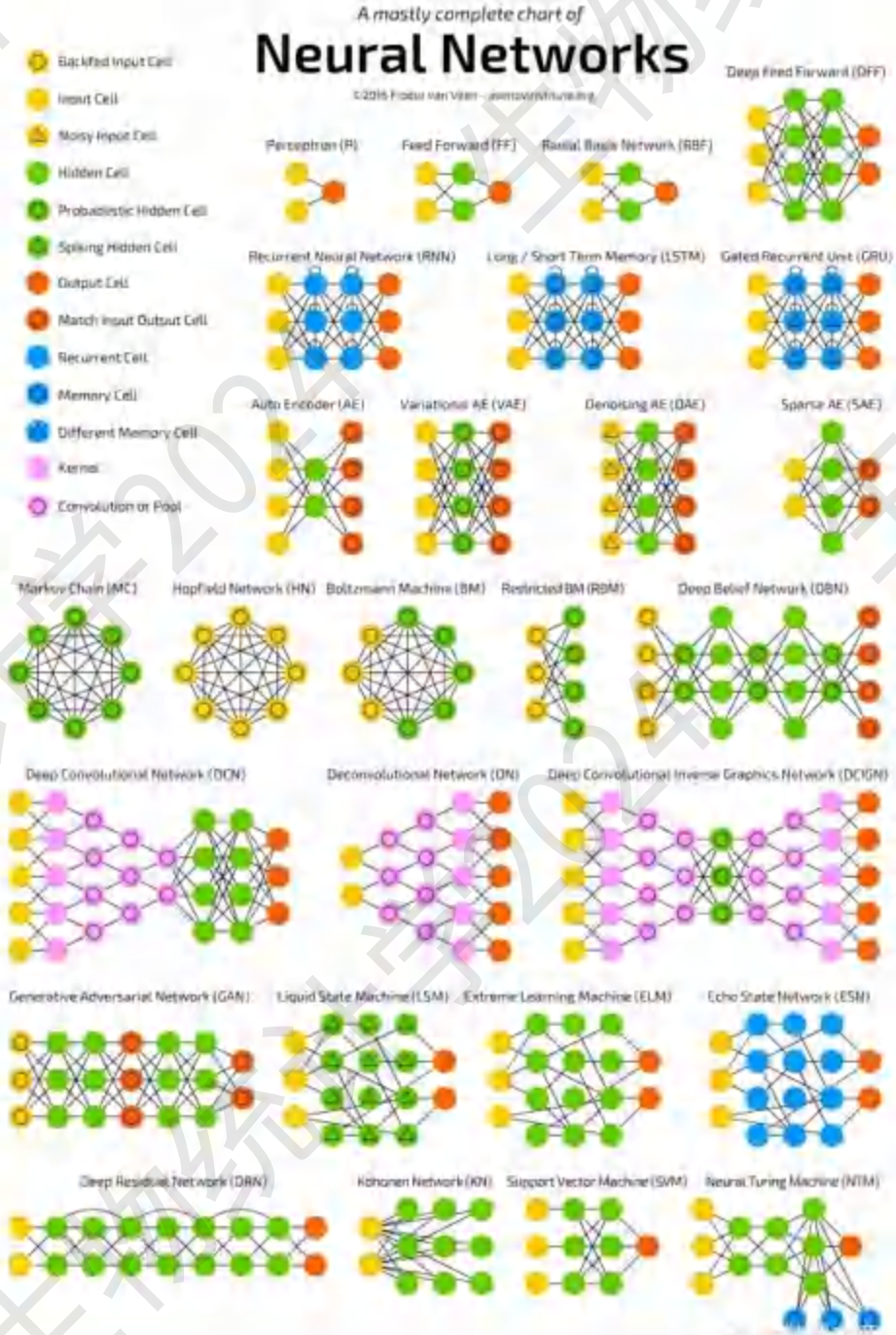
  Once the weights are adjusted the network is ready for inferencing new objects .

生物统计学2024

Neural network

Deep Neural Network (DNN)



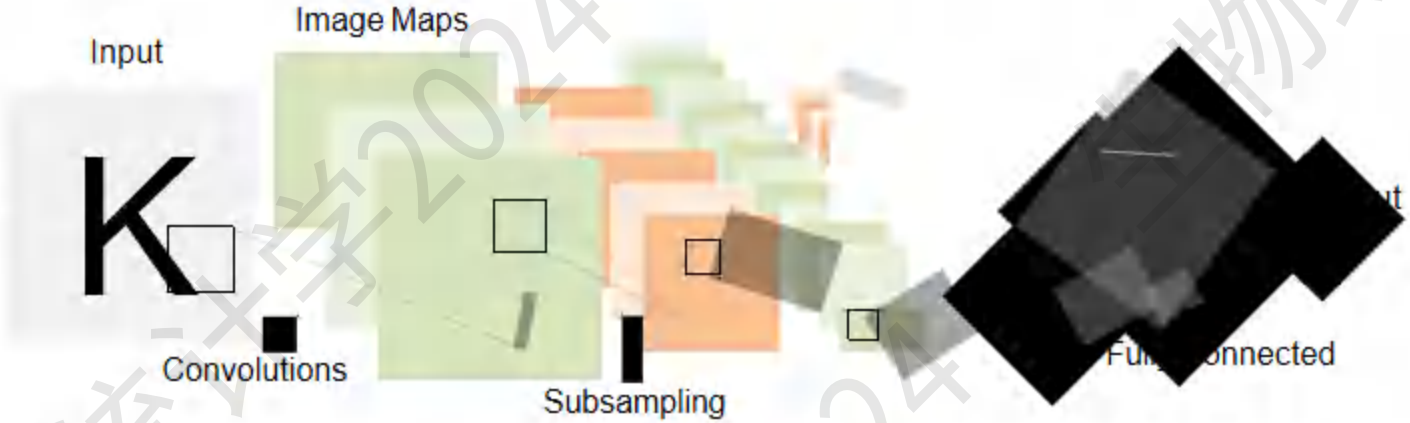


Adapted from: http://www.asimovinstitute.org/neural-network-zoo/

生物统计学2024

1998

LeCun et al.



# of transistors

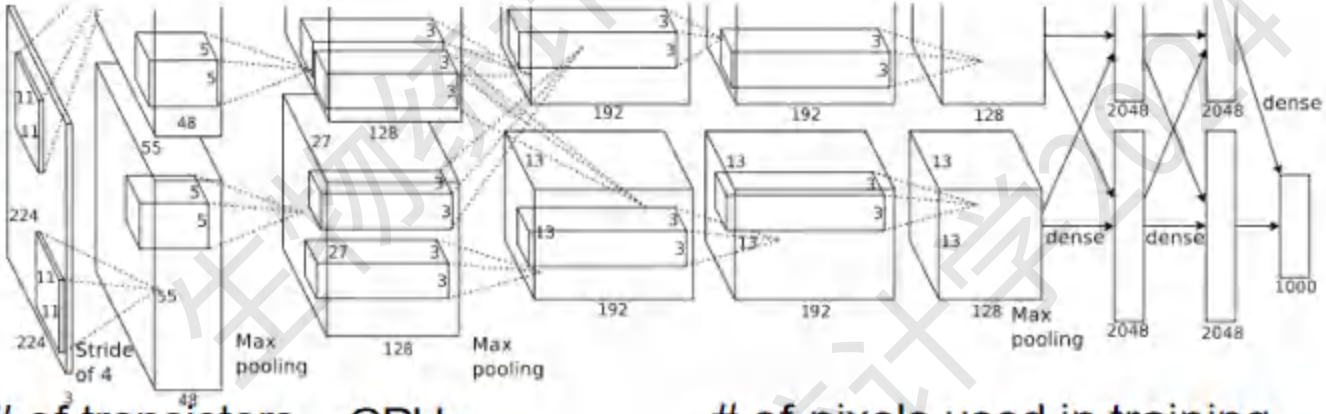
10^6

# of pixels used in training

10^7 NIST

2012

Krizhevsky et al.



# of transistors

10^9

GPUs

# of pixels used in training

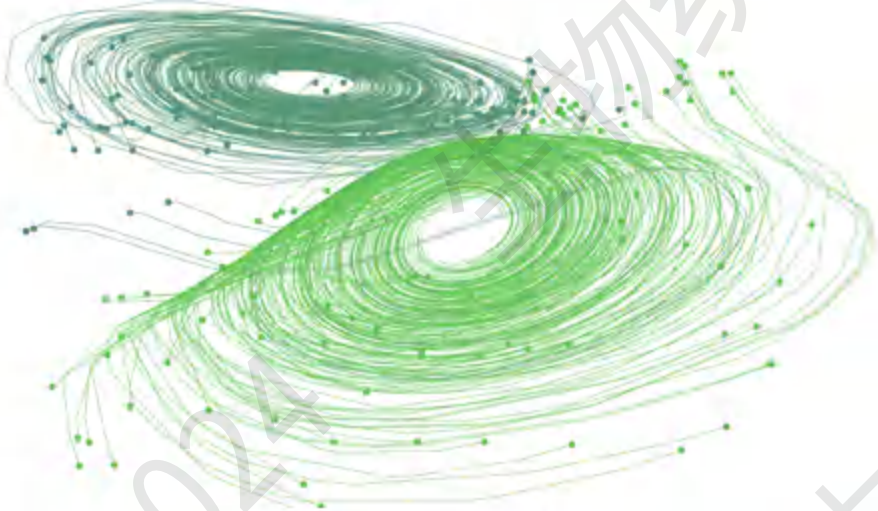
10^{14} IMAGENET

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Convolutional Neural Networks (CNN) were NOT invented overnight!

生物统计学2024

Neural network



Intelligent Machines

A radical new neural network design could overcome big challenges in AI

Researchers borrowed equations from calculus to redesign the core machinery of deep learning so it can model continuous processes like changes in health.

by Karen Hao December 12, 2018

Advertisement

深度学习的基本过程

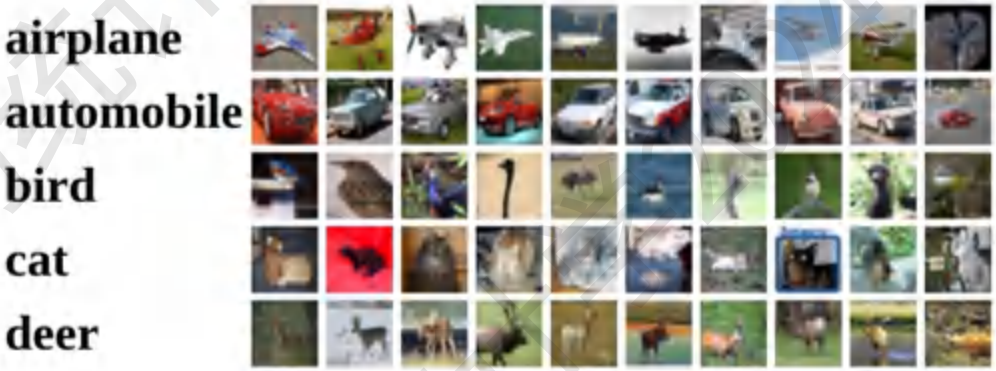
生物统计学2024

Data-Driven Approach

1. Collect a dataset of images and labels
2. Use Machine Learning to train a classifier
3. Evaluate the classifier on new images

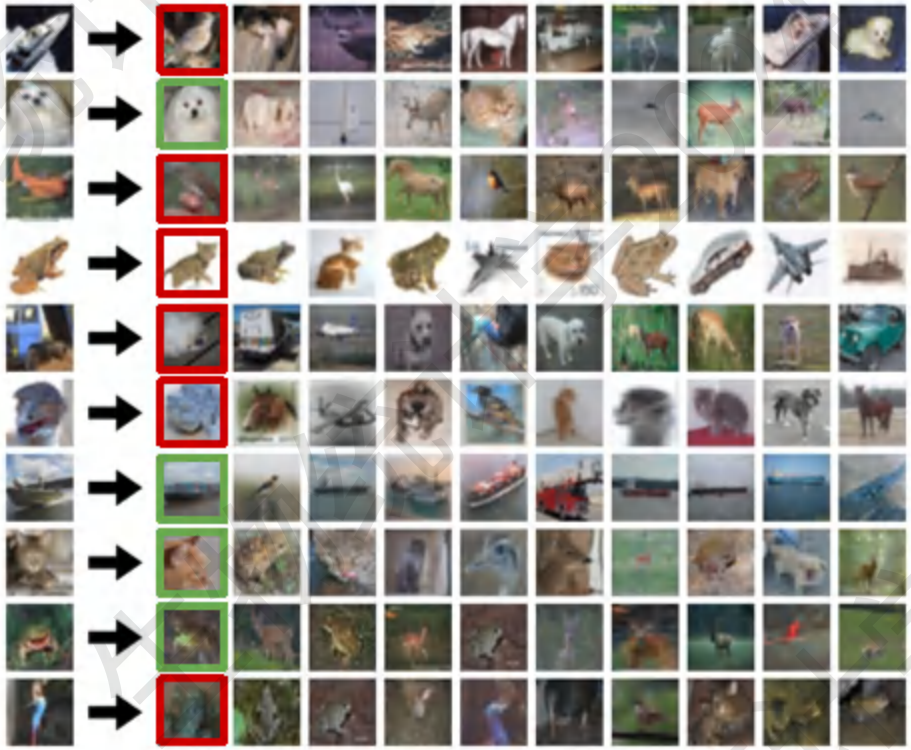
```
def train(images, labels):
  # Machine learning!
  return model


def predict(model, test_images):
  # Use model to predict labels
  return test_labels
```

Example training set



生物统计学2024

What does this look like?



生物统计学2024

Parametric Approach: Linear Classifier

3072x1

$$f(x,W) = \boxed{W}\boxed{x}$$

10x1 10x3072

Image

Array of **32x32x3** numbers
(3072 numbers total)

$f(x,W)$ → **10** numbers giving class scores

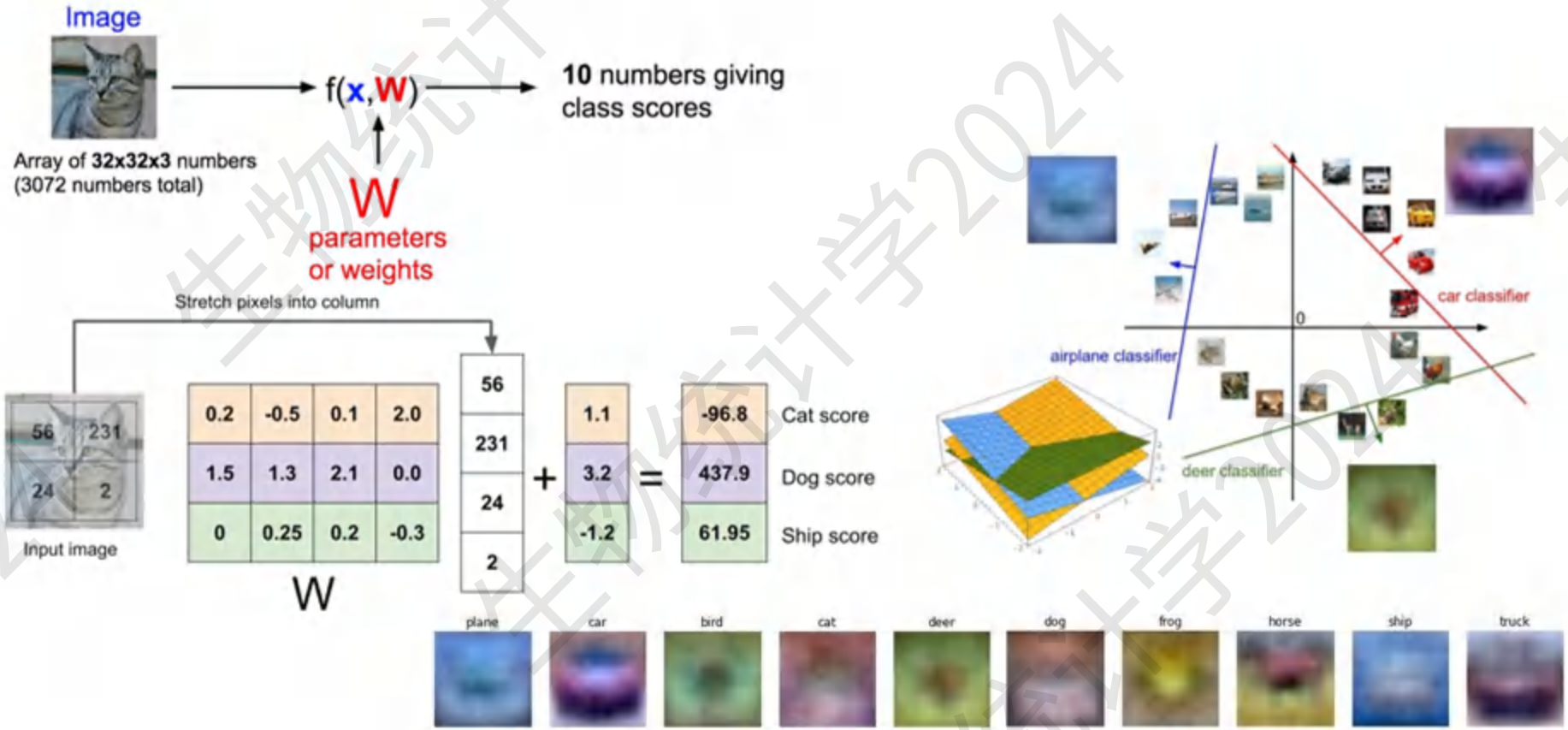
W

parameters
or weights

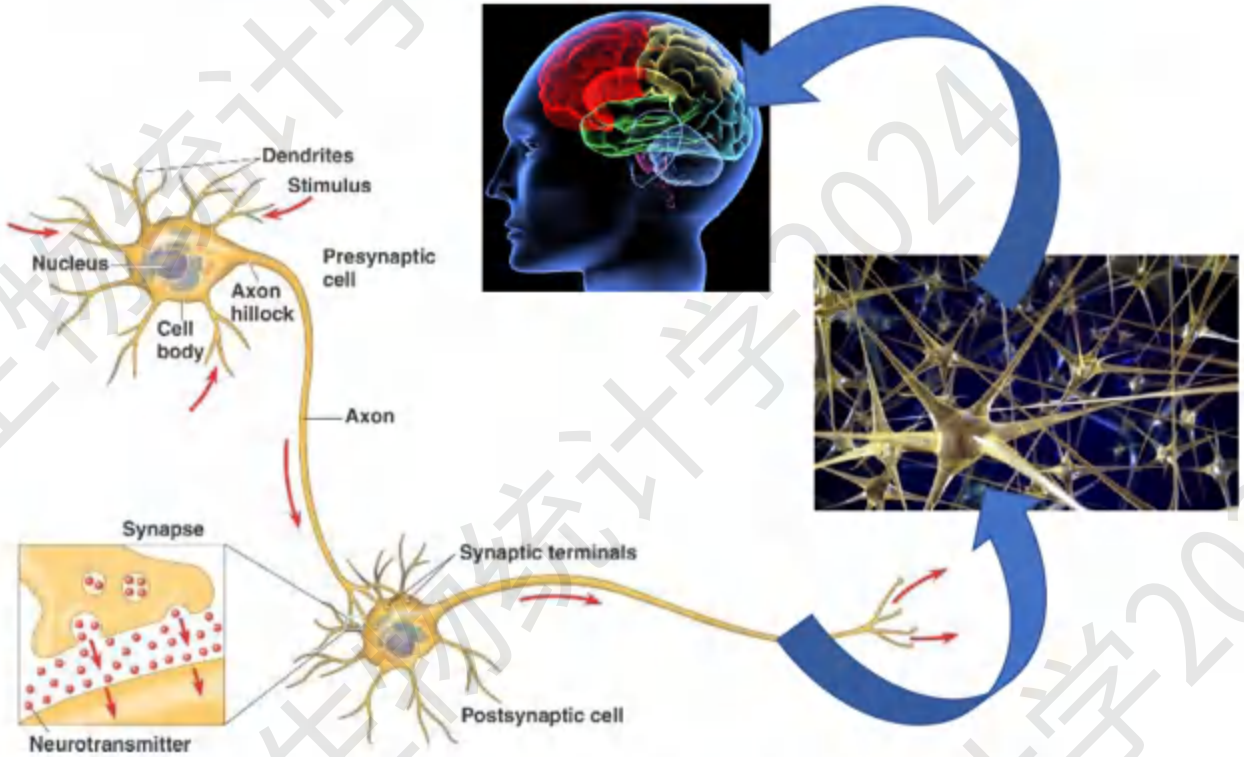
生物统计学2024

Parametric Approach: Linear Classifier

$$f(x,W) = Wx + b$$



Neural Network introduction



Neural network model is inspired by human neuron

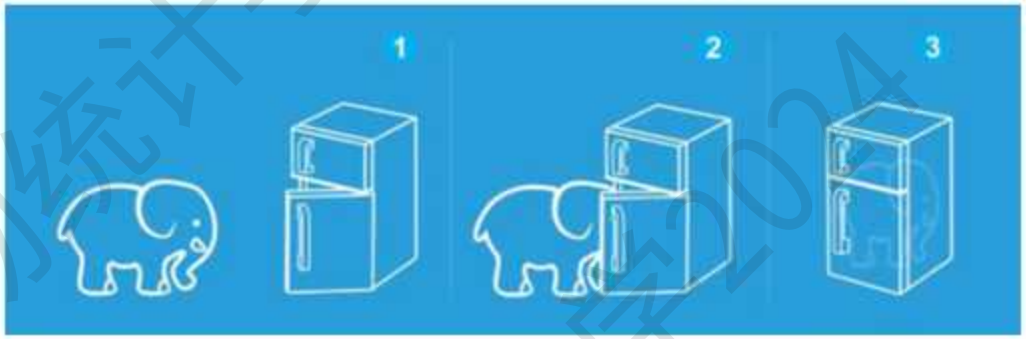
Each neuron is a judgement unit

Hubel & Wiesel, 1959

生物统计学2024

Neural Network workflow

Neural network is so simple……



Feature Selection → Model Construction → Pick the best model

The key point

Deep learning should be a problem-oriented tool.

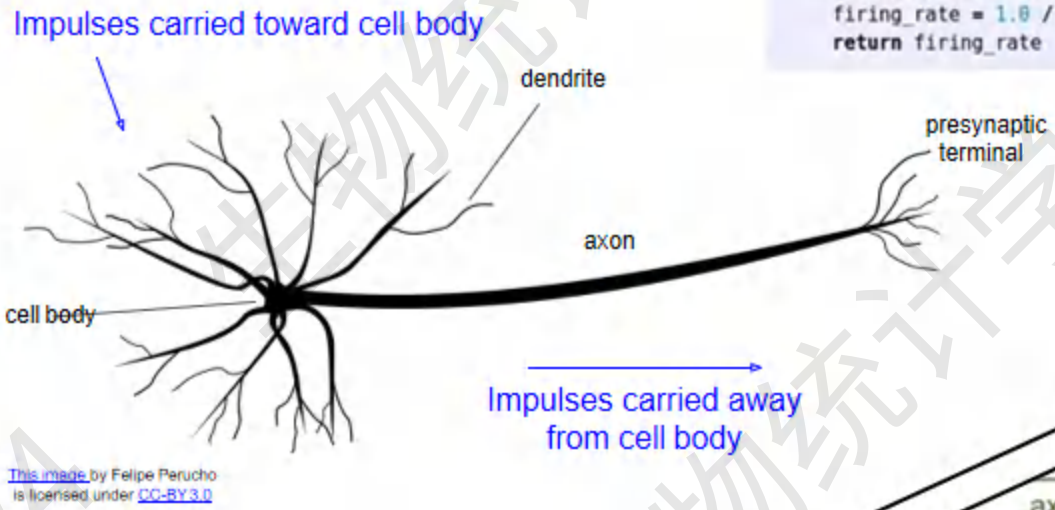
Number of layers
Functions Selection

Parameters Adjustment

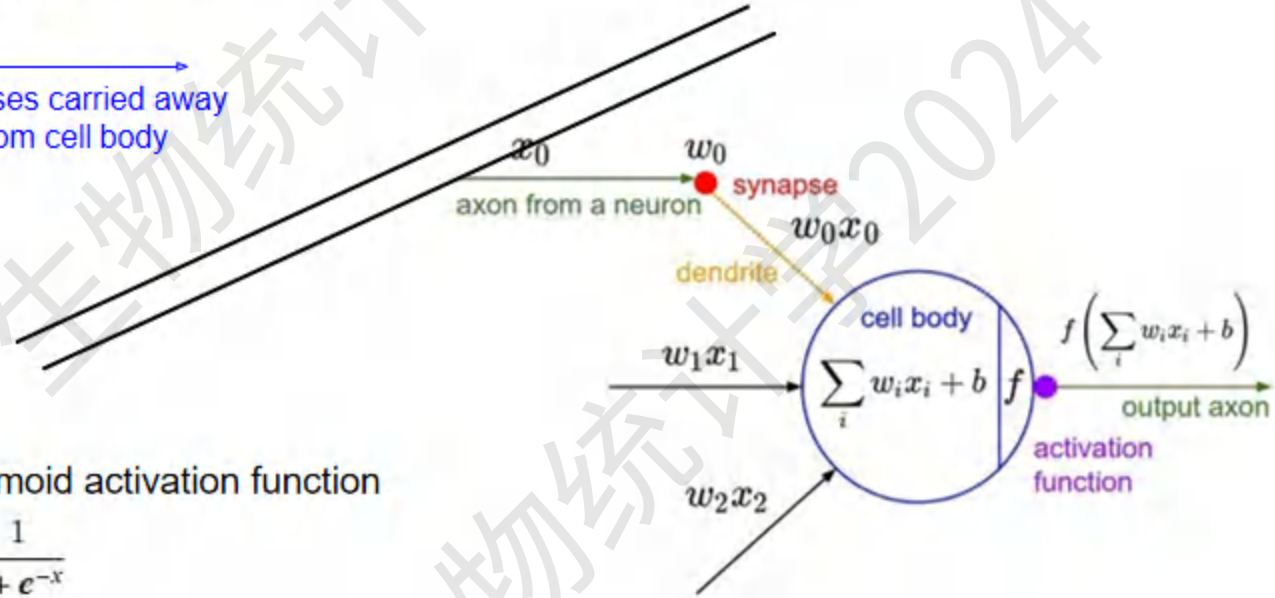
生物统计学2024

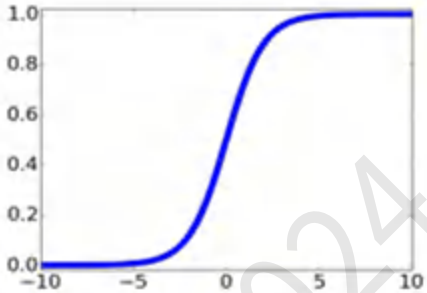
Neuron Network

```python
class Neuron:
  # ...
  def neuron_tick(inputs):
    """ assume inputs and weights are 1-D numpy arrays and bias is a number """
    cell_body_sum = np.sum(inputs * self.weights) + self.bias
    firing_rate = 1.0 / (1.0 + math.exp(-cell_body_sum)) # sigmoid activation function
    return firing_rate
```



This image by Felipe Perucho is licensed under CC-BY 3.0





sigmoid activation function

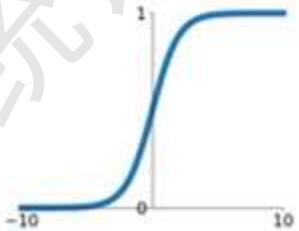
$$\frac{1}{1+e^{-x}}$$

生物统计学2024

Activation Functions

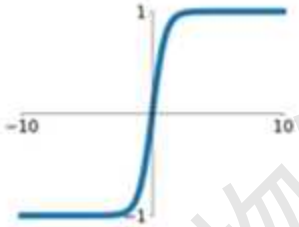
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



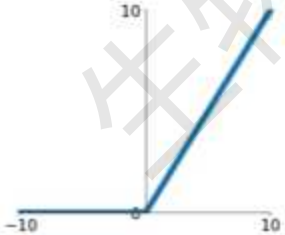
tanh

$\tanh(x)$



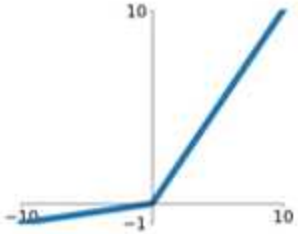
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

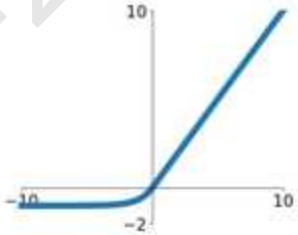


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

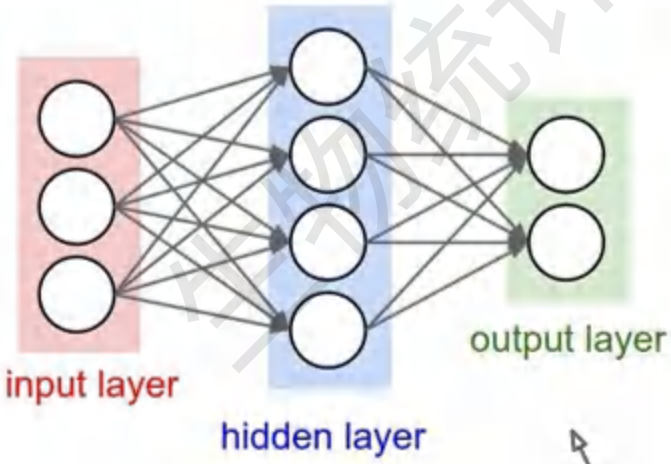
ELU

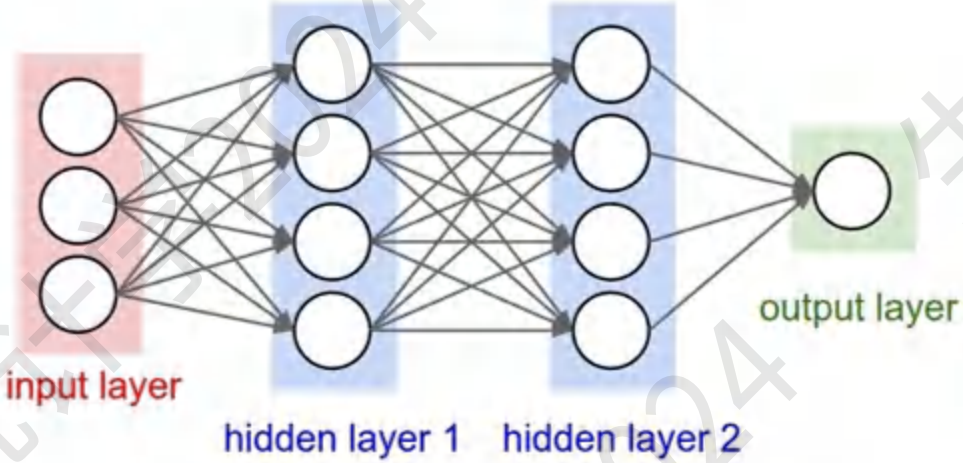
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



生物统计学2024

Neural networks: Architectures





"2-layer Neural Net", or
"1-hidden-layer Neural Net"

"3-layer Neural Net", or
"2-hidden-layer Neural Net"

"Fully-connected" layers

生物统计学2024

Convolutional Neural Networks

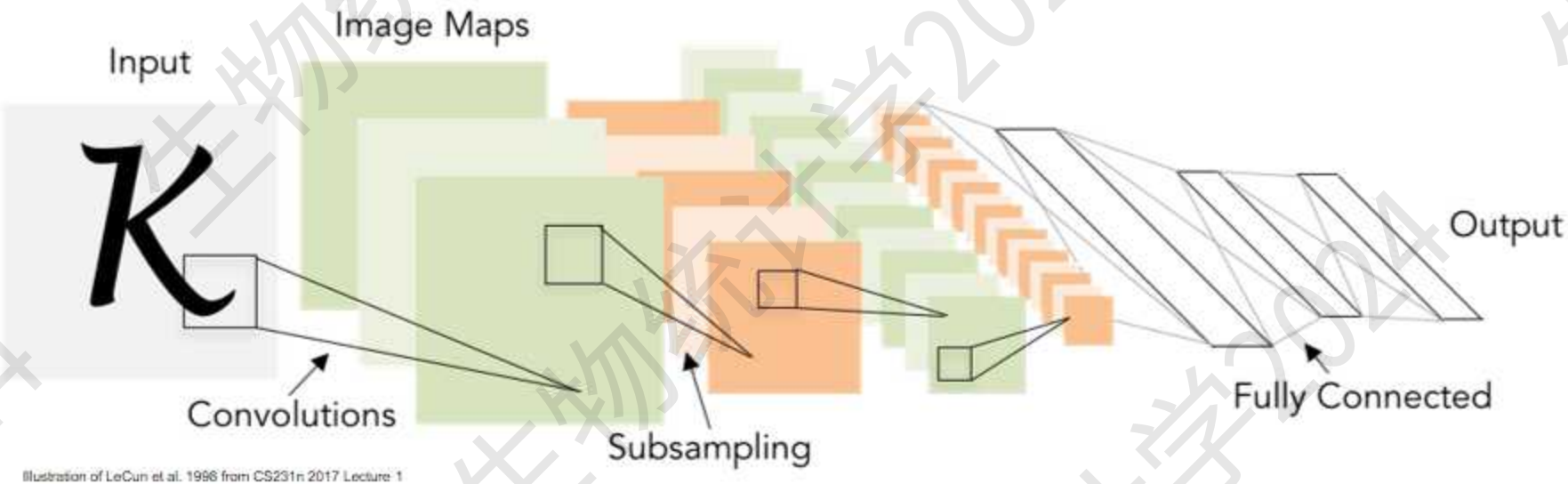
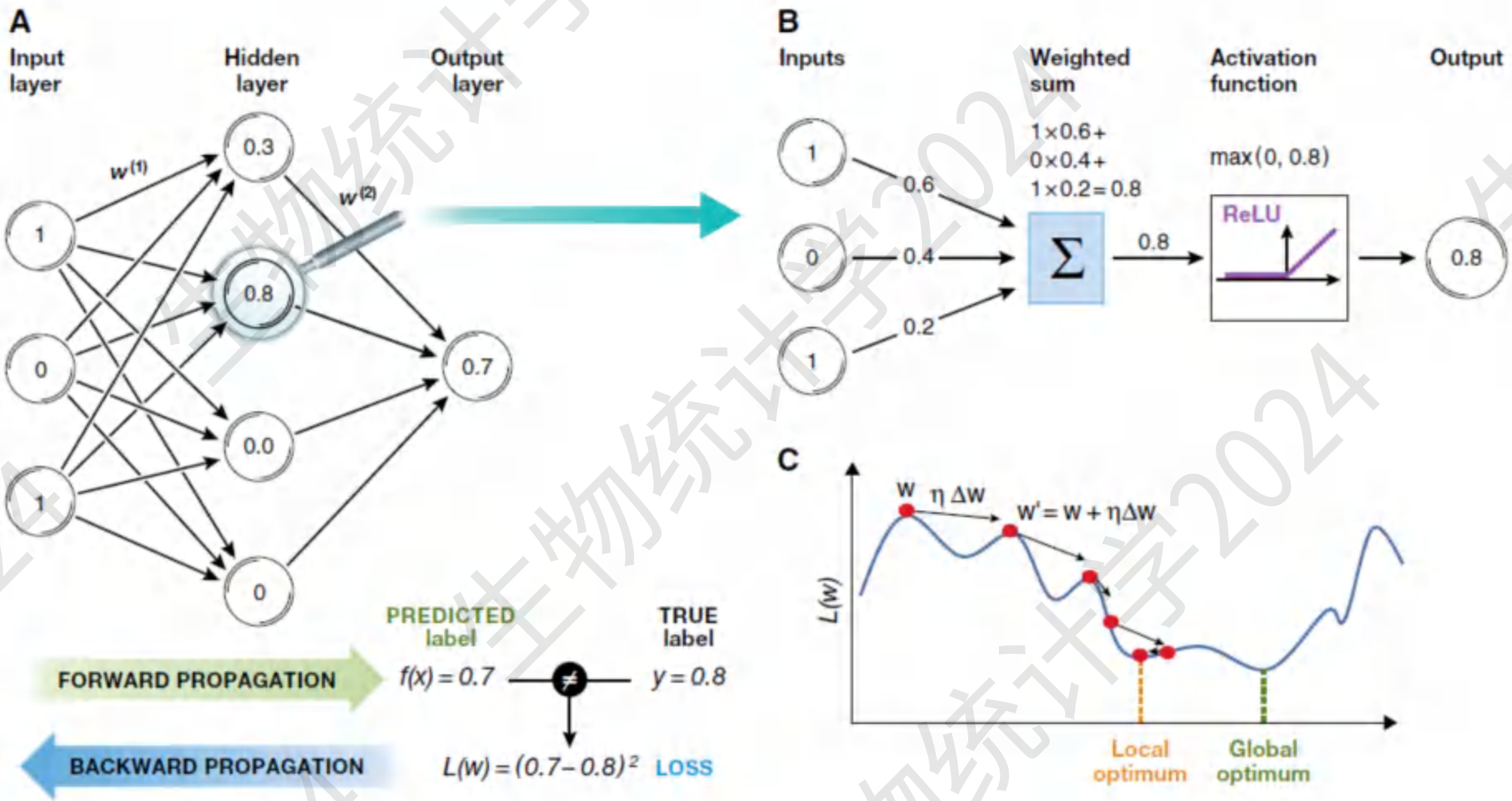


Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

生物统计学2024

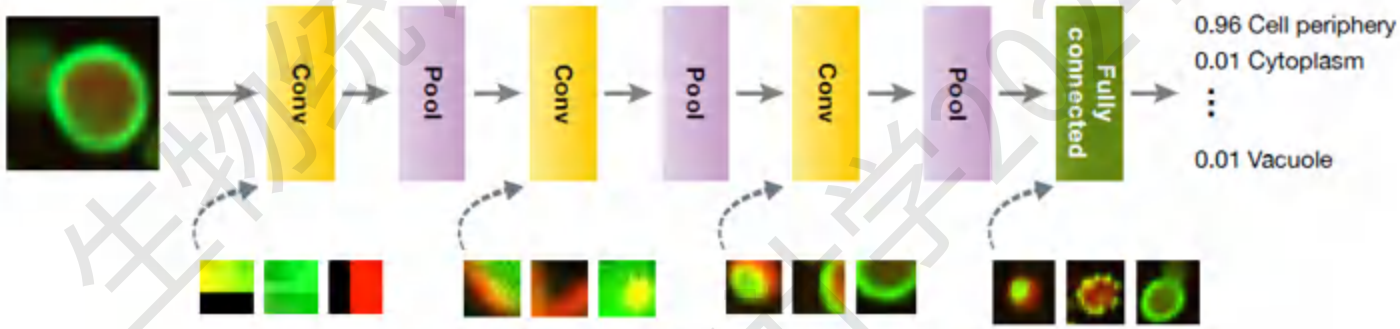
Artificial Neural Network



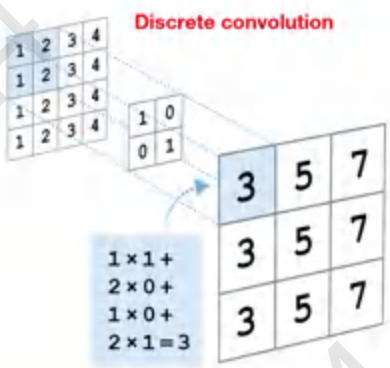
Adapted from "Deep learning for computational biology. Molecular Systems Biology, 2016."

Convolution Neural Network

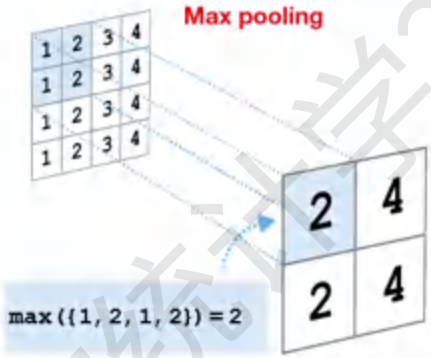
Convolution neural network is similar to neural network, it is a good approach to process figure



Convolution and pooling process : edge features extraction



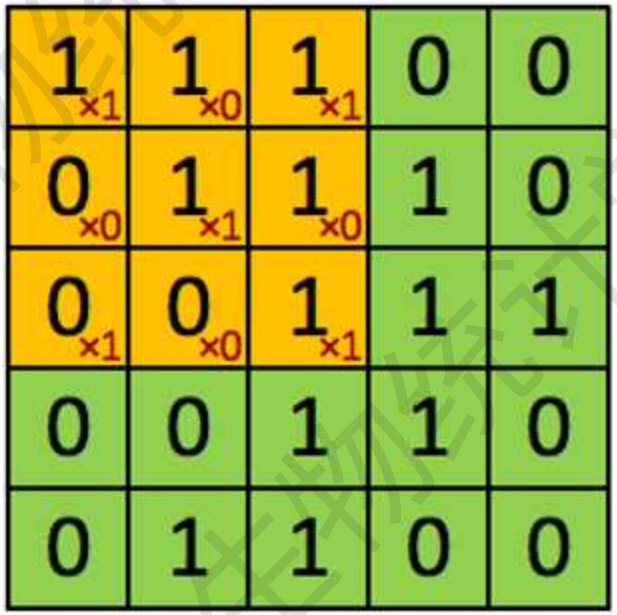
Convolution



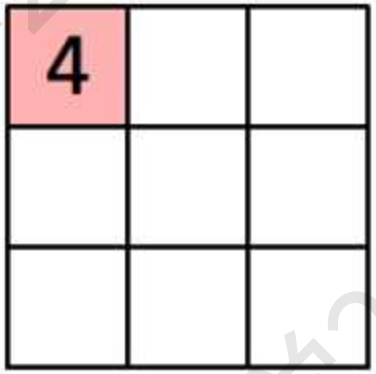
Pooling

生物统计学2024

Convolution Neural Network



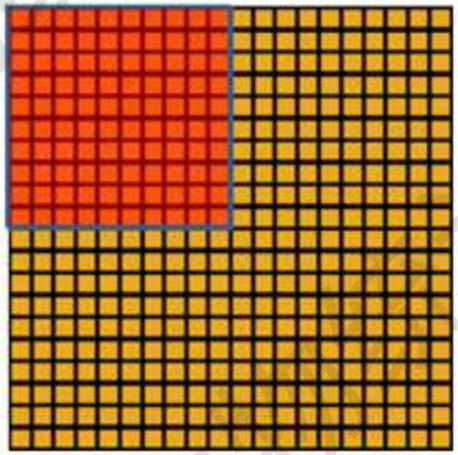
Image



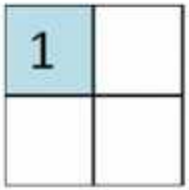
Convolved Feature

生物统计学2024

Convolution Neural Network



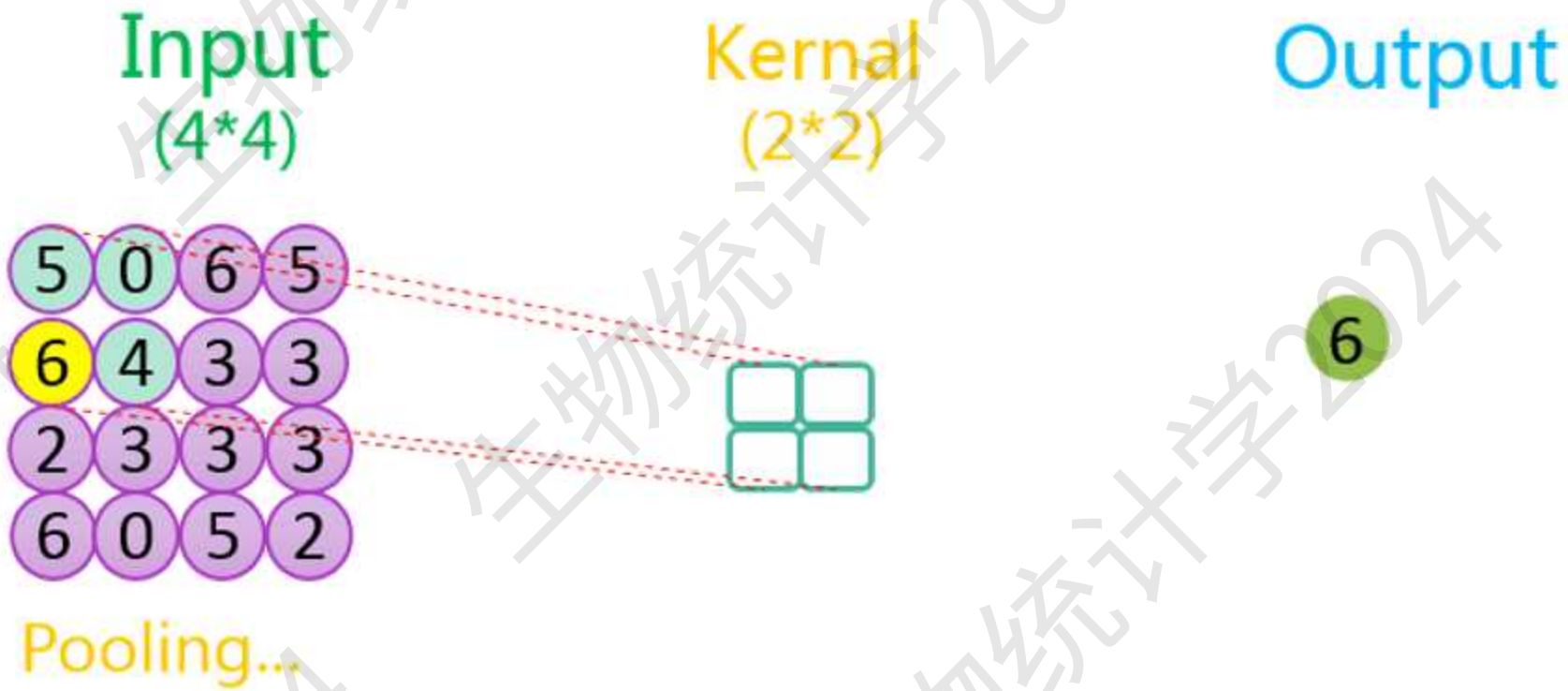
Convolved feature



Pooled feature

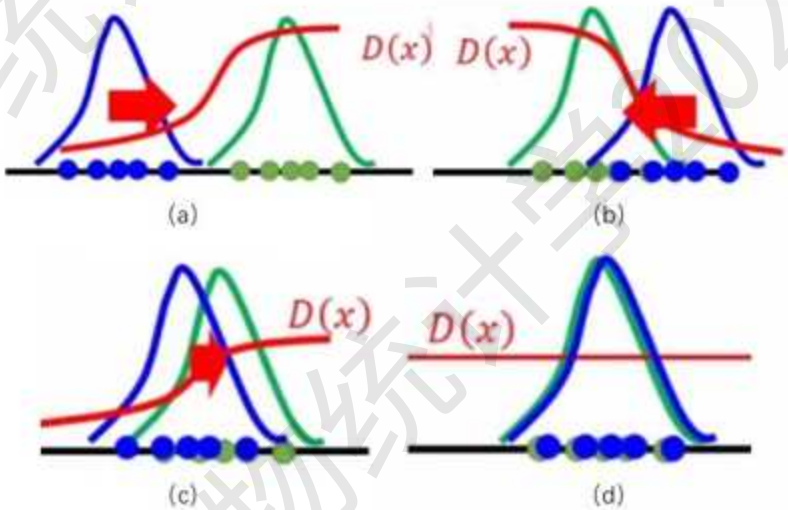
生物统计学2024

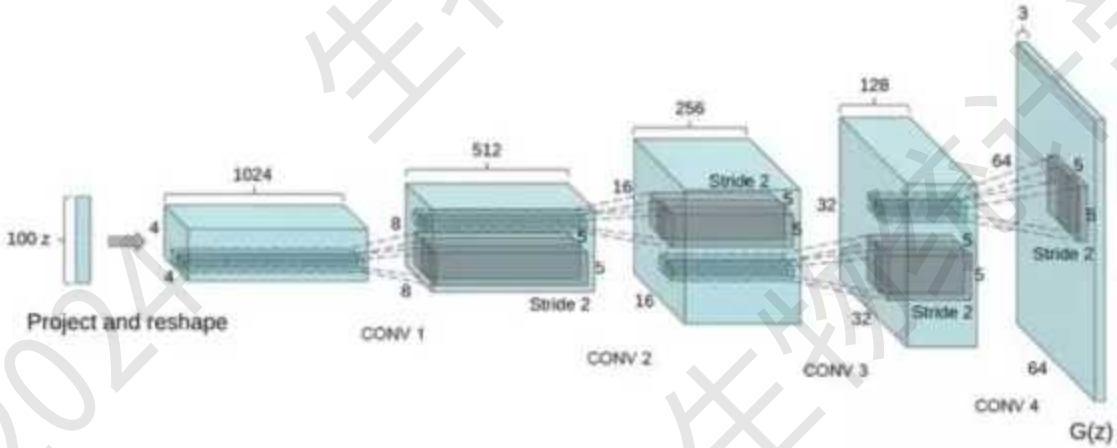
Convolution Neural Network



生物统计学2024

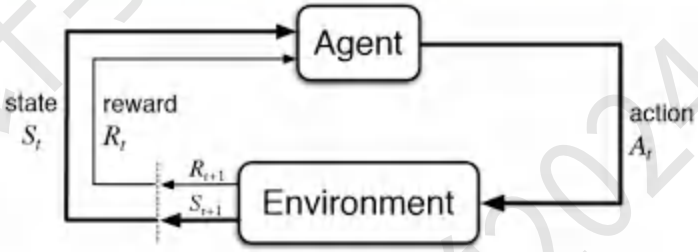
Generative adversarial network, GAN

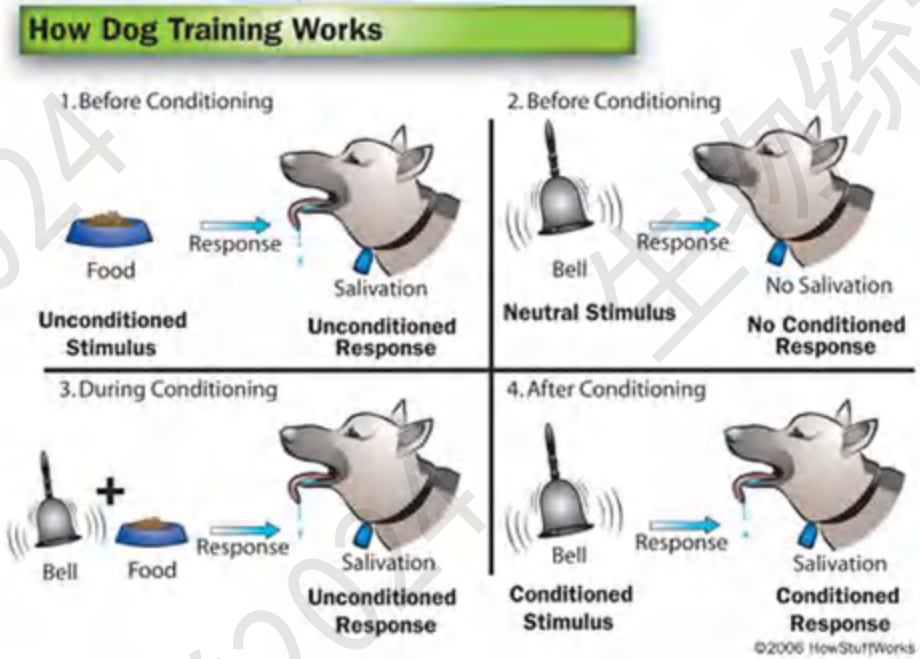


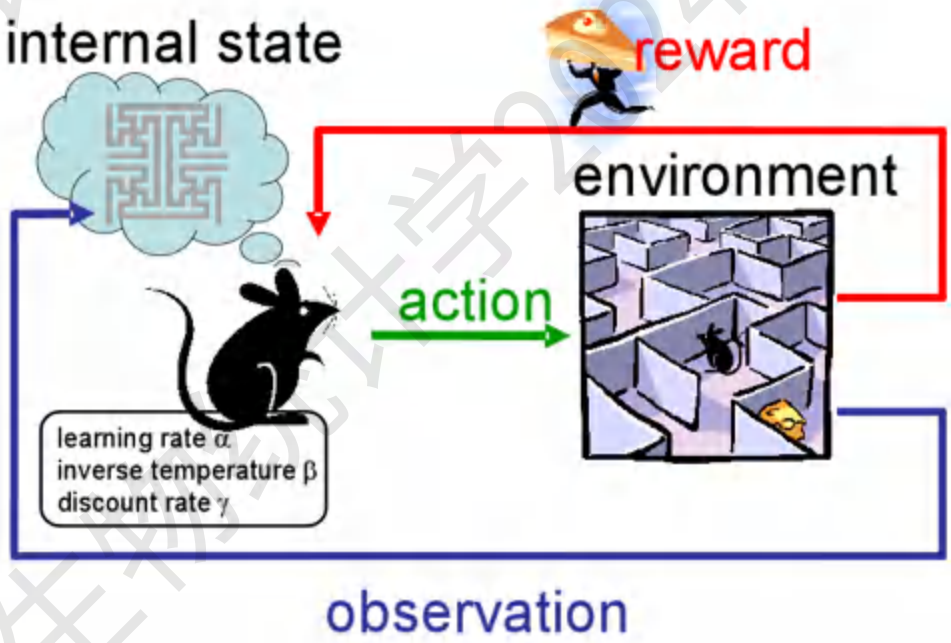


生物统计学2024

Reinforcement learning

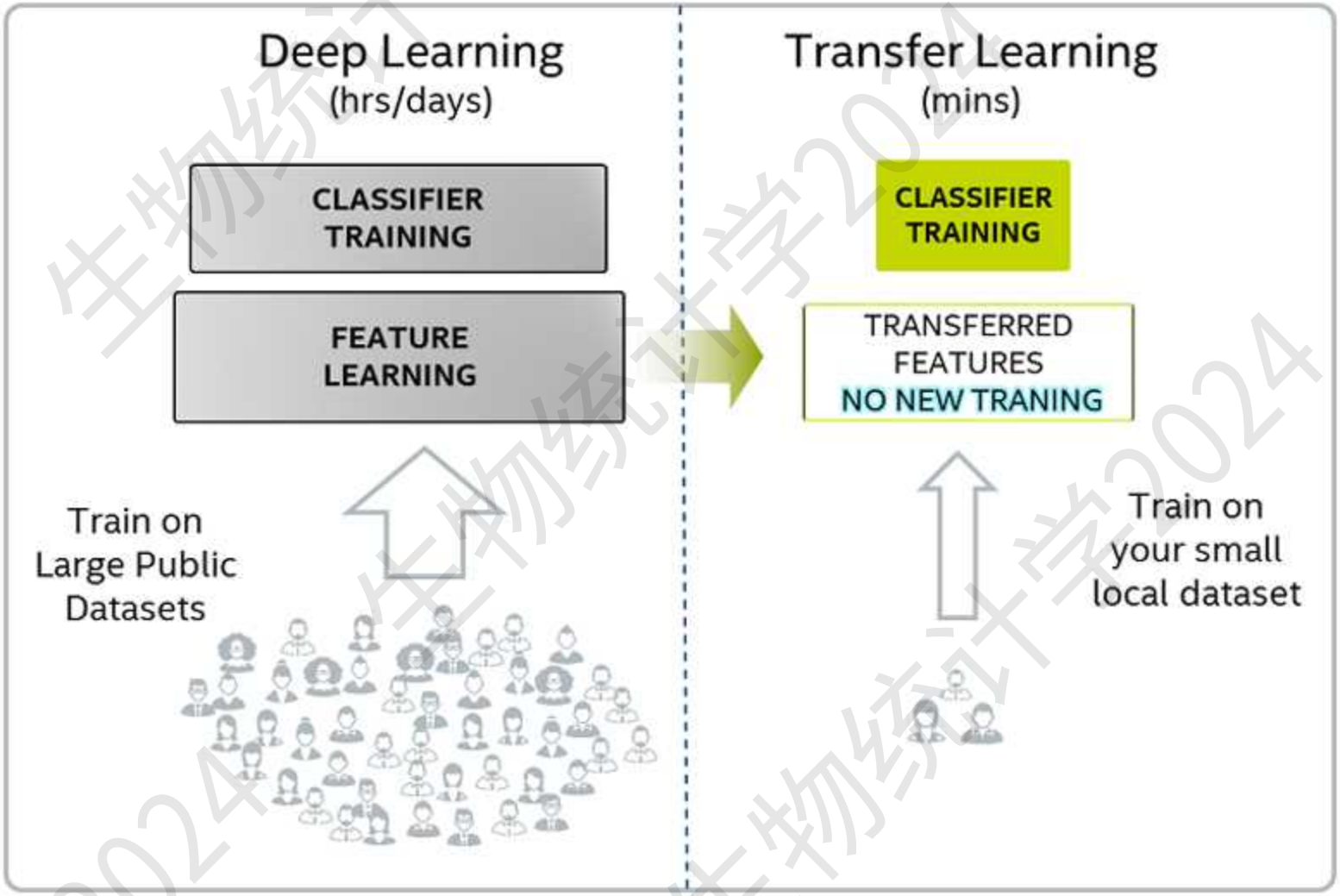






生物统计学2024

Transfer learning



生物统计学2024

深度学习的特点

深度学习常用算法介绍

深度学习常用框架介绍

生物统计学2024

深度学习

- 2006年，加拿大多伦多大学教授、机器学习领域的泰斗Geoffrey Hinton在《科学》上发表论文提出深度学习主要观点：

1）多隐层的人工神经网络具有优异的特征学习能力，学习得到的特征对数据有更本质的刻画，从而有利于可视化或分类；

2）深度神经网络在训练上的难度，可以通过“逐层初始化”（layer-wise pre-training）来有效克服，逐层初始化可通过无监督学习实现的。

生物统计学2024

深度学习

- 本质：通过构建多隐层的模型和海量训练数据（可为无标签数据），来学习更有用的特征，从而最终提升分类或预测的准确性。“深度模型”是手段，“特征学习”是目的。
- 与浅层学习区别：

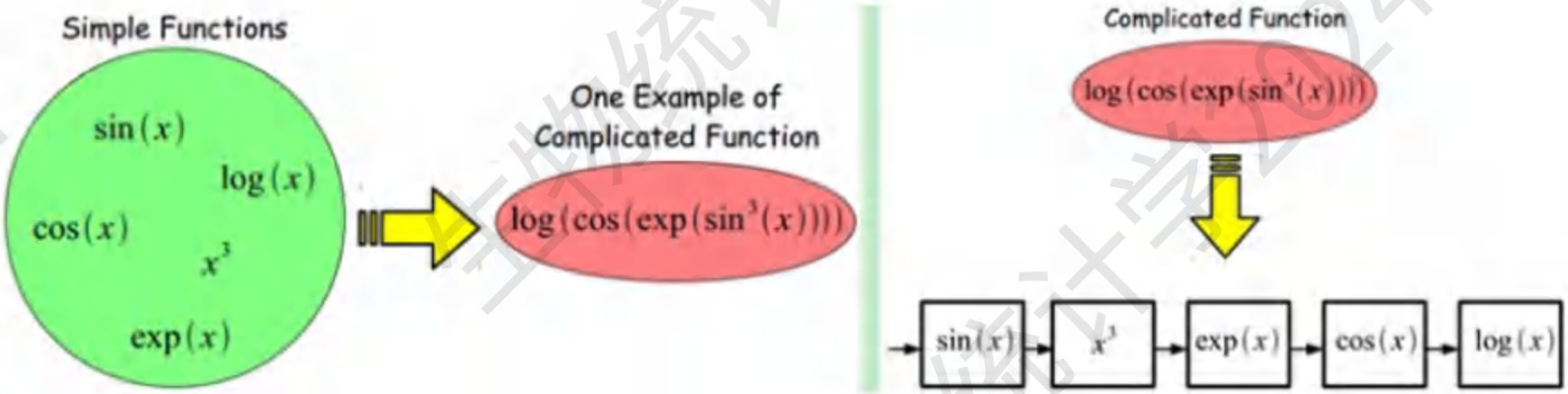
1）强调了模型结构的深度，通常有5-10多层的隐层节点；

2）明确突出了特征学习的重要性，通过逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间，从而使分类或预测更加容易。与人工规则构造特征的方法相比，利用大数据来学习特征，更能够刻画数据的丰富内在信息。

生物统计学2024

深度学习

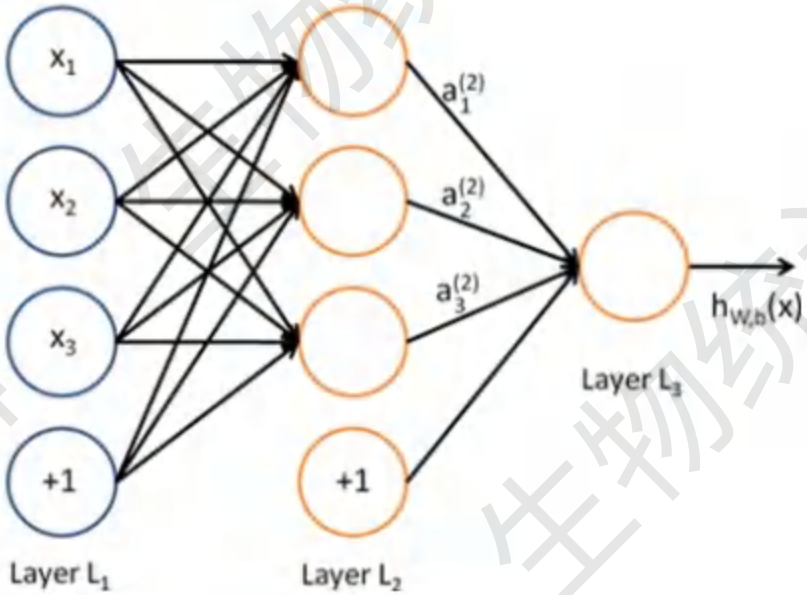
- 好处：可通过学习一种深层非线性网络结构，实现复杂函数逼近，表征输入数据分布式表示。



生物统计学2024

深度学习 vs. 神经网络

神经网络：



深度学习：

含多个隐层的深度学习模型

生物统计学2024

深度学习 vs. 神经网络

相同点： 二者均采用分层结构，系统包括输入层、隐层（多层）、输出层组成的多层网络，只有相邻层节点之间有连接，同一层以及跨层节点之间相互无连接，每一层可以看作是一个logistic 回归模型。

不同点：

神经网络：采用BP算法调整参数，即采用迭代算法来训练整个网络。随机设定初值，计算当前网络的输出，然后根据当前输出和样本真实标签之间的差去改变前面各层的参数，直到收敛；

深度学习：采用逐层训练机制。采用该机制的原因在于如果采用BP机制，对于一个deep network（7层以上），残差传播到最前面的层将变得很小，出现所谓的gradient diffusion（梯度扩散）。

深度学习 vs. 神经网络

- 神经网络的局限性：

1）比较容易过拟合，参数比较难调整，而且需要不少技巧；

2）训练速度比较慢，在层次比较少（小于等于3）的情况下效果并不比其它方法更优；

生物统计学2024

深度学习的特点

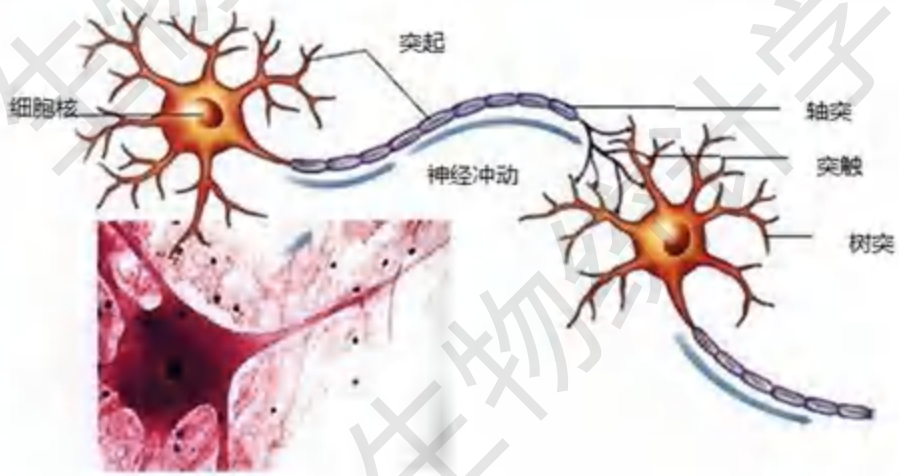
深度学习常用算法介绍

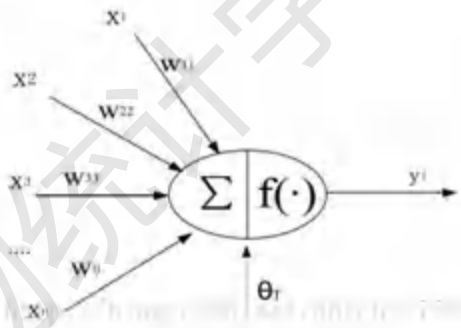
深度学习常用框架介绍

生物统计学2024

人工神经网络（ANN）

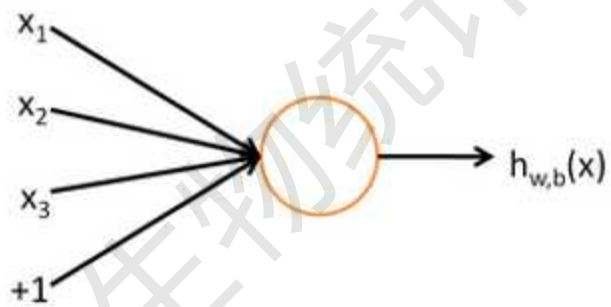
- 人工神经网络（Artificial Neural Networks）是一种模仿生物神经网络行为特征，进行分布式并行信息处理的算法数学模型。这种网络依靠系统的复杂程度，通过调整内部大量节点（神经元）之间相互连接的权重，从而达到处理信息的目的。



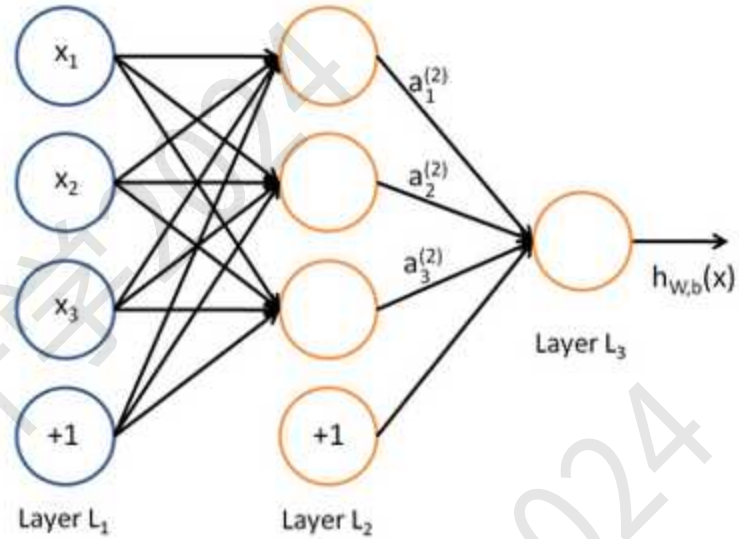


人工神经网络（ANN）

- 神经网络



$$h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{3} W_i x_i + b)$$



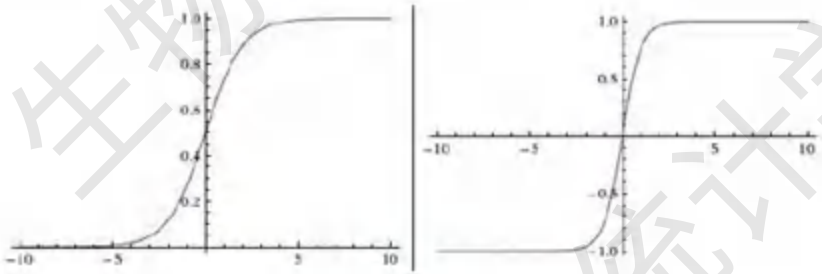
$$a_1^{(2)} = f(W_{11}^{(1)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)})$$
$$a_2^{(2)} = f(W_{21}^{(1)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)})$$
$$a_3^{(2)} = f(W_{31}^{(1)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(1)} x_3 + b_3^{(1)})$$
$$h_{W,b}(x) = a_1^{(3)} = f(W_{11}^{(2)} a_1^{(2)} + W_{12}^{(2)} a_2^{(2)} + W_{13}^{(2)} a_3^{(2)} + b_1^{(2)})$$

人工神经网络（ANN）

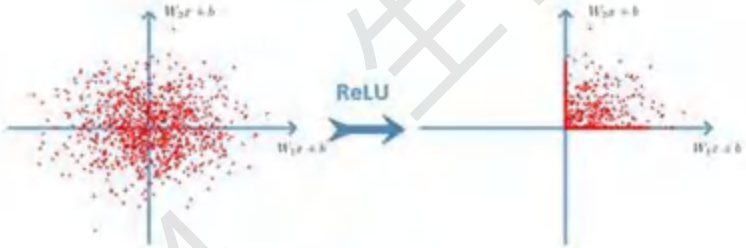
- 人工神经网络的重要概念：

  1 权值矩阵：相当于神经网络的记忆！在训练的过程中，动态调整和适应。

  2 激励函数：

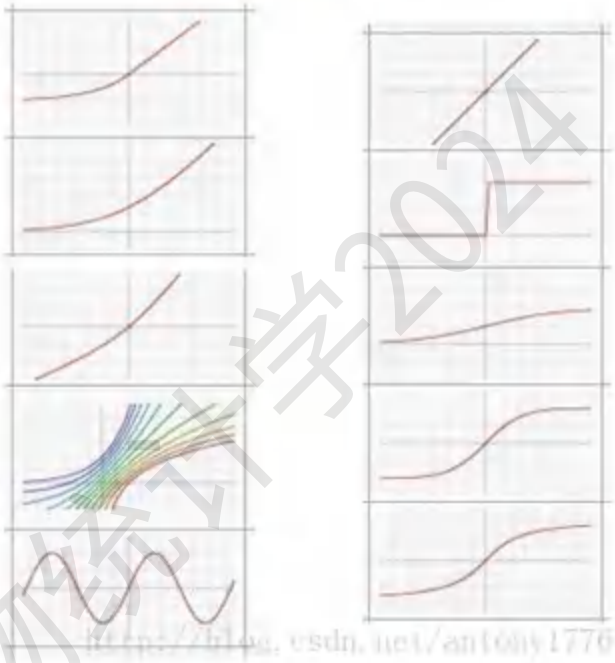


Sigmoid



Relu

Others



人工神经网络（ANN）

- 人工神经网络的重要概念：

激励函数很重要，无论是对建立神经网络的模型，还是理解神经网络。首先要了解，它有以下几个影响：

1 如何能更好的求解目标函数的极值！——高等数学中求解函数极值的知识！可微，单调！

2 如何提升训练效率，让梯度的优化方法更稳定；

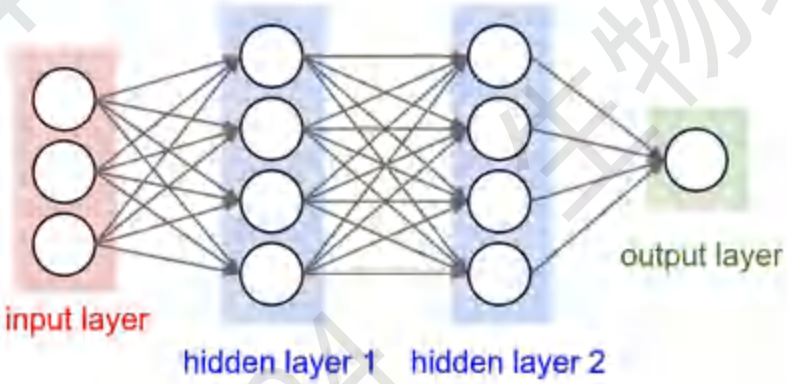
3 权值的初始值，不影响训练结果！

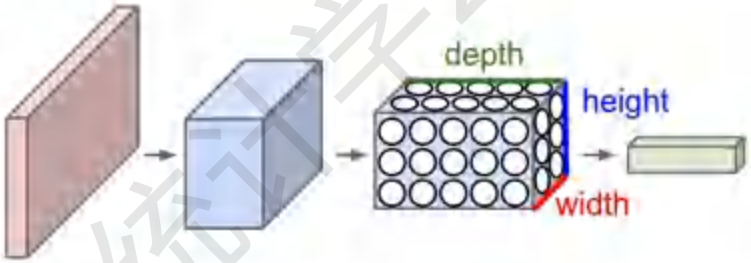
生物统计学2024

卷积神经网络（CNN）

- 卷积神经网络（Convolutional Neural Networks / CNNs / ConvNets）与普通神经网络非常相似，它们都由具有可学习的权重和偏置常量(biases)的神经元组成。每个神经元都接收一些输入，并做一些点积计算，输出是每个分类的分数，普通神经网络里的一些计算技巧到这里依旧适用。

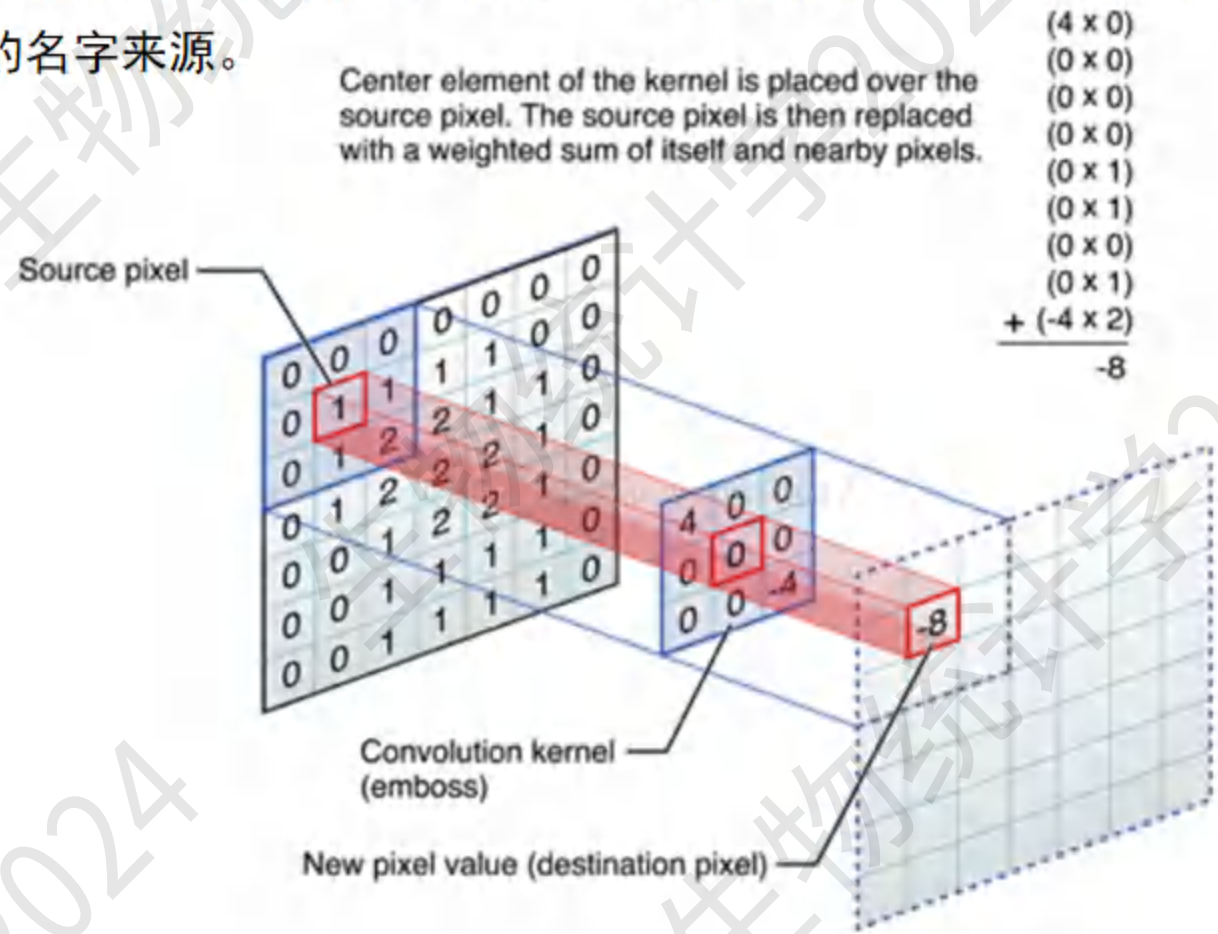
- 与普通神经网络不同之处：卷积神经网络默认输入是图像，可以让我们把特定的性质编码入网络结构，使是我们的前馈函数更加有效率，并减少了大量参数。



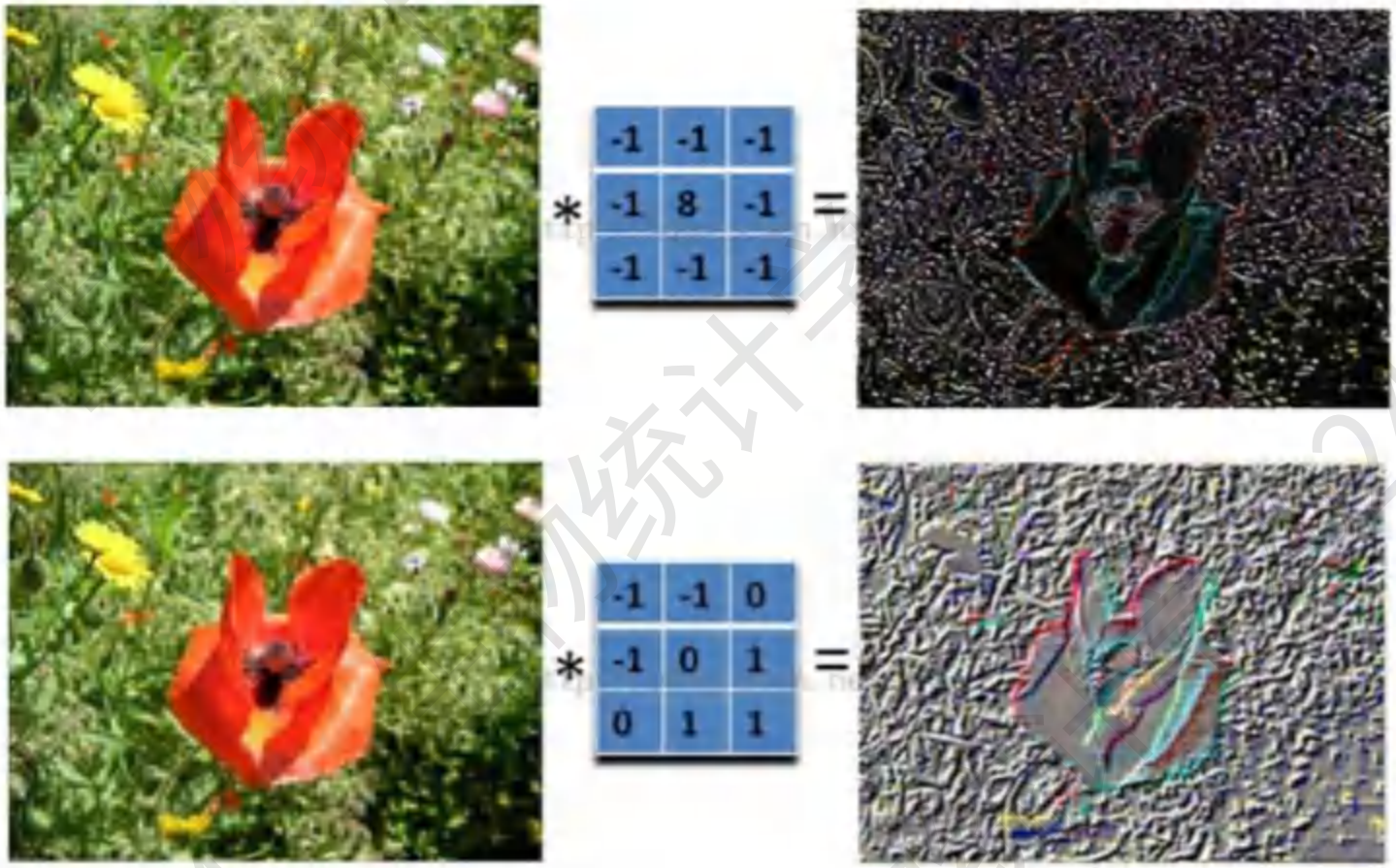


卷积神经网络（CNN）

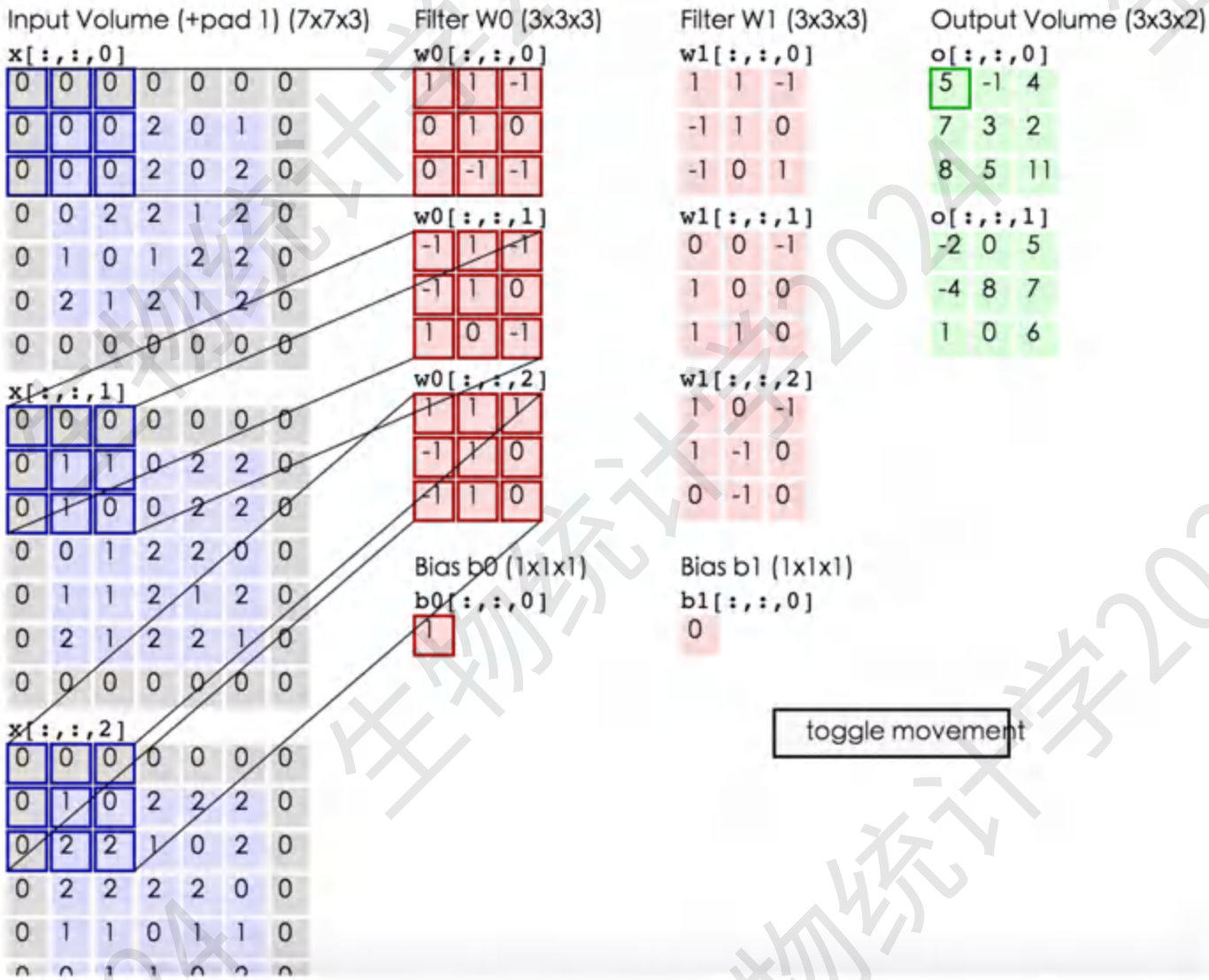
- 卷积操作：对图像（不同的数据窗口数据）和滤波矩阵（一组固定的权重：因为每个神经元的权重固定，所以又可以看做一个恒定的滤波器filter）做内积（逐个元素相乘再求和）的操作就是所谓的『卷积』操作，也是卷积神经网络的名字来源。



卷积神经网络（CNN）



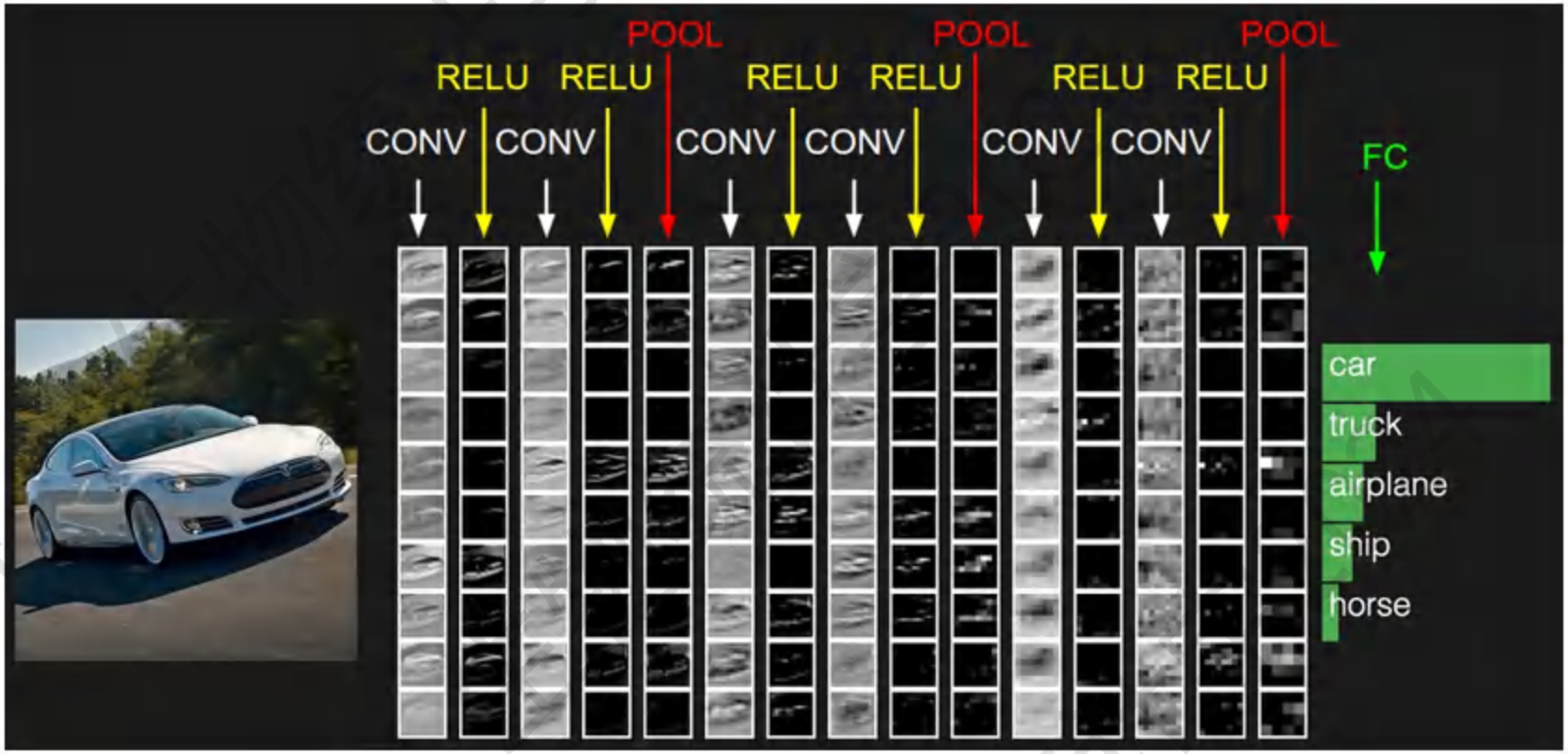
卷积神经网络（CNN）


Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 0 1 0
0 0 0 2 0 2 0
0 0 2 2 1 2 0
0 1 0 1 2 2 0
0 2 1 2 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 1 1 0 2 2 0
0 1 0 0 2 2 0
0 0 1 2 2 0 0
0 1 1 2 1 2 0
0 2 1 2 2 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 1 0 2 2 2 0
0 2 2 1 0 2 0
0 2 2 2 2 0 0
0 1 1 0 1 1 0
Filter W0 (3x3x3)
w0[:,:,0]
1 1 -1
0 1 0
0 -1 -1
w0[:,:,1]
-1 1 -1
-1 1 0
1 0 -1
w0[:,:,2]
1 1 1
-1 1 0
-1 1 0
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 1 -1
-1 1 0
-1 0 1
w1[:,:,1]
0 0 -1
1 0 0
1 1 0
w1[:,:,2]
1 0 -1
1 -1 0
0 -1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
5 -1 4
7 3 2
8 5 11
o[:,:,1]
-2 0 5
-4 8 7
1 0 6
toggle movement

卷积神经网络（CNN）

- 卷积层（Convolutional layer），卷积神经网路中每层卷积层由若干卷积单元组成，每个卷积单元的参数都是通过反向传播算法优化得到的。卷积运算的目的是提取输入的不同特征，第一层卷积层可能只能提取一些低级的特征如边缘、线条和角等层级，更多层的网络能从低级特征中迭代提取更复杂的特征。
- 线性整流层（Rectified Linear Units layer, ReLU layer），这一层神经的激励函数（Activation function）使用线性整流（Rectified Linear Units, ReLU）f(x)=max(0,x)。
- 池化层（Pooling layer），通常在卷积层之后会得到维度很大的特征，将特征切成几个区域，取其最大值或平均值，得到新的、维度较小的特征。
- 全连接层（ Fully-Connected layer），把所有局部特征结合变成全局特征，用来计算最后每一类的得分。

卷积神经网络（CNN）


POOL
POOL
POOL
RELU
RELU
RELU
RELU
RELU
RELU
CONV
CONV
CONV
CONV
CONV
CONV
FC
car
truck
airplane
ship
horse

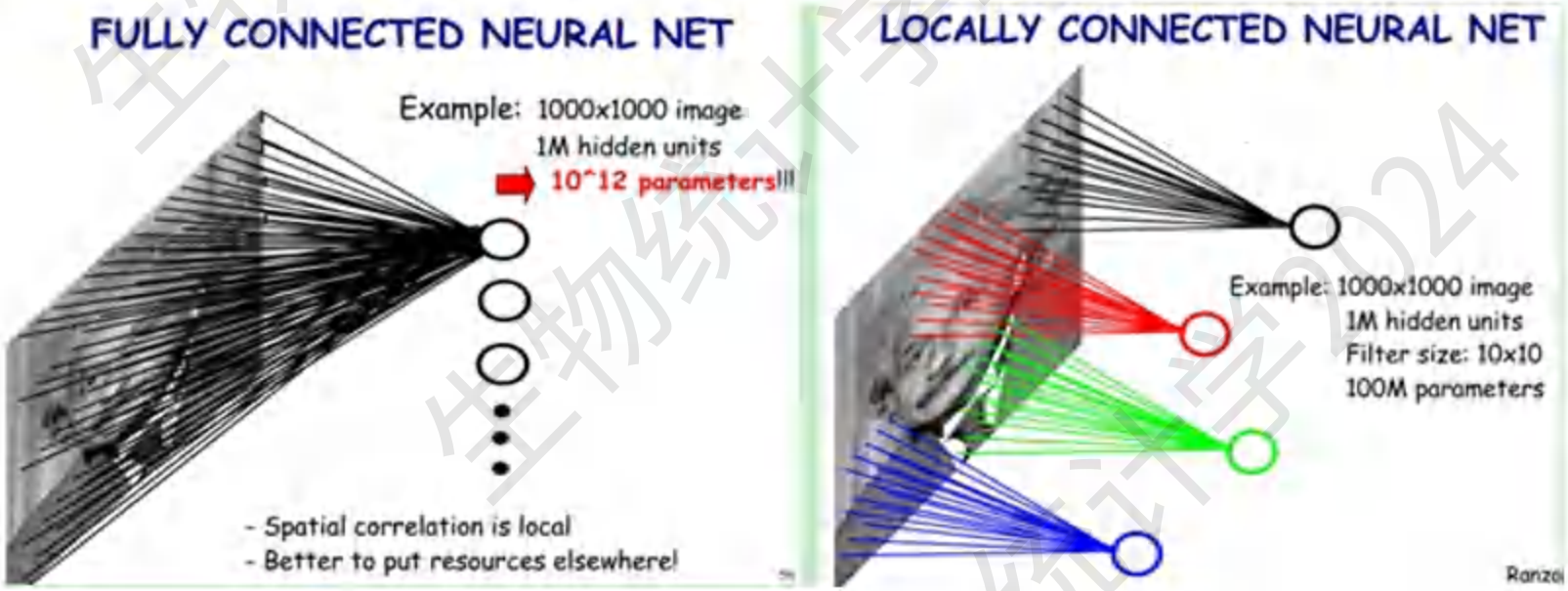
卷积神经网络（CNN）

- 在图像处理中，往往把图像表示为像素的向量，比如一个1000×1000的图像，可以表示为一个1000000的向量。在上一节中提到的神经网络中，如果隐含层数目与输入层一样，即也是1000000时，那么输入层到隐含层的参数数据为1000000×1000000=10^12，这样就太多了，基本没法训练。所以图像处理要想练成神经网络大法，必先减少参数加快速度。就跟辟邪剑谱似的，普通人练得很挫，一旦自宫后内力变强剑法变快，就变的很牛了。

生物统计学2024

卷积神经网络（CNN）

- 卷积神经网络有两种神器可以降低参数数目，第一种神器叫做局部感知。
- 在下方右图中，假如每个神经元只和10×10个像素值相连，那么权值数据为1000000×100个参数，减少为原来的万分之一。而那10×10个像素值对应的10×10个参数，其实就相当于卷积操作。

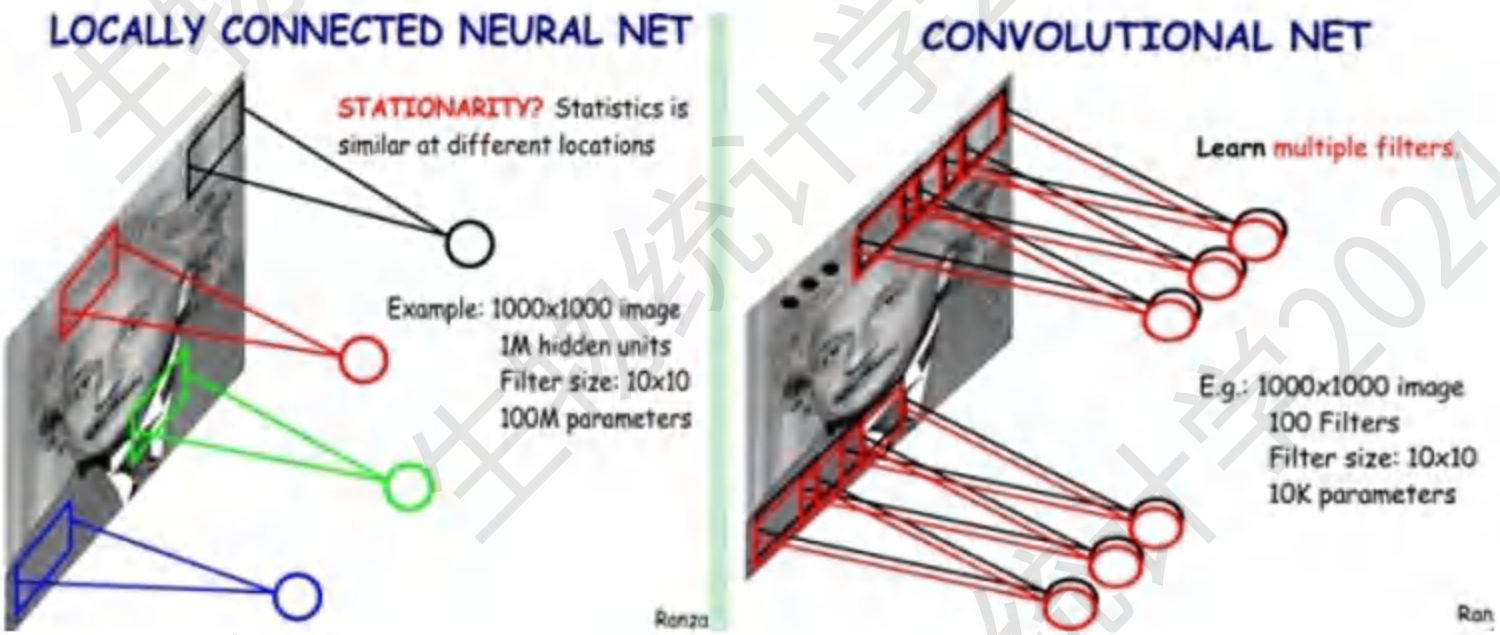


卷积神经网络（CNN）

- 这样的话参数仍然过多，那么就启动第二级神器，即权值共享。在上面的局部连接中，每个神经元都对应100个参数，一共1000000个神经元，如果这1000000个神经元的100个参数都是相等的，那么参数数目就变为100了。
- 怎么理解权值共享呢？我们可以这100个参数（也就是卷积操作）看成是提取特征的方式，该方式与位置无关。这其中隐含的原理则是：图像的一部分的统计特性与其他部分是一样的。这也意味着我们在这一部分学习的特征也能用在另一部分上，所以对于这个图像上的所有位置，我们都能使用同样的学习特征。

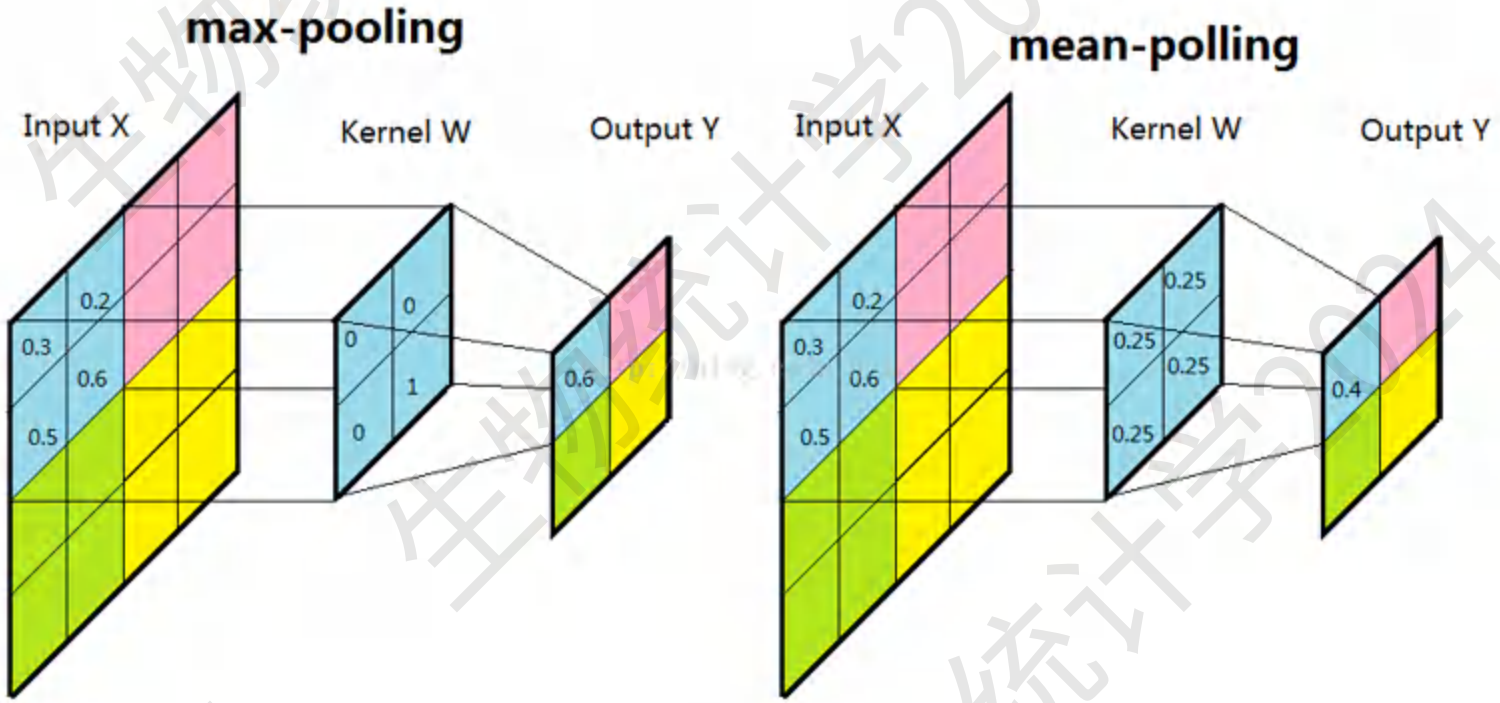
卷积神经网络（CNN）

- 上面所述只有100个参数时，表明只有1个100*100的卷积核，显然，特征提取是不充分的，我们可以添加多个卷积核，比如32个卷积核，可以学习32种特征。在有多个卷积核时，如下图所示：



卷积神经网络（CNN）

- 池化，也称作下采样，可以实现降维。常用有最大值池化和均值池化。



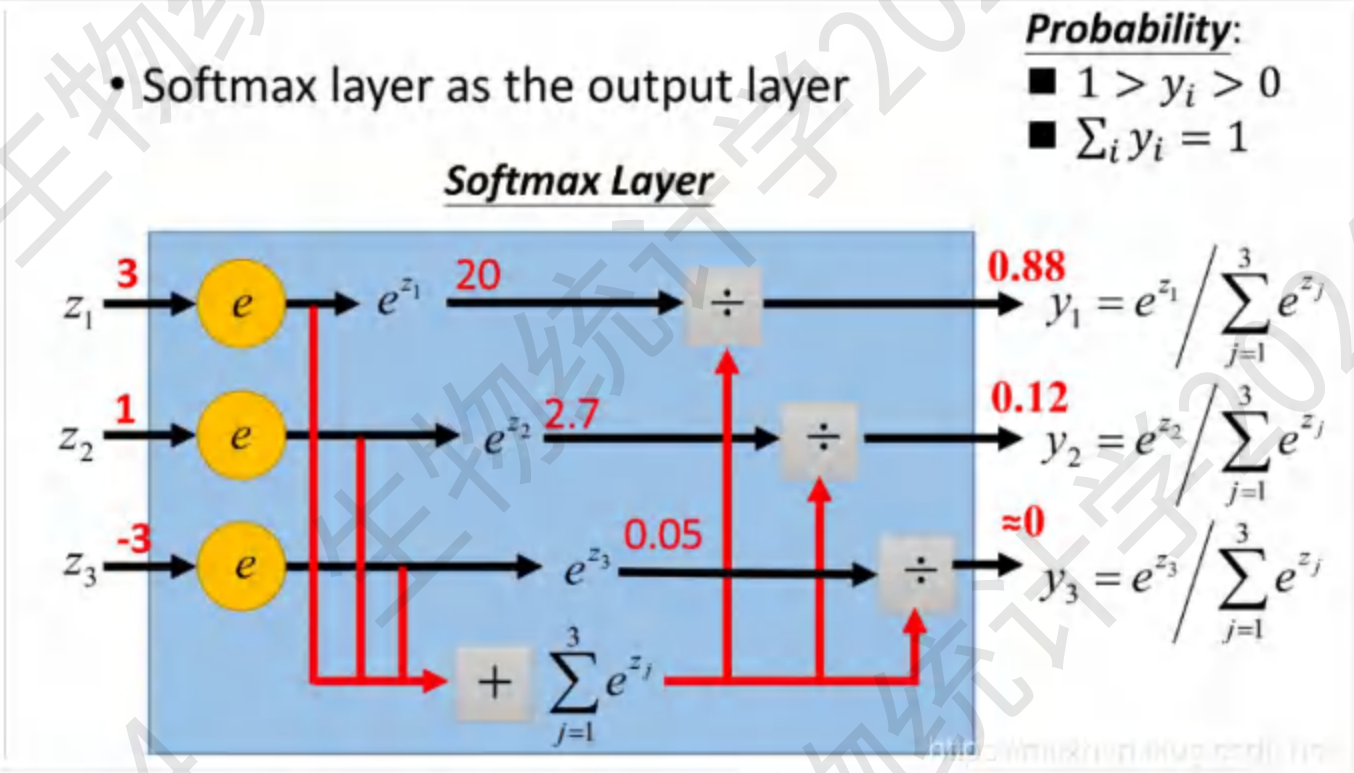
卷积神经网络（CNN）

- 全连接层：连接所有的特征，将输出值送给分类器（如softmax分类器），最终得出识别结果。

生物统计学2024

softmax分类器

$$softmax(x_0) = \frac{e^{x_0}}{e^{x_0} + e^{x_1} + e^{x_2}}$$

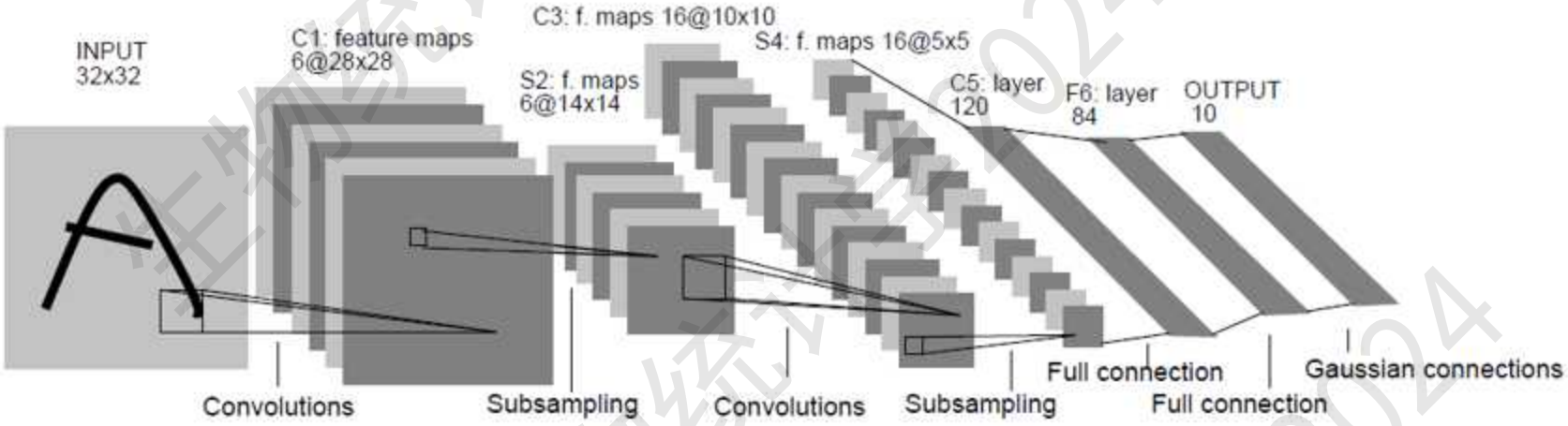


softmax对神经元的输出信号进行加工，输出为分类的概率值。

生物统计学2024

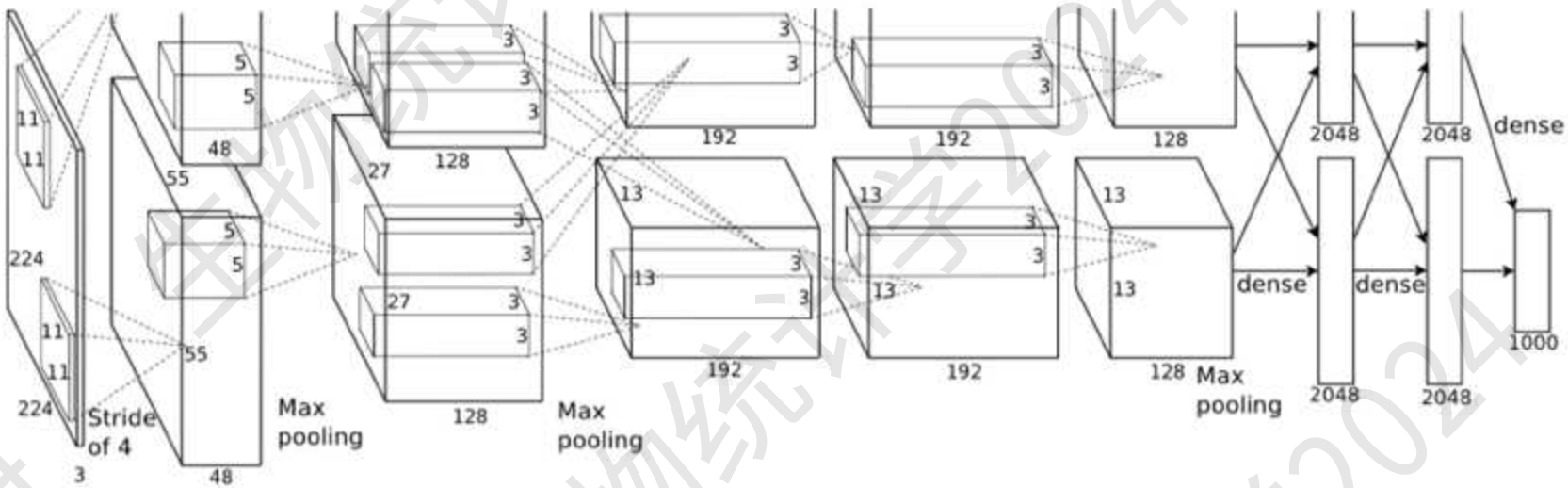
常见网络模型

- LeNet



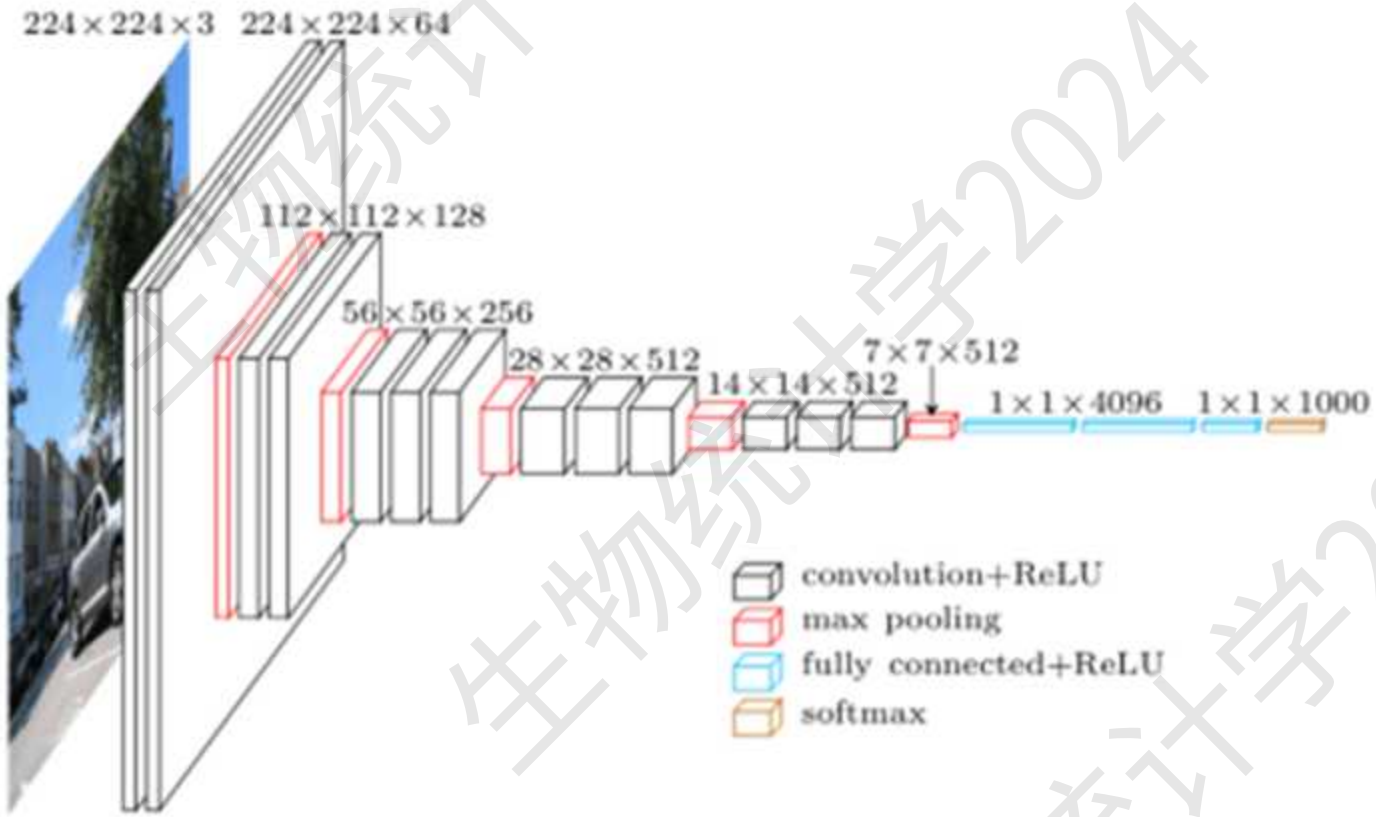
常见网络模型

- AlexNet



常见网络模型

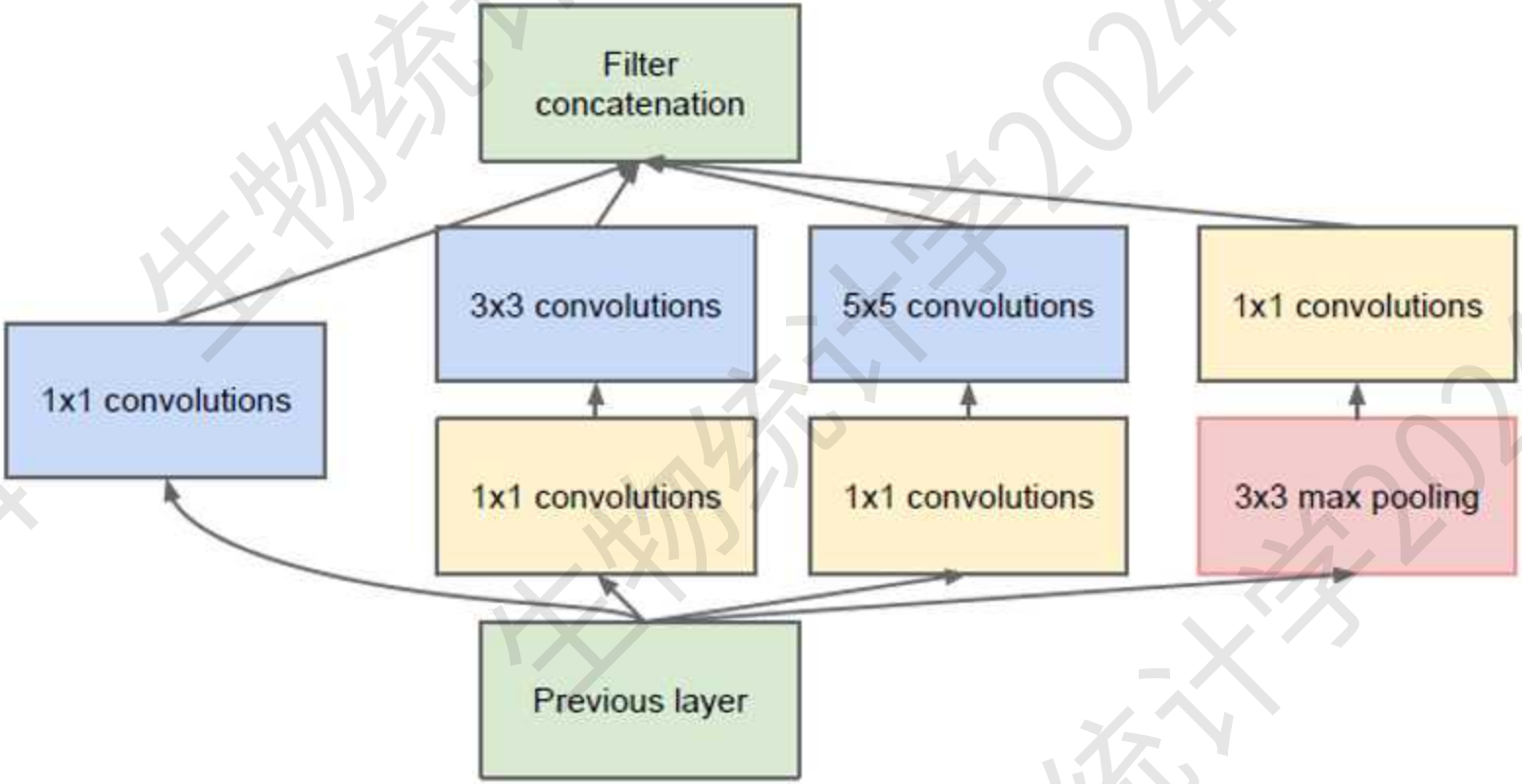
- VGG16



生物统计学2024

常见网络模型

- GoogleNet （InceptionV4）



常见网络模型

CNN模型比较.doc

- 比较

模型名	AlexNet	VGG	GoogLeNet	ResNet
初入江湖	2012	2014	2014	2015
层数	8	19	22	152
Top-5错误	16.4%	7.3%	6.7%	3.57%
Data Augmentation	+	+	+	+
Inception(NIN)	-	-	+	-
卷积层数	5	16	21	151
卷积核大小	11,5,3	3	7,1,3,5	7,1,3,5
全连接层数	3	3	1	1
全连接层大小	4096,4096,1000	4096,4096,1000	1000	1000
Dropout	+	+	+	+
Local Response Normalization	+	-	+	-
Batch Normalization	-	-	-	+

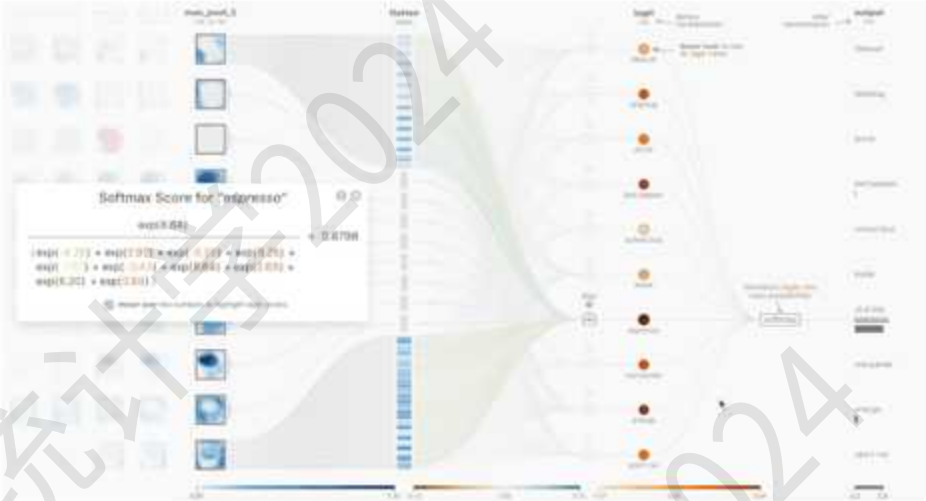
Re-cap

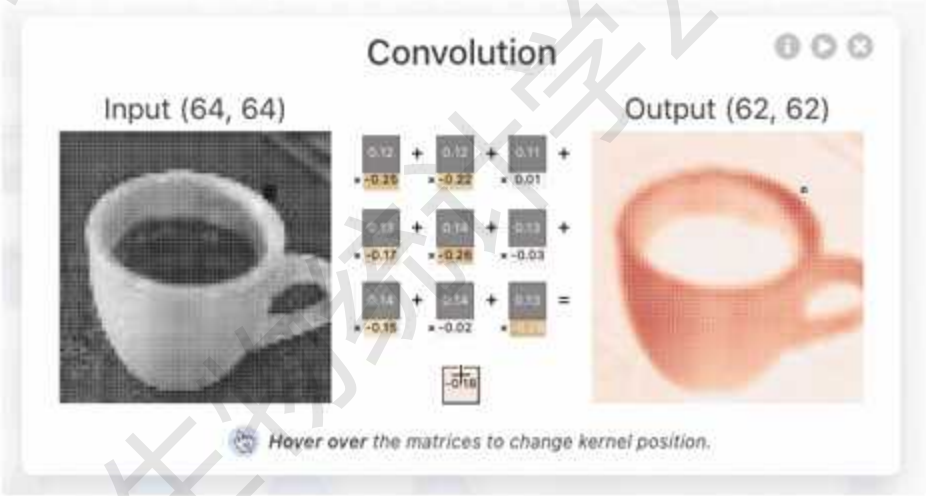
Overall steps

生物统计学2024

Re-cap

convolution, softmax, parameters...





生物统计学2024

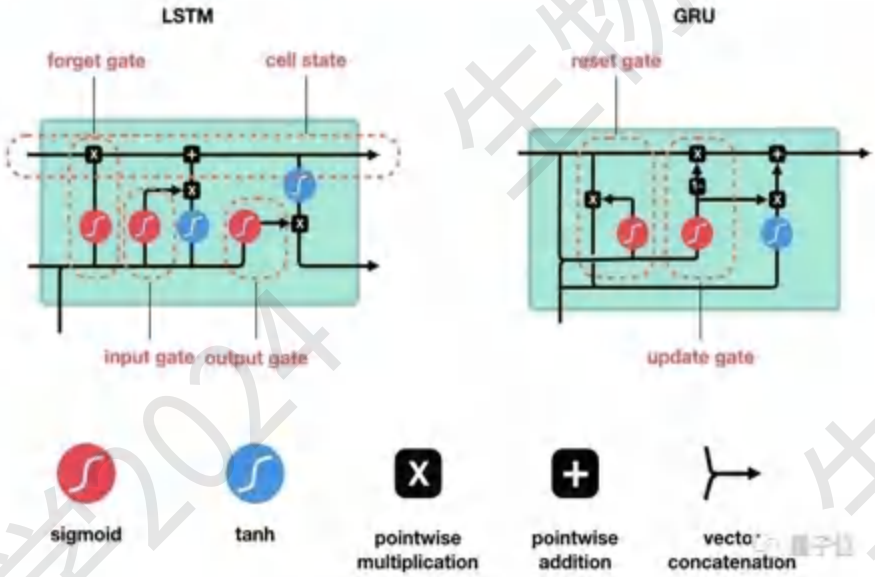
Re-cap

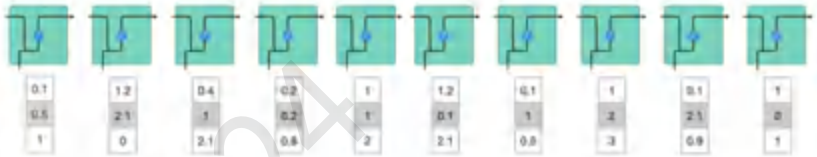
pooling and output

生物统计学2024

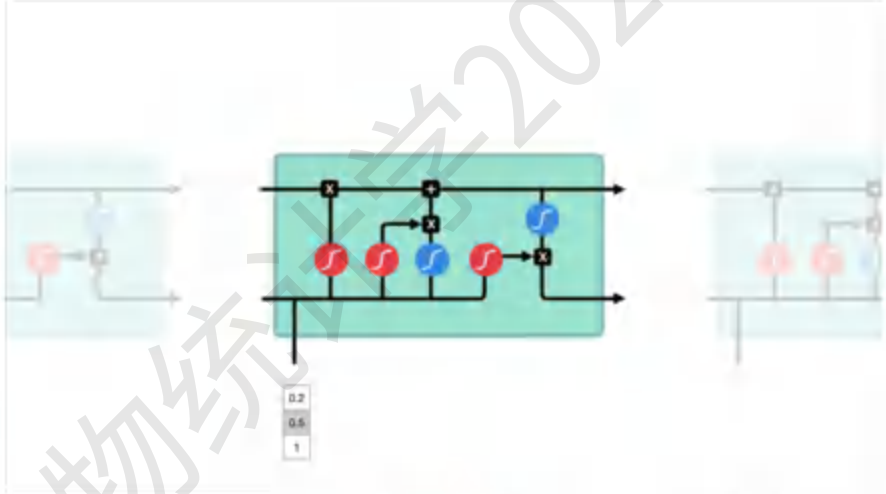
其他深度学习算法

- 自动编码器（AutoEncoder）
- 稀疏编码（Sparse Coding）
- 限制玻尔兹曼机（RBM）
- 。 。 。
- 循环神经网络（RNN）
- 门控循环单元（GRU）
- 长短期记忆网络（LSTM）
- BERT
- Transformer





RNN



LSTM

生物统计学2024

Transformer

- Large Language Model，大语言模型

https://ig.ft.com/generative-ai/ https://mp.weixin.qq.com/s/KGU3uekq585dTCel9y9

生物统计学2024

Transformer

- Large Language Model，大语言模型

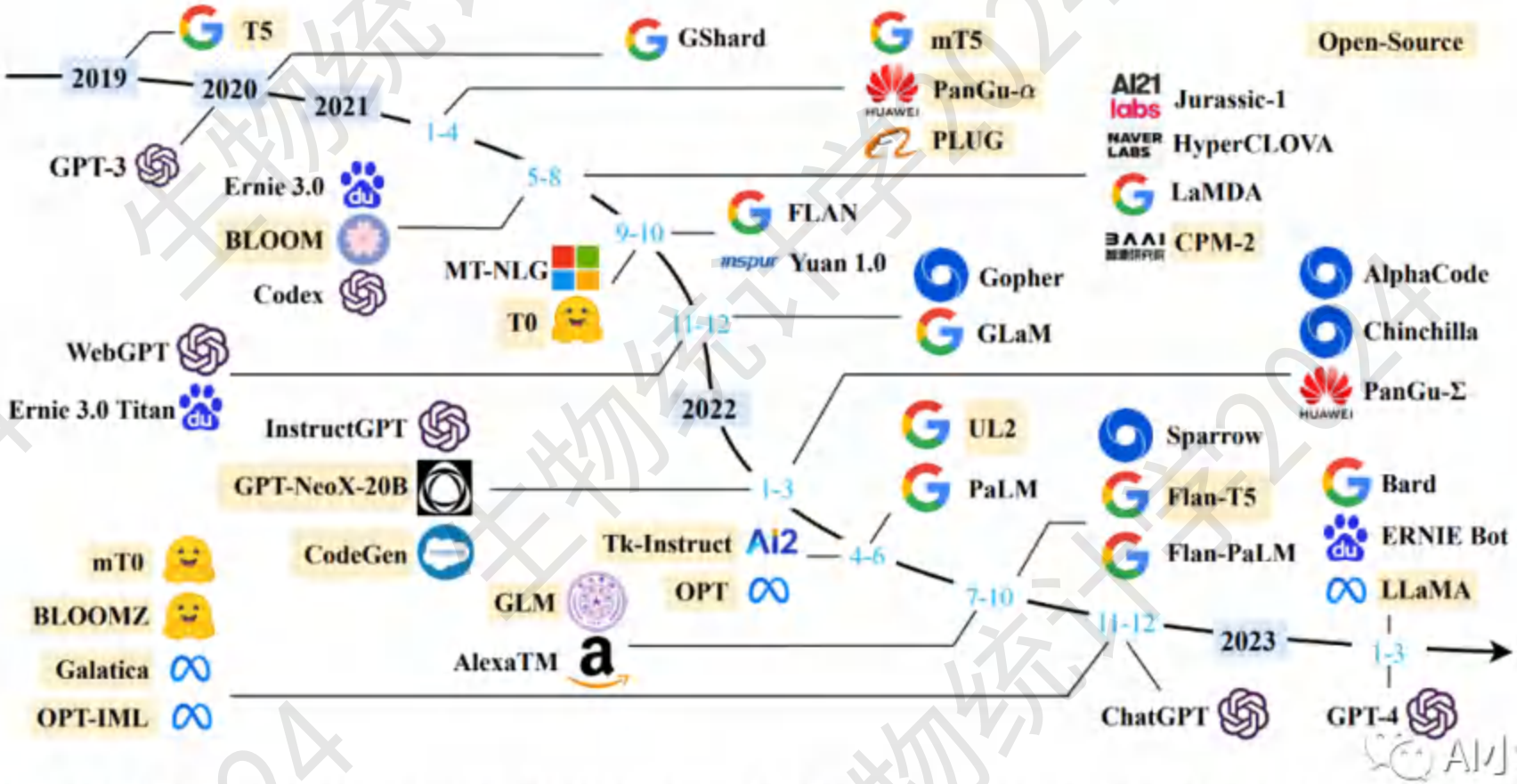
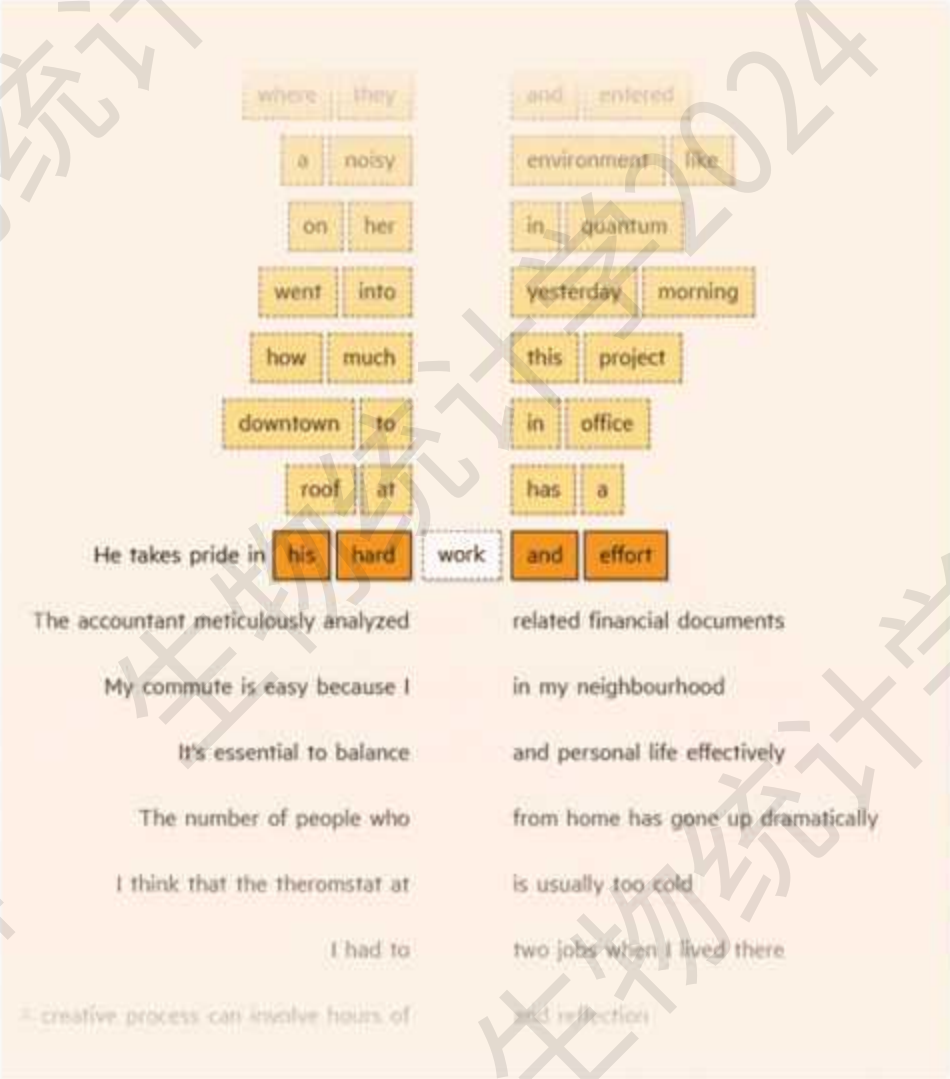


Fig. 1. A timeline of existing large language models (having a size larger than 10B) in recent years. We mark the open-source LLMs in yellow color.

Transformer

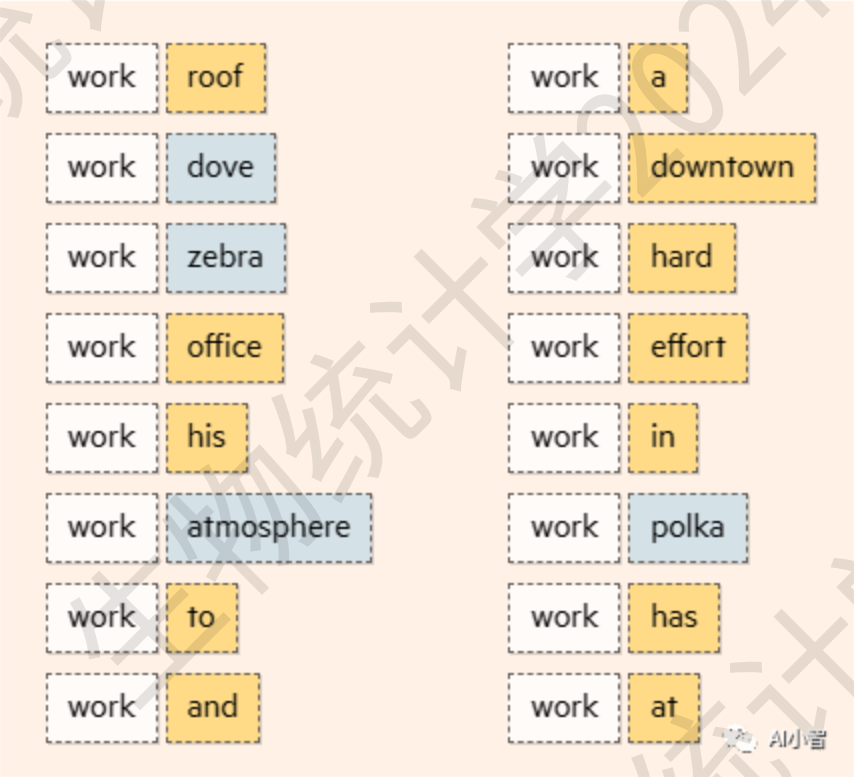
- 令牌(tokens) ——可以编码的基本单位



生物统计学2024

Transformer

- 令牌(tokens) ——可以编码的基本单位



生物统计学2024

Transformer

- 词嵌入(embedding) ——每个词在训练数据中与work的邻近程度

生物统计学2024

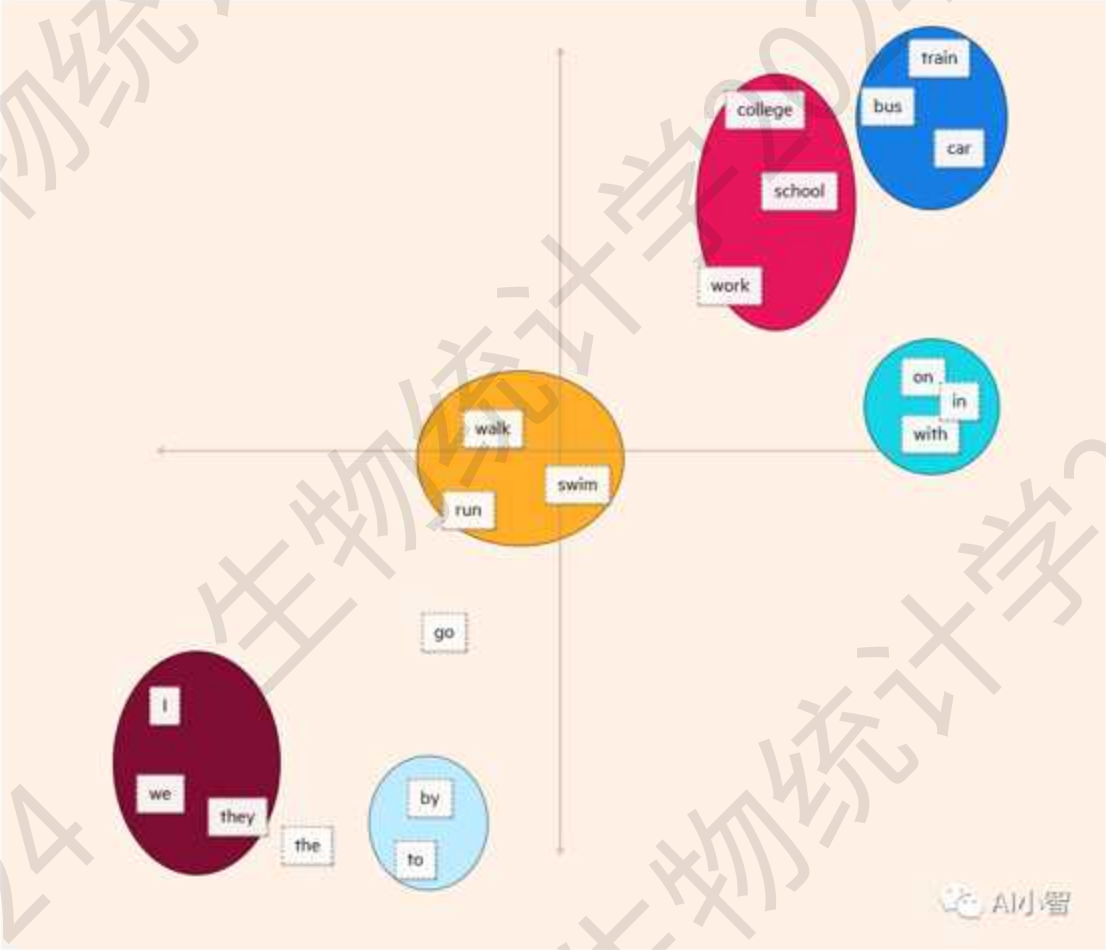
Transformer

- 词嵌入(embedding) ——每个词在训练数据中与work的邻近程度

生物统计学2024

Transformer

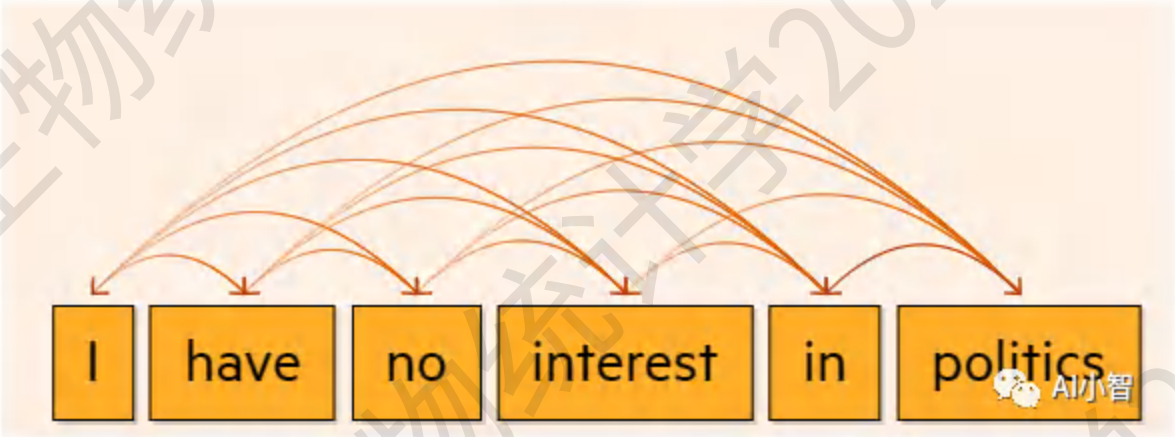
- 词之间的距离（降维分析）

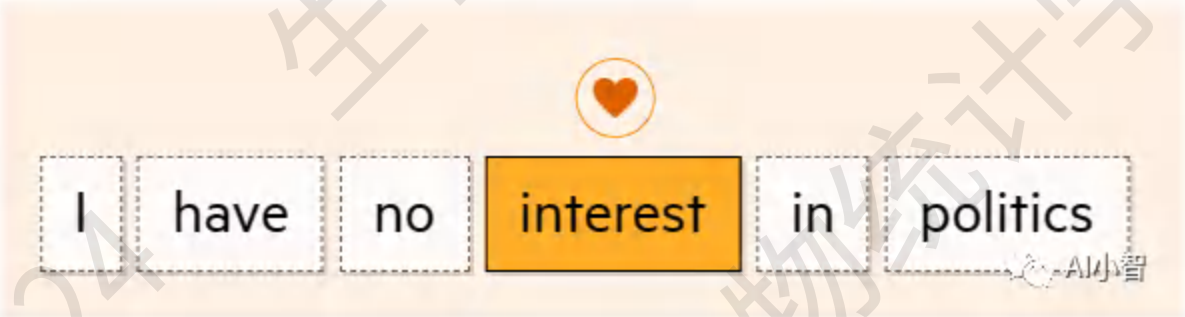


生物统计学2024

Transformer

- 自注意力(Attention)——允许LLM理解词之间关系
- 自注意力查看文本中的每个令牌(token)，并决定哪些对理解其含义最重要





生物统计学2024

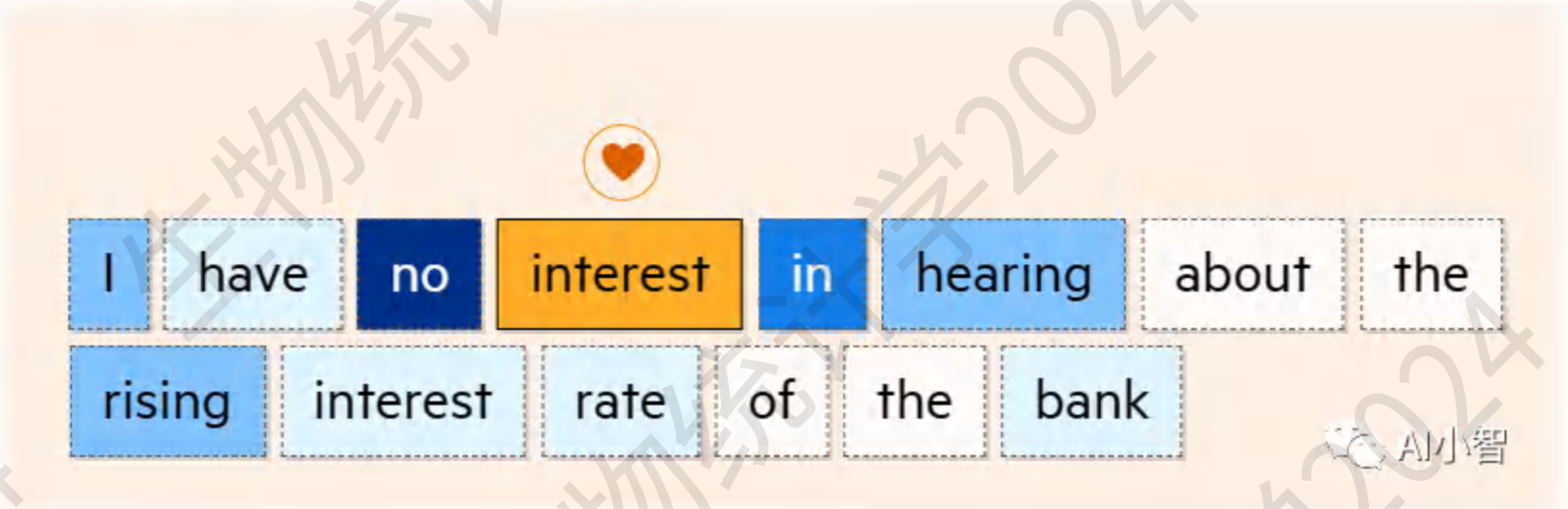
Transformer

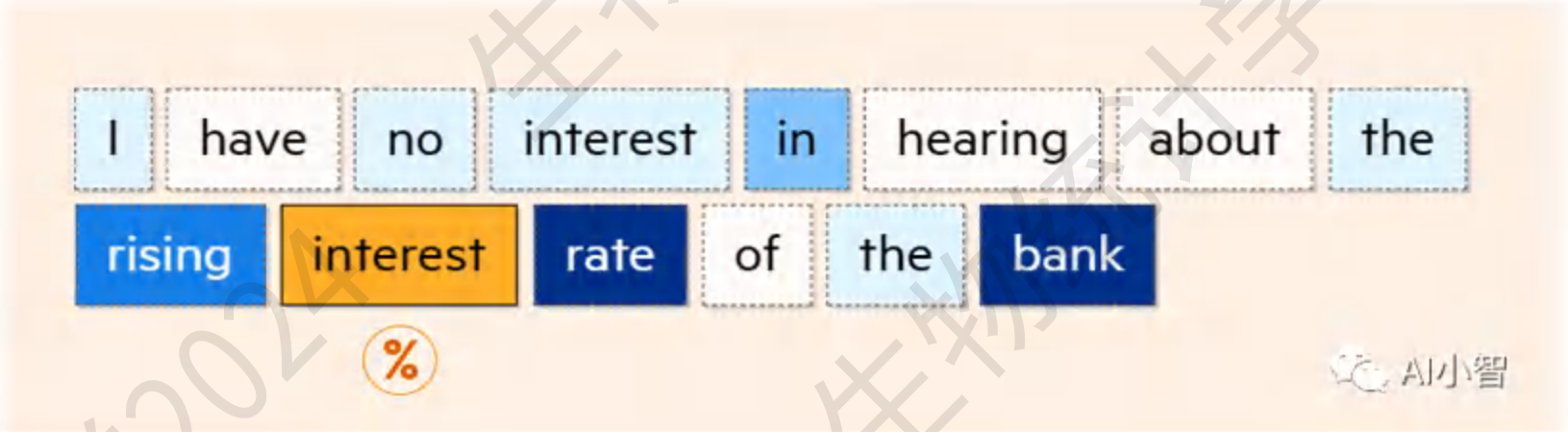
- 自注意力(Attention)——允许LLM理解词之间关系
- 自注意力查看文本中的每个令牌(token)，并决定哪些对理解其含义最重要

生物统计学2024

Transformer

- 组合这些句子时,模型仍然能够识别每个词的正确含义
- 归功于它对伴随文本的注意力

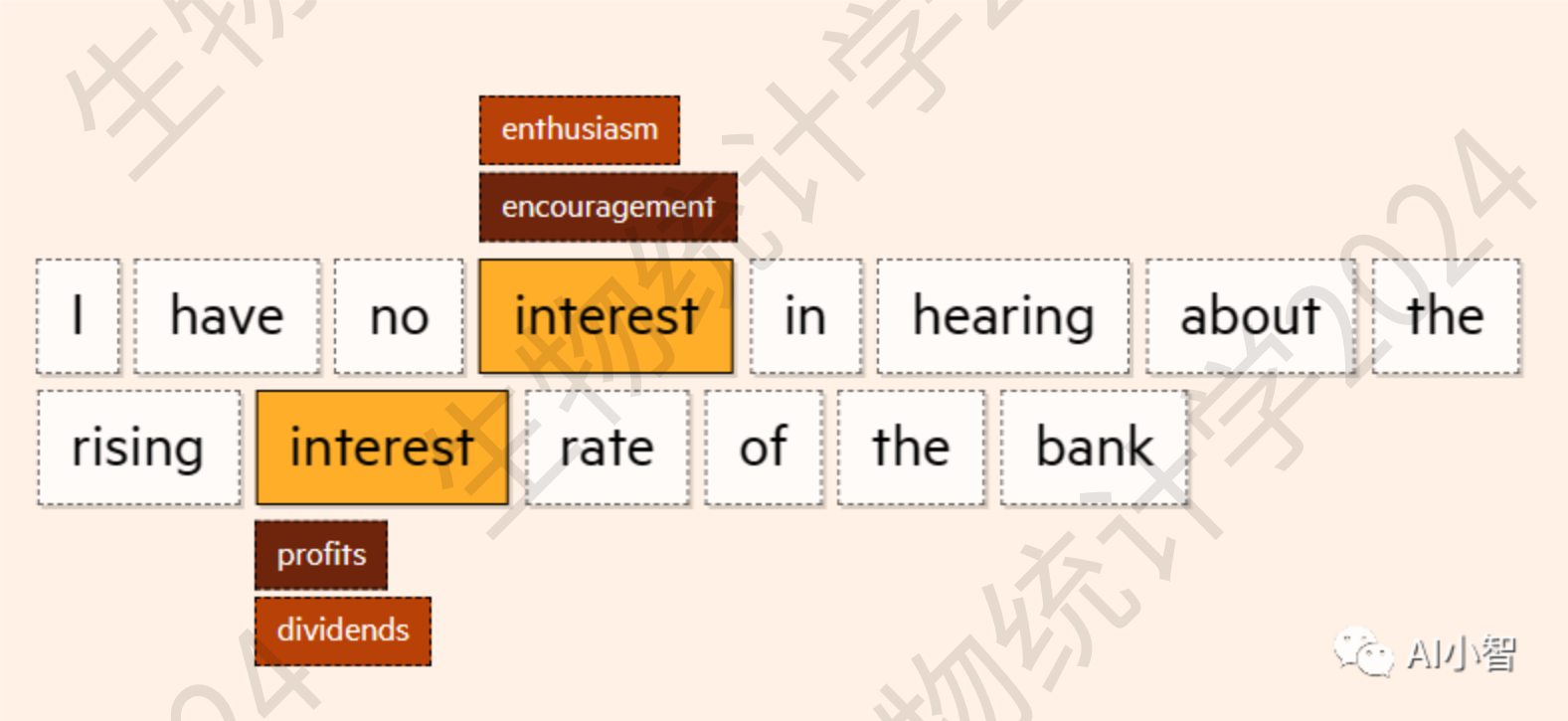




生物统计学2024

Transformer

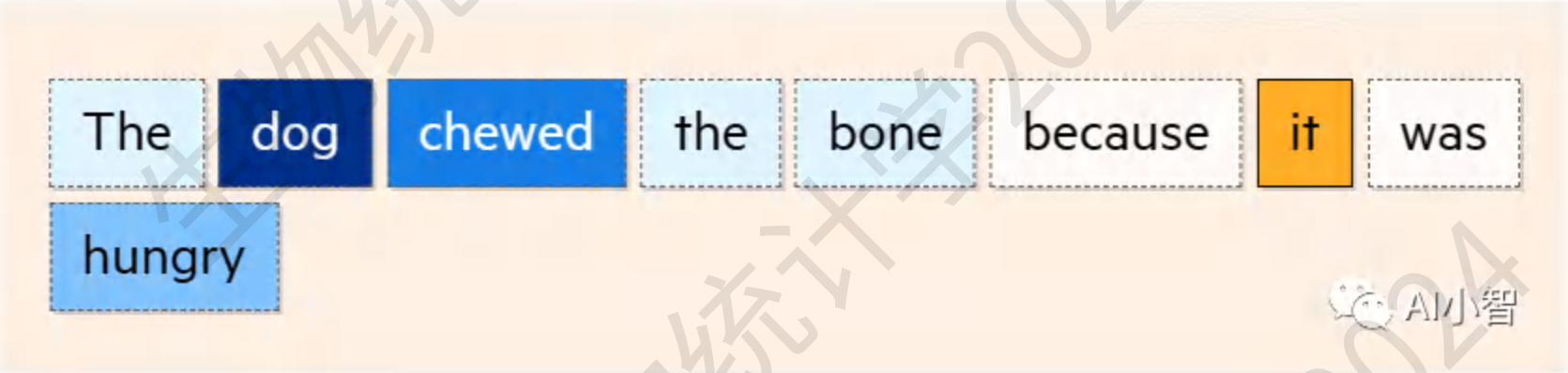
- 在某些上下文中可以互换的词，可以被互换
- 在其他上下文中不可以的词，不会被错误使用

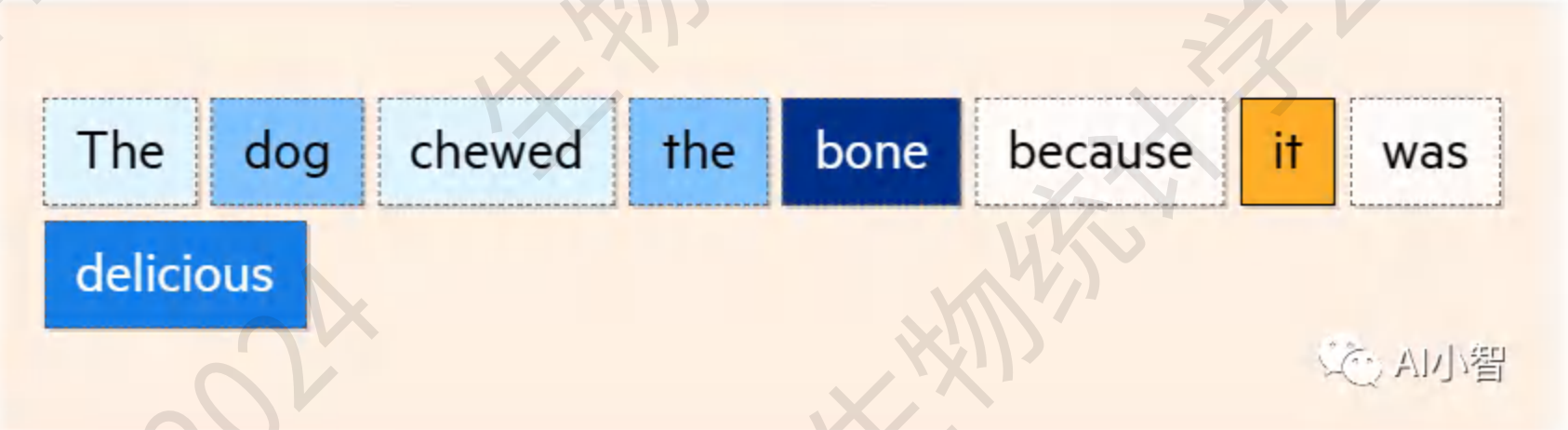


生物统计学2024

Transformer

- 在某些上下文中可以互换的词，可以被互换
- 在其他上下文中不可以的词，不会被错误使用





生物统计学2024

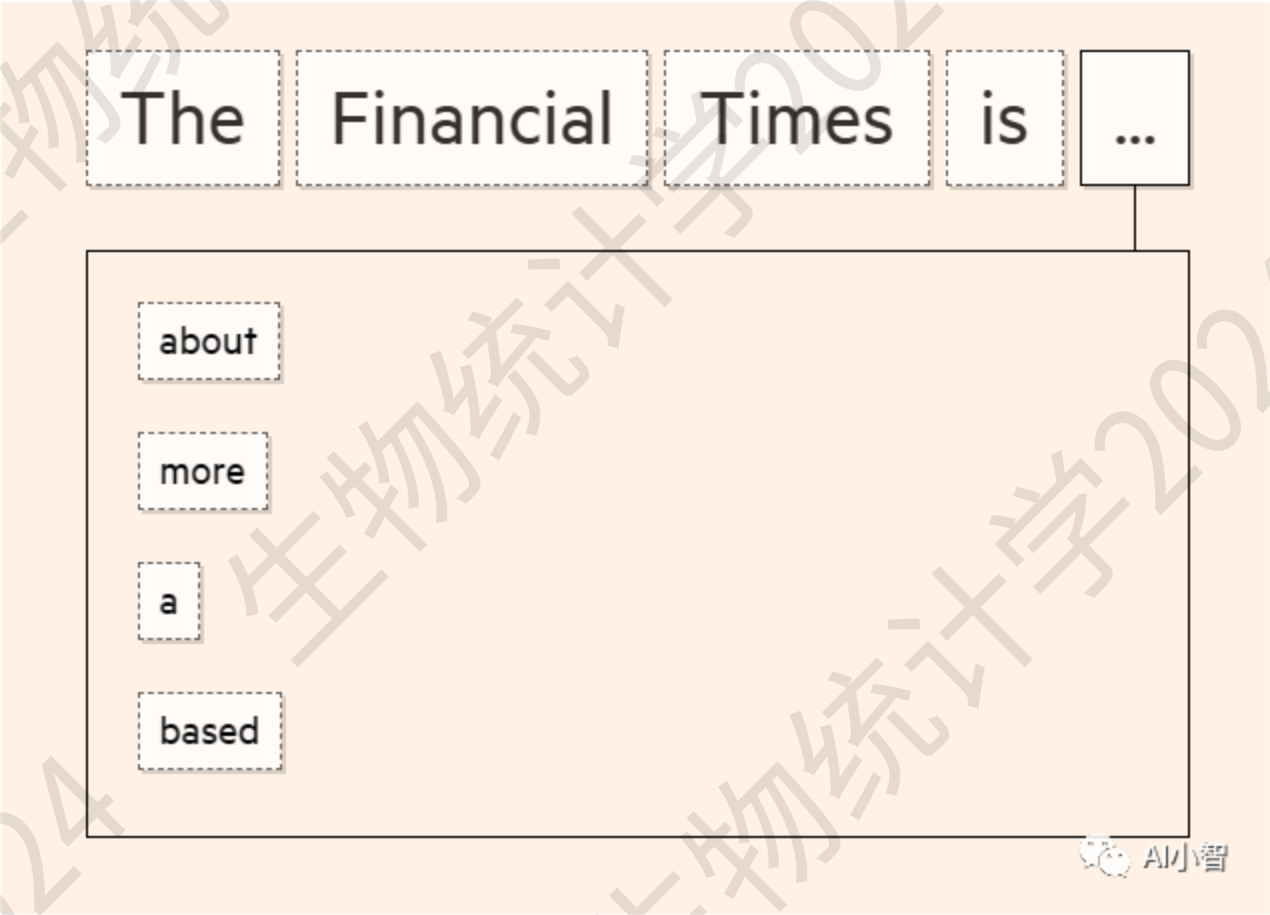
Transformer

- 从这巨大的语料库中,模型学会识别模式,最终预测下一个最佳选项
- 涌现智能——成为最像人的人工智能

生物统计学2024

Transformer

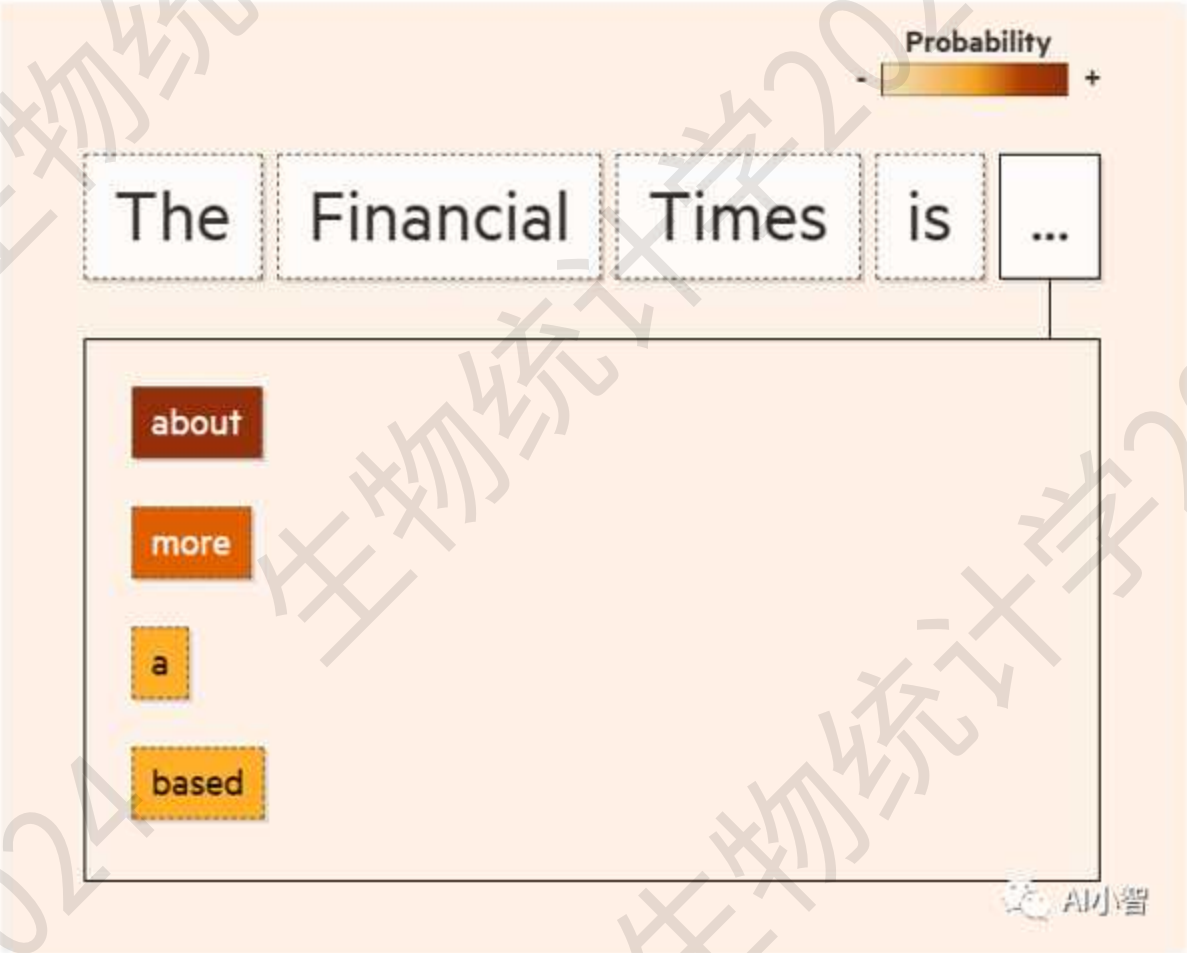
- 将模型的目标设定为预测一个序列中的下一个词,并重复此过程直到输出完成



生物统计学2024

Transformer

- 模型给每个令牌一个概率分数(probability score),表示它是序列中下一个词的可能性



生物统计学2024

Transformer

- 模型给每个令牌一个概率分数(probability score),表示它是序列中下一个词的可能性

生物统计学2024

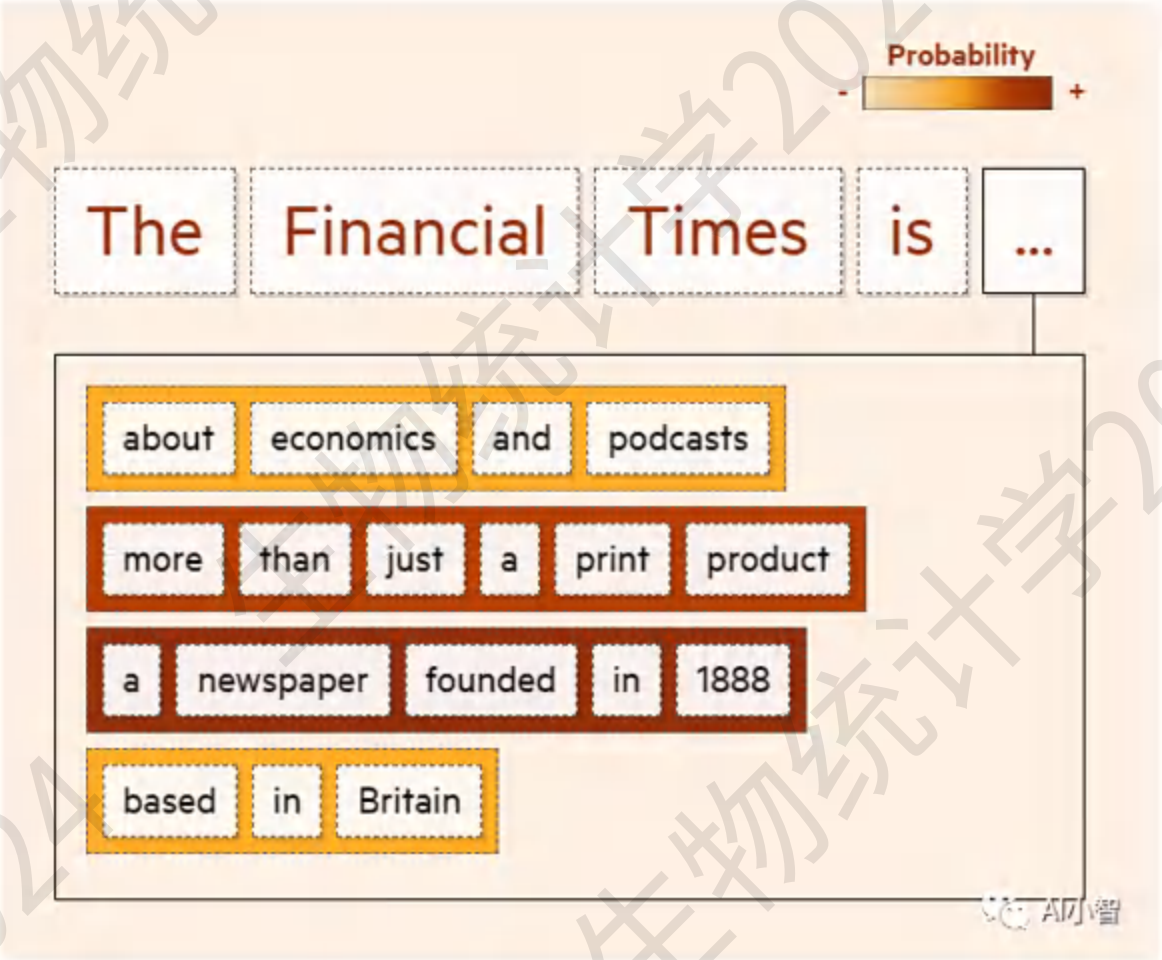
Transformer

- 模型给每个令牌一个概率分数(probability score),表示它是序列中下一个词的可能性（贪心算法）

生物统计学2024

Transformer

- 虽然每个令牌可能是下一个最佳选择,但整个短语可能不太相关
- 束搜索——考虑一组较大令牌集合的概率



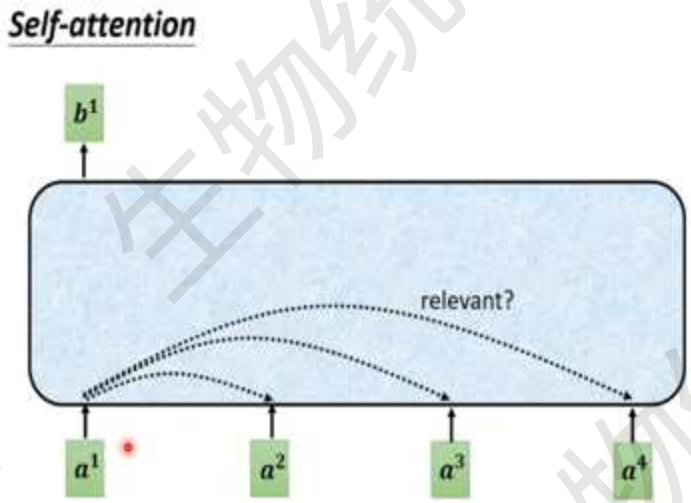
生物统计学2024

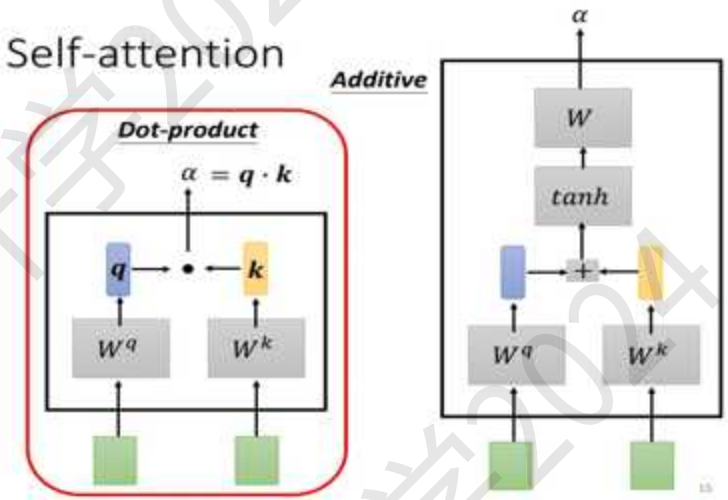
Transformer

- 虽然每个令牌可能是下一个最佳选择,但整个短语可能不太相关
- 束搜索——考虑一组较大令牌集合的概率

生物统计学2024

Self-Attention

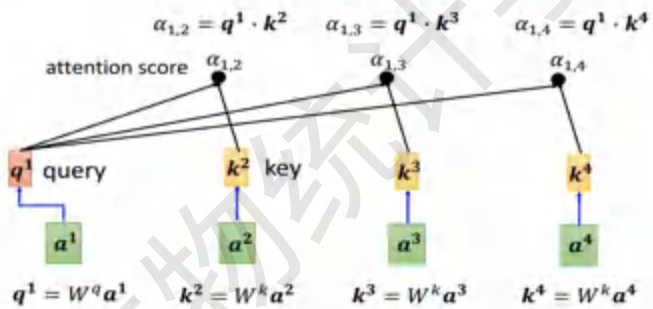




生物统计学2024

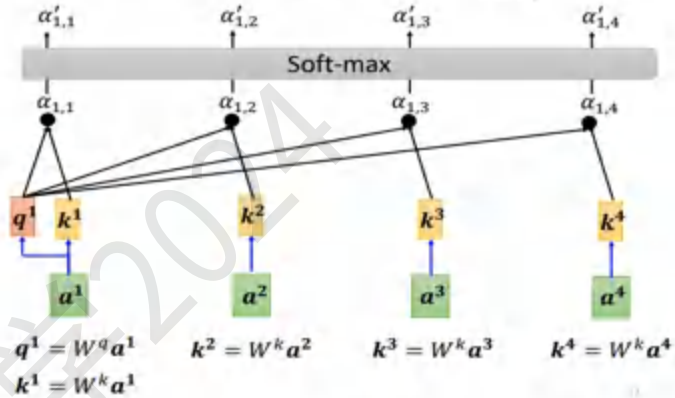
Self-Attention

***Step* 1**：q、k、v的矩阵形式生成



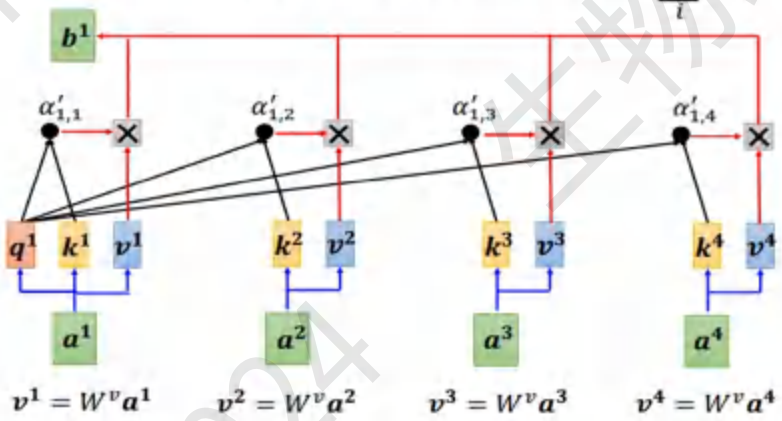
Self-attention

$$\alpha'_{1,i} = exp(\alpha_{1,i})/\sum_j exp(\alpha_{1,j})$$

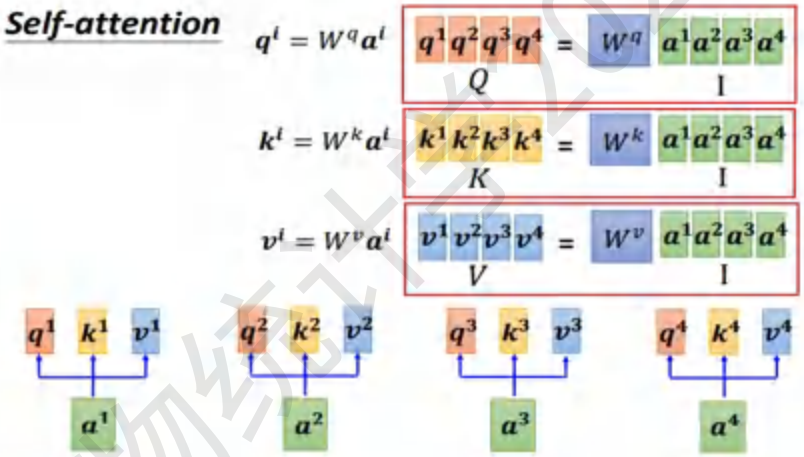


Self-attention Extract information based on attention scores

$$b^1 = \sum_i \alpha'_{1,i} v^i$$



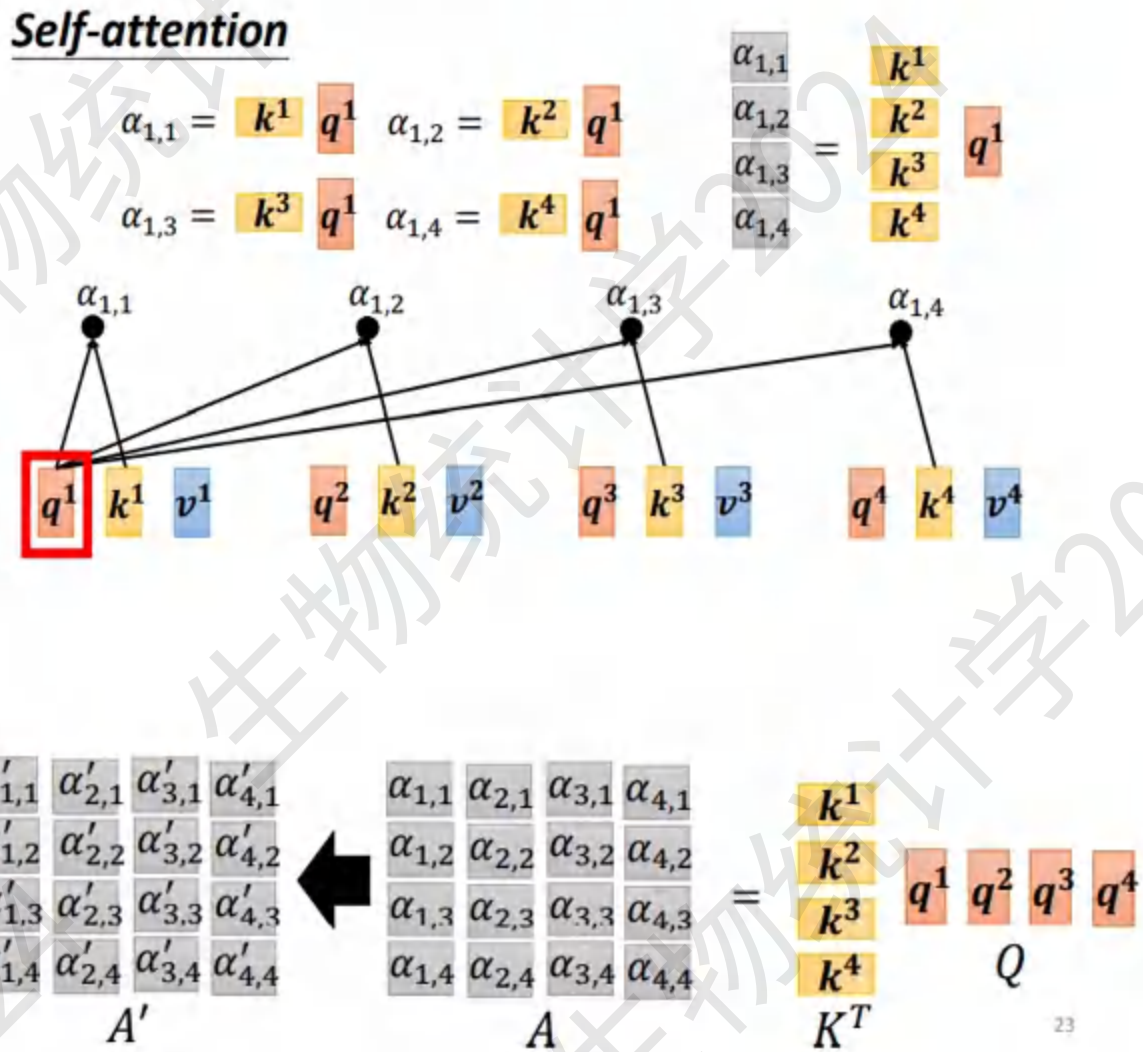
Self-attention



生物统计学2024

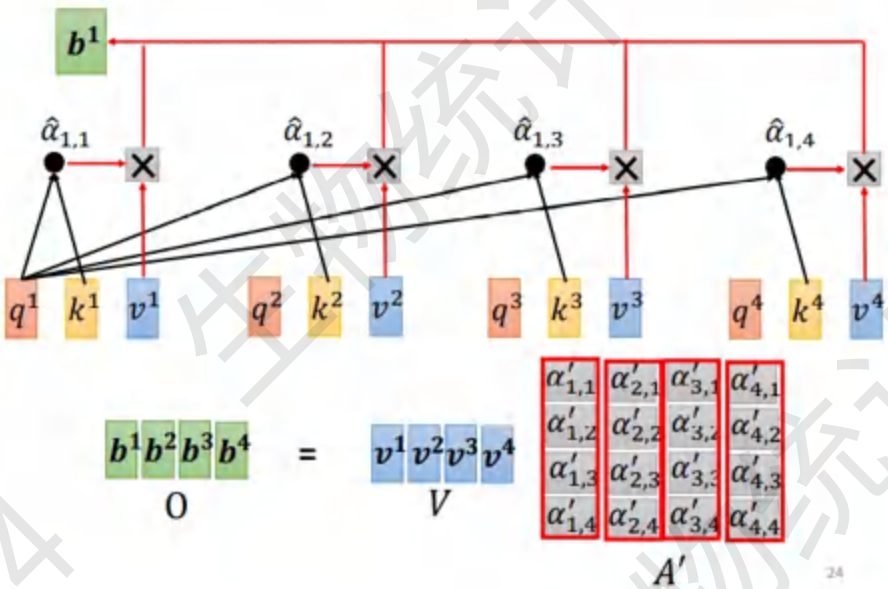
Self-Attention

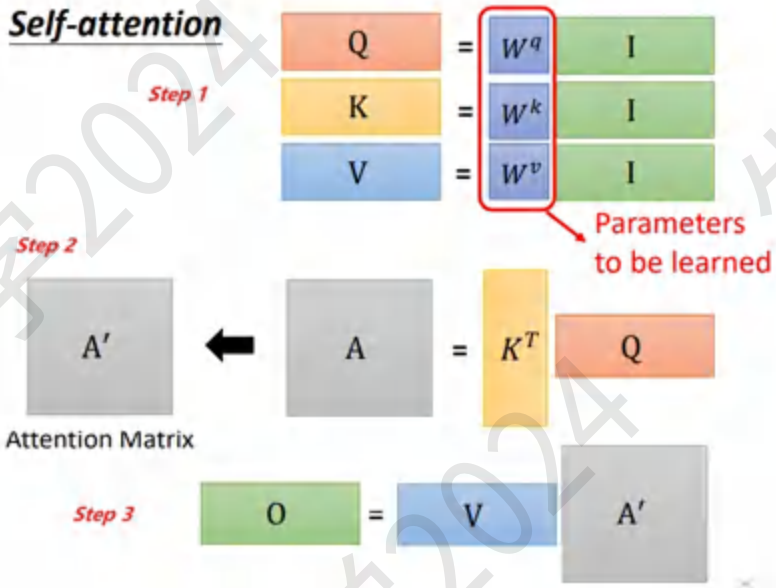
Step 2：利用得到的Q和K计算每两个输入向量之间的相关性，也就是计算attention的值α



Self-Attention

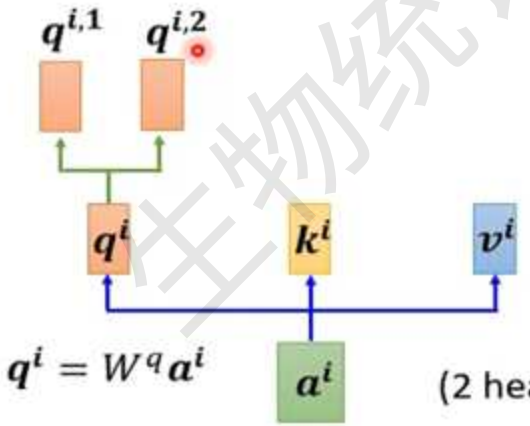
***Step* 3**：利用得到的A'和V，计算每个输入向量a对应的self-attention层的输出向量b



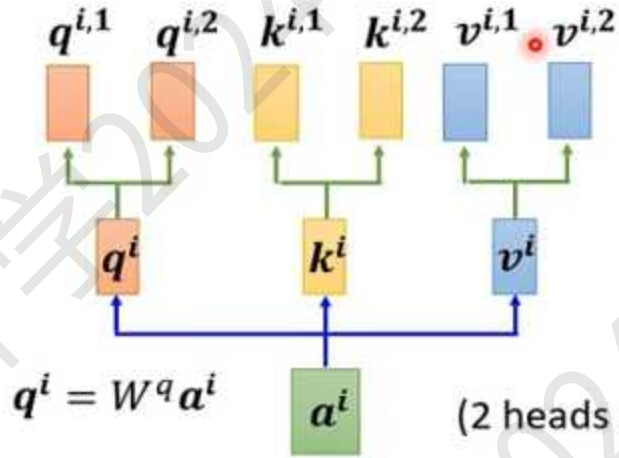


Self-Attention

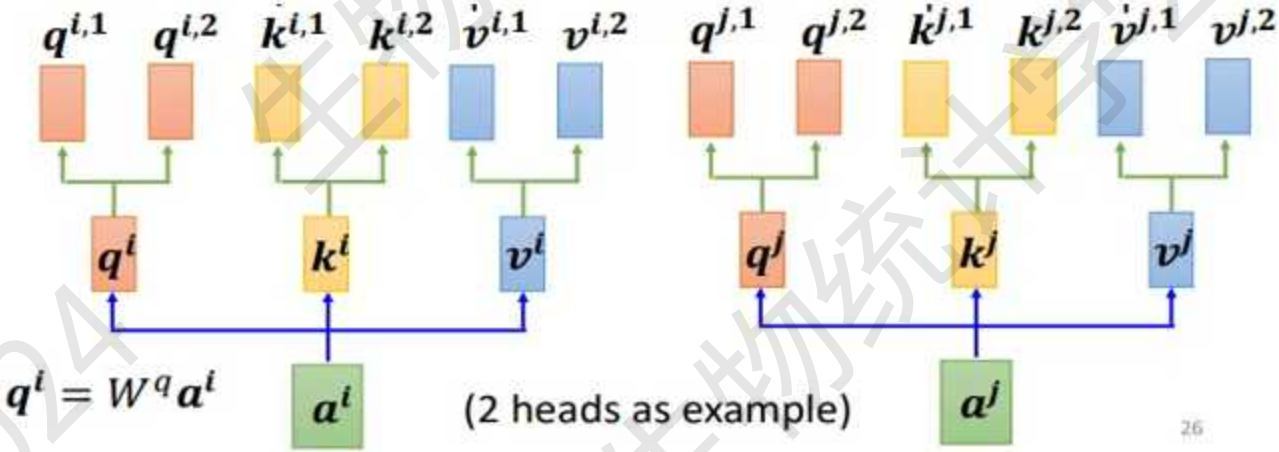
Step 4: Multi-head Self-attention



(2 heads as example)



(2 heads as example)

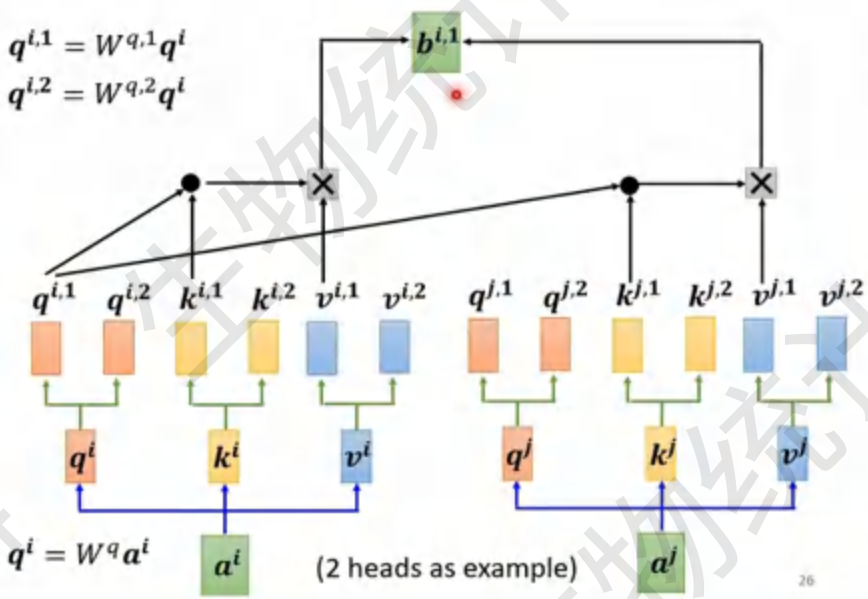


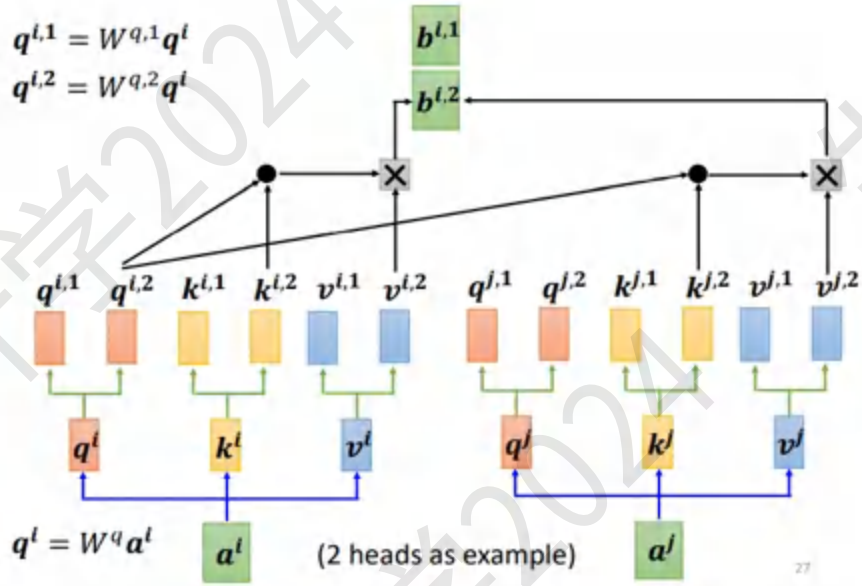
(2 heads as example)

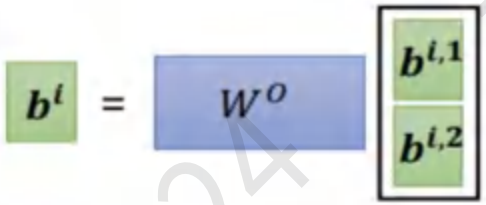
生物统计学2024

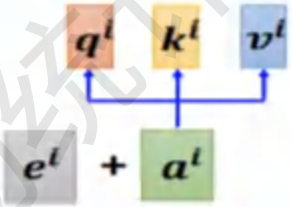
Self-Attention

Step 4: Multi-head Self-attention









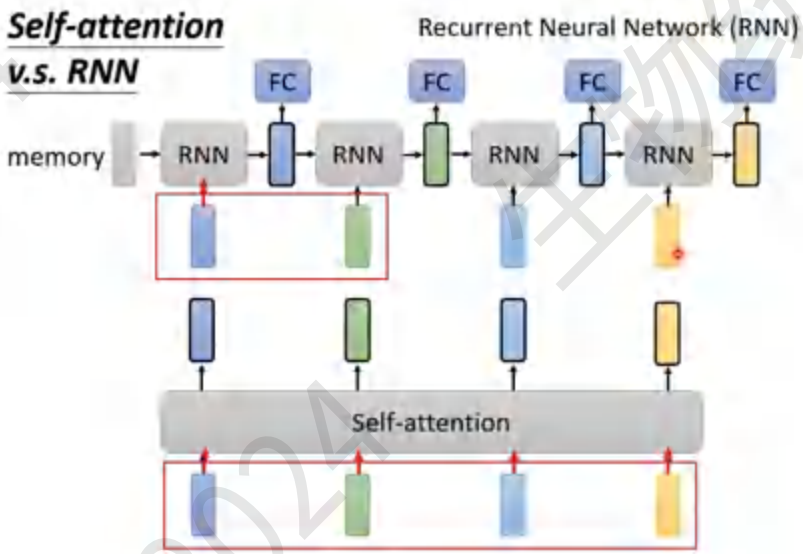
Position Encoding

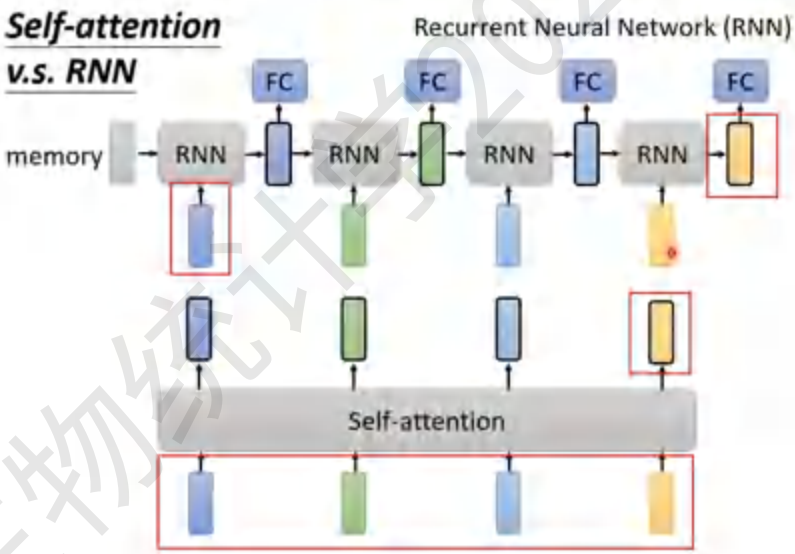
生物统计学2024

Self-Attention

Self-Attention和RNN的区别

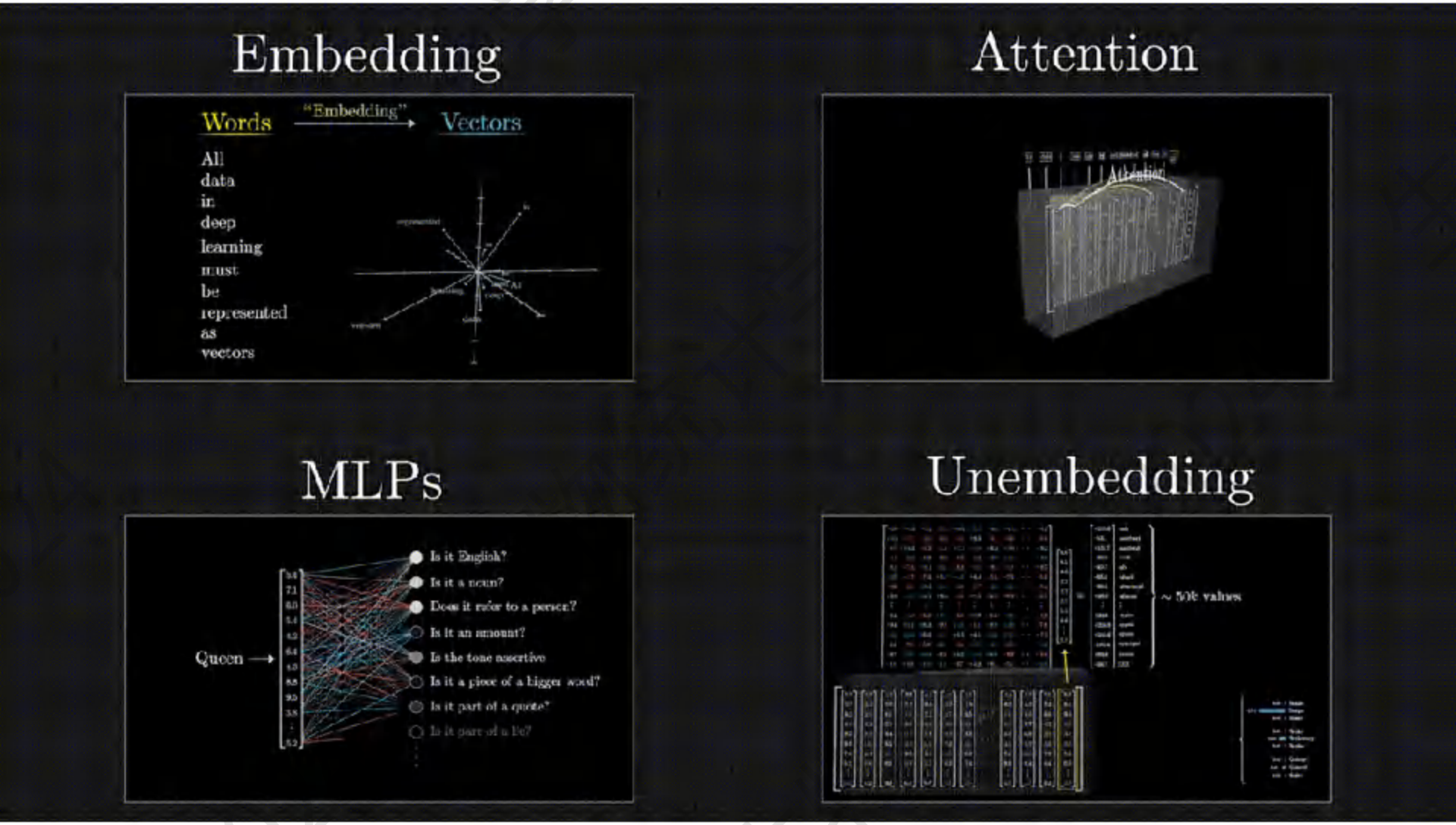
- Self-attention可以考虑全部的输入，而RNN似乎只能考虑之前的输入（左边）。但是当使用双向RNN的时候可以避免这一问题。
- Self-attention可以容易地考虑比较久之前的输入，而RNN的最早输入由于经过了很多层网络的处理变得较难考虑。
- Self-attention可以并行计算，而RNN不同层之间具有先后顺序。





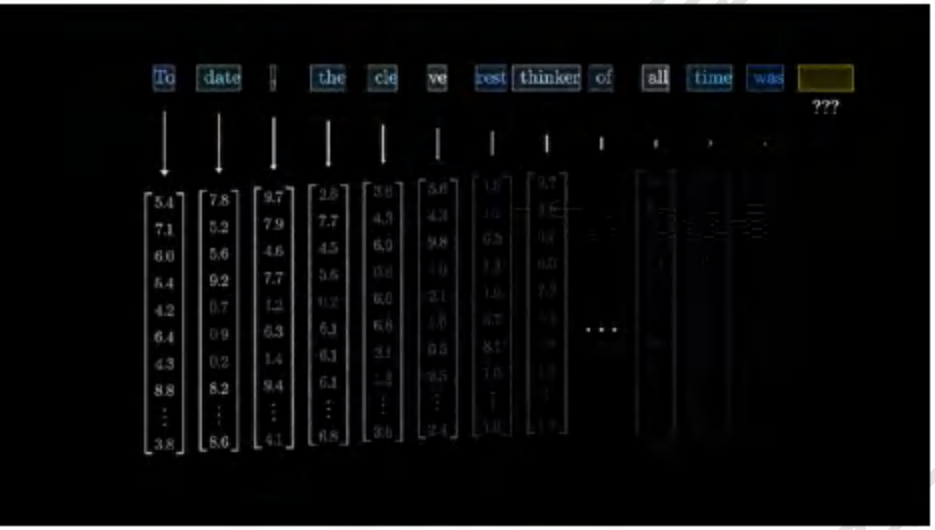
生物统计学2024

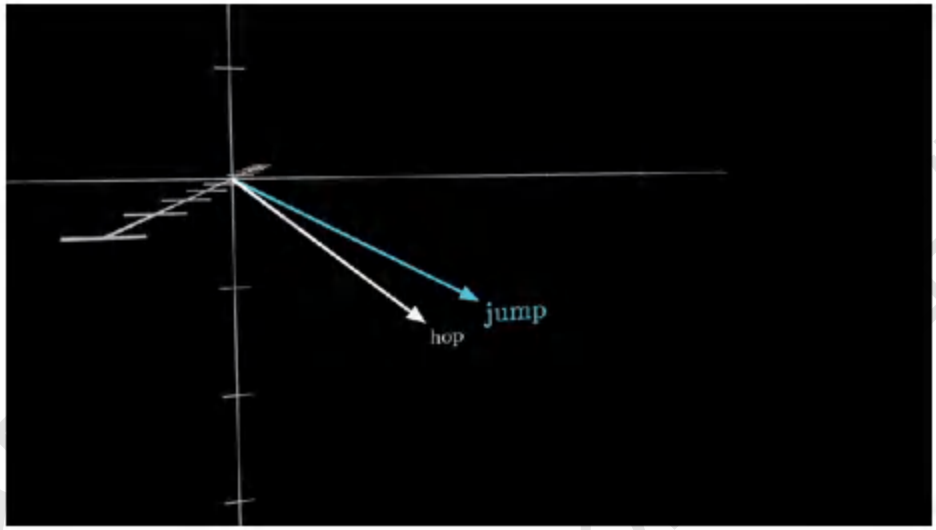
Transformer



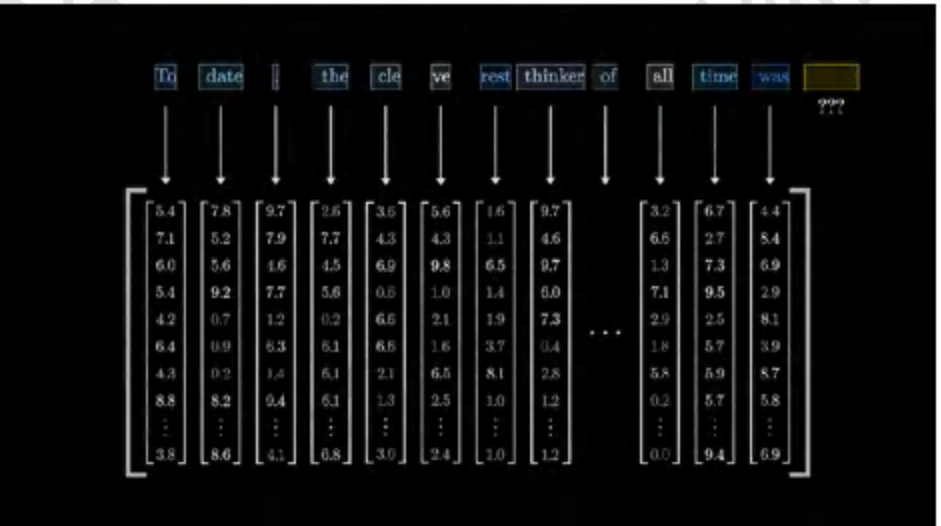
Transformer

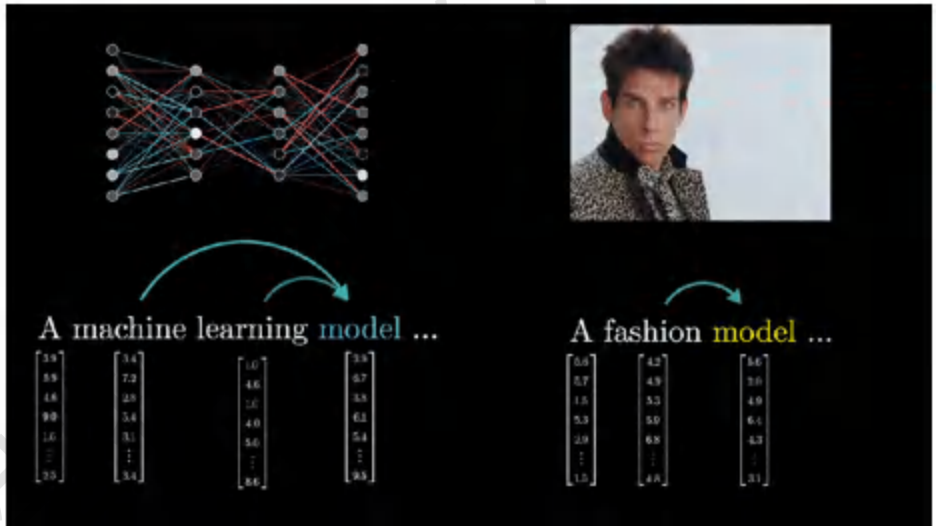
- Enbedding





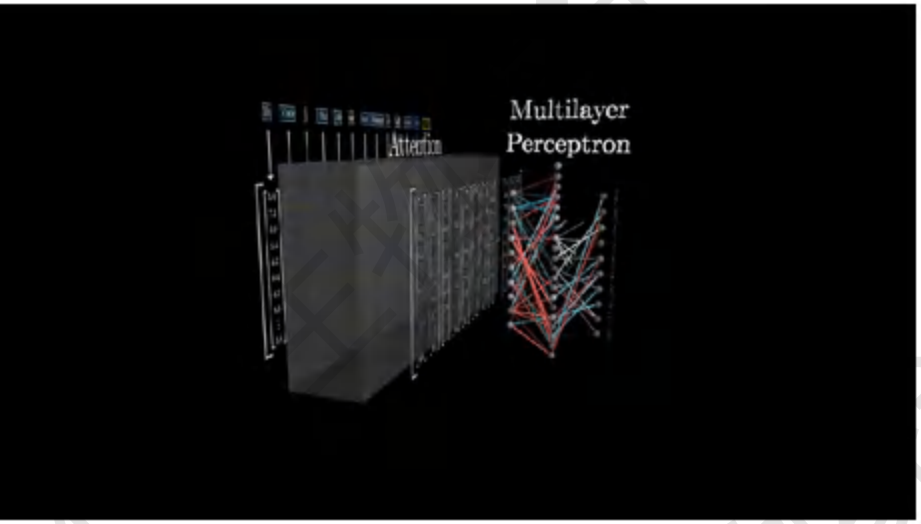
- Attention-语义交流

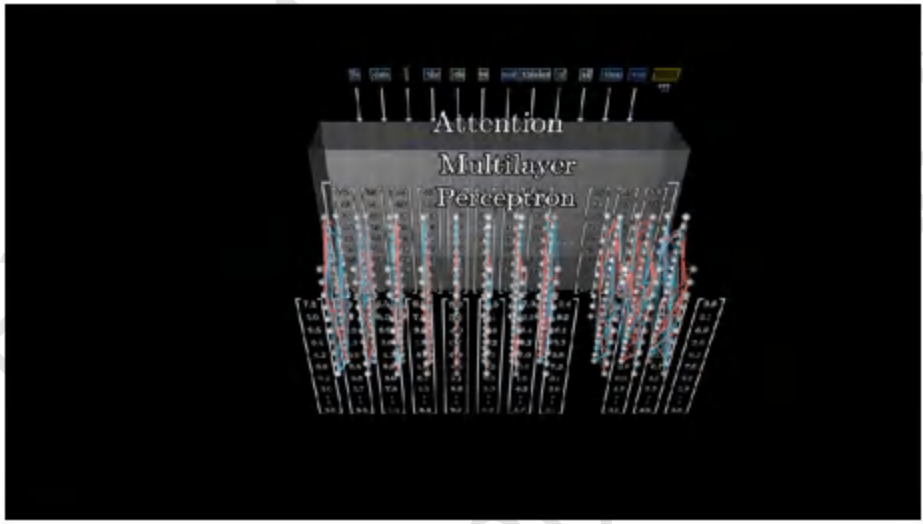


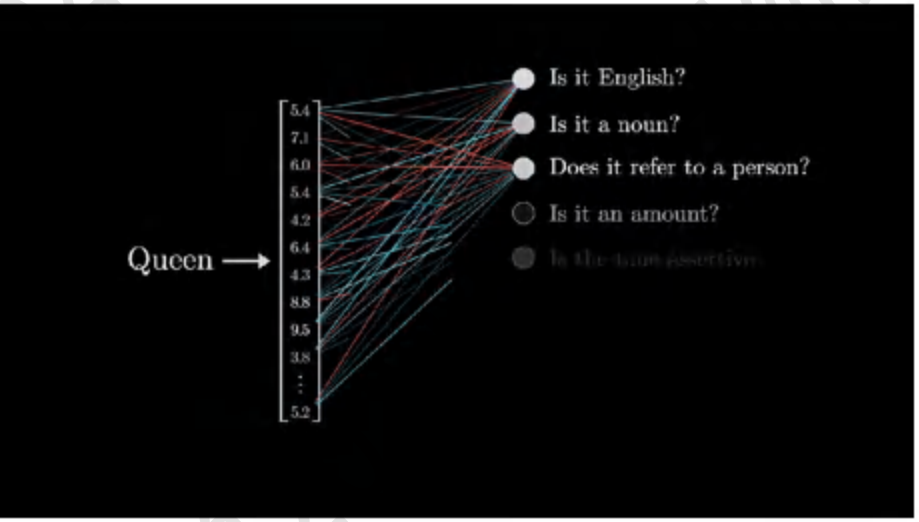


Transformer

- MLPs-感知反馈

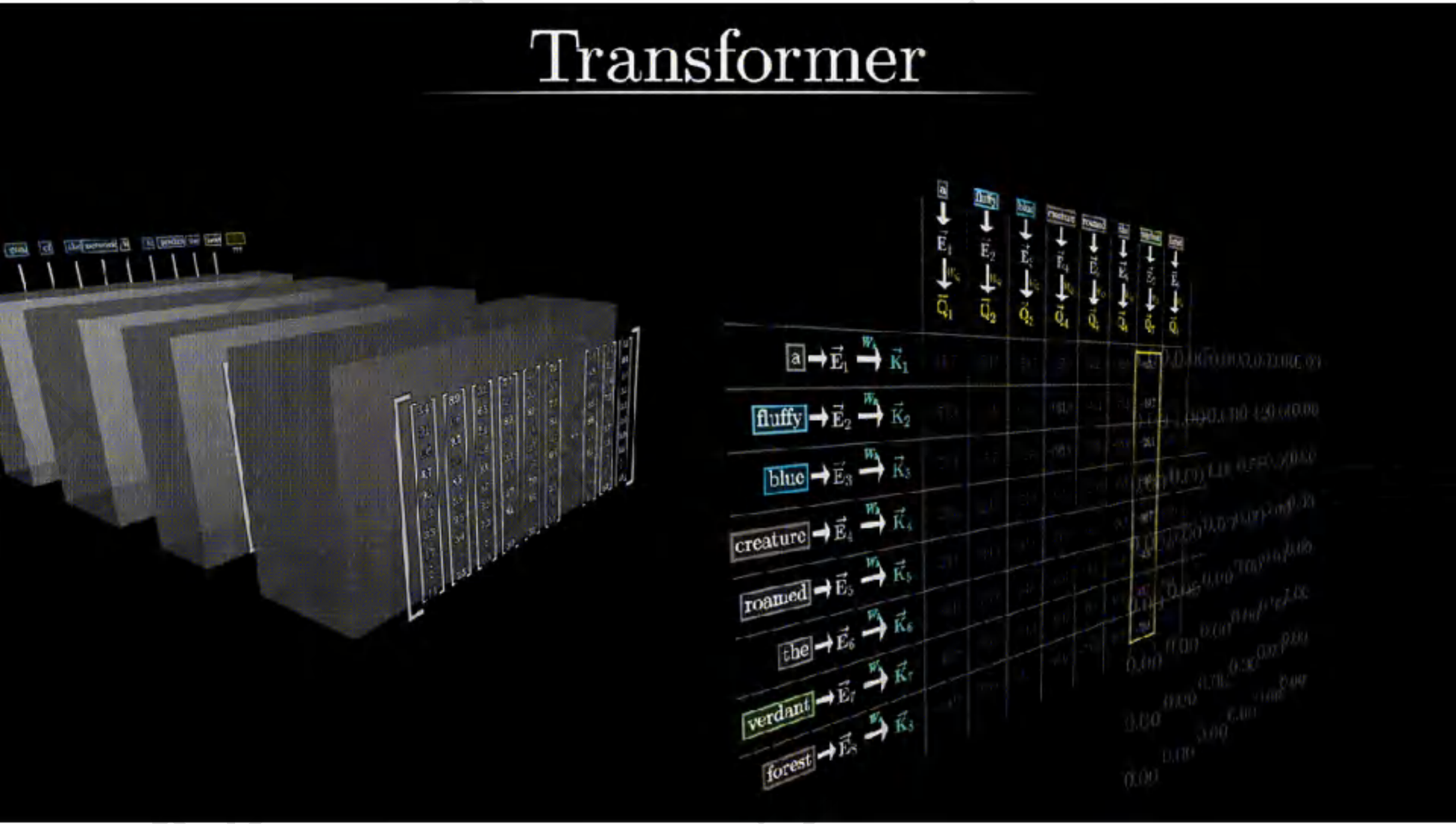






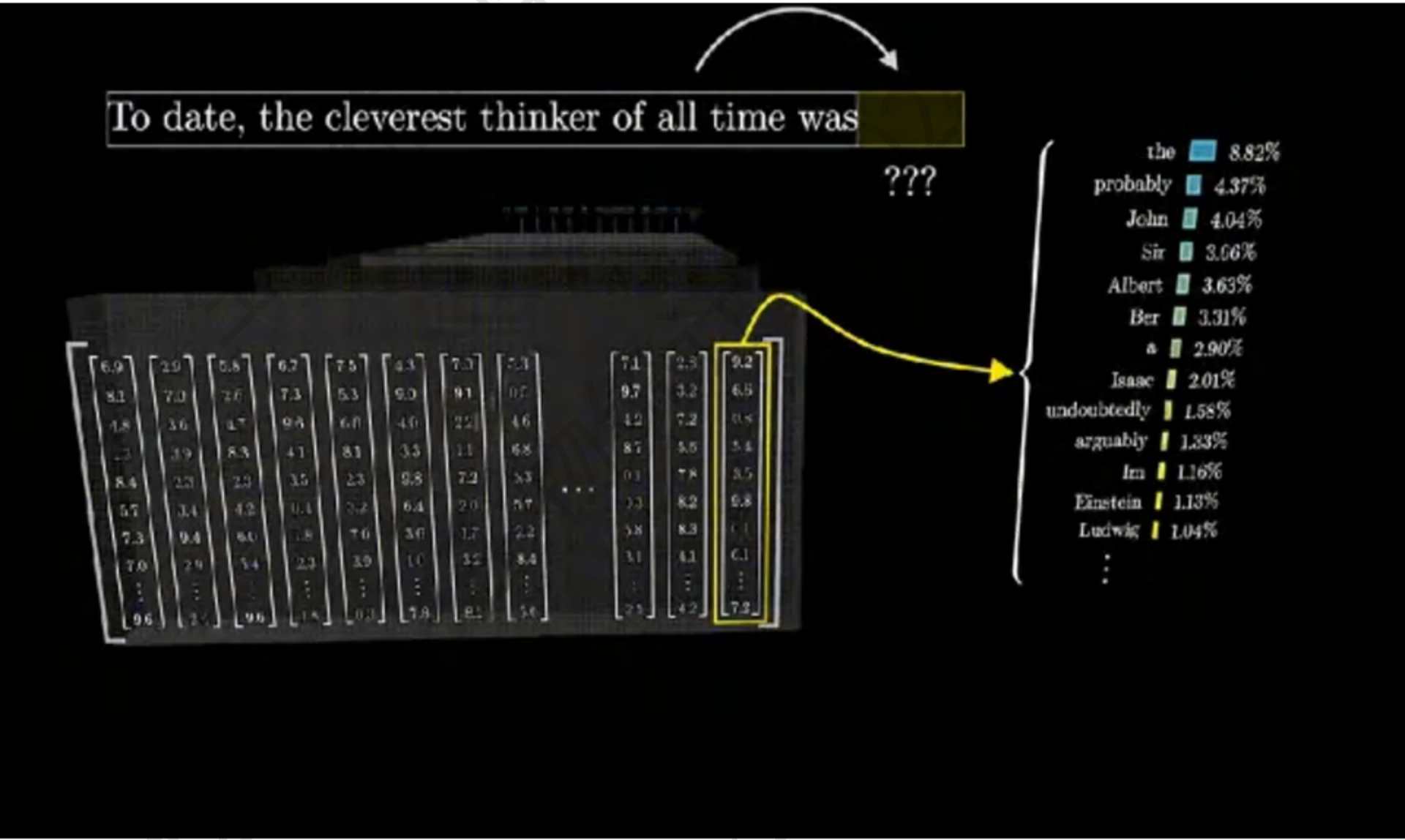
Transformer

- Unembedding-概率



上学2024 生物

Transformer



Transformer

- Transformer已经引领了各种尖端的AI应用程序的创建
  - ChatGPT这样的聊天机器人
  - 从图片中的像素,使用Dall-E、Midjourney和Stable Diffusion等工具,到计算机代码使用GitHub Copilot等生成器
  - 甚至可以预测蛋白质来帮助设计药物分子
  - 。。。
- 统一的模型：总结、翻译、搜索和检索
  - 通过一个统一的语言模型，实现了从图像，音乐，视频多模态的应用，并且强于以往所有的AI应用

生物统计学2024

常用的定理

- 没有免费午餐定理（No Free Lunch Theorem，NFL）
  - 对于基于迭代的最优化算法，不存在某种算法对所有问题（有限的搜索空间内）都有效。如果一个算法对某些问题有效，那么它一定在另外一些问题上比纯随机搜索算法更差。

生物统计学2024

常用的定理

- 丑小鸭定理(Ugly Duckling Theorem)
  - 丑小鸭与白天鹅之间的区别和两只白天鹅之间的区别一样大.

生物统计学2024

常用的定理

- 奥卡姆剃刀原理(Occam's Razor)
  - 如无必要，勿增实体

生物统计学2024

深度学习的特点

深度学习常用算法介绍

深度学习常用框架介绍

生物统计学2024

开源框架概述

- 深度学习研究的热潮持续高涨，各种开源深度学习框架也层出不穷，其中包括TensorFlow、Caffe、Keras、CNTK、Torch7、MXNet、Leaf、Theano、DeepLearning4、Lasagne、Neon等等。下图是各个开源框架在GitHub上的数据统计（2017年初）。

框　架	机　构	支持语言	Stars	Forks	Contributors
TensorFlow	Google	Python/C++/Go/...	41628	19339	568
Caffe	BVLC	C++/Python	14956	9282	221
Keras	fchollet	Python	10727	3575	322
CNTK	Microsoft	C++	9063	2144	100
MXNet	DMLC	Python/C++/R/...	7393	2745	241
Torch7	Facebook	Lua	6111	1784	113
Theano	U. Montreal	Python	5352	1868	271
Deeplearning4J	DeepLearning4J	Java/Scala	5053	1927	101
Leaf	AutumnAI	Rust	4562	216	14
Lasagne	Lasagne	Python	2749	761	55
Neon	NervanaSystems	Python	2633	573	52

开源框架概述

- Google、Microsoft、Facebook等巨头都参与了这场深度学习框架大战，此外，还有毕业于伯克利大学的贾扬清主导开发的Caffe，蒙特利尔大学Lisa Lab团队开发的Theano，以及其他个人或商业组织贡献的框架。下表是主流深度学习框架在各个维度的评分。

	模型设计	接　口	部　署	性　能	架构设计	总体评分
TensorFlow	80	80	90	90	100	88
Caffe	60	60	90	80	70	72

续表

	模型设计	接　口	部　署	性　能	架构设计	总体评分
CNTK	50	50	70	100	60	66
Theano	80	70	40	50	50	58
Torch	90	70	60	70	90	76
MXNet	70	100	80	80	90	84
DeepLearning4J	60	70	80	80	70	72

TensorFlow

- TensorFlow最初是由研究人员和Google Brain团队针对机器学习和深度神经网络进行研究所开发的，目前开源之后可以在几乎各种领域适用。
- TensorFlow灵活的架构可以部署在一个或多个CPU、GPU的台式以及服务器中，或者使用单一的API应用在移动设备中。

生物统计学2024

Torch

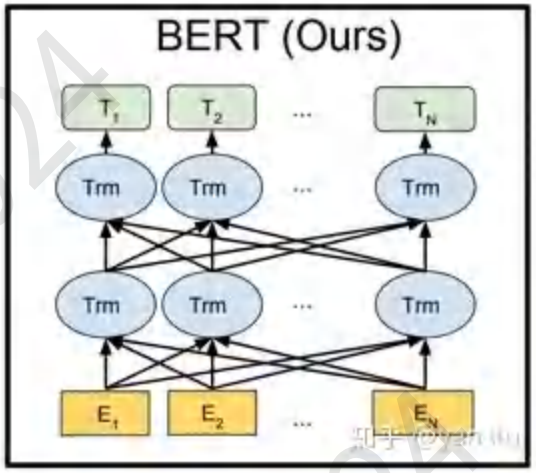
- Torch是一个有大量机器学习算法支持的科学计算框架，其诞生已经有十年之久，但是真正起势得益于Facebook开源了大量Torch的深度学习模块和扩展。Torch另外一个特殊之处是采用了编程语言Lua(该语言曾被用来开发视频游戏)。

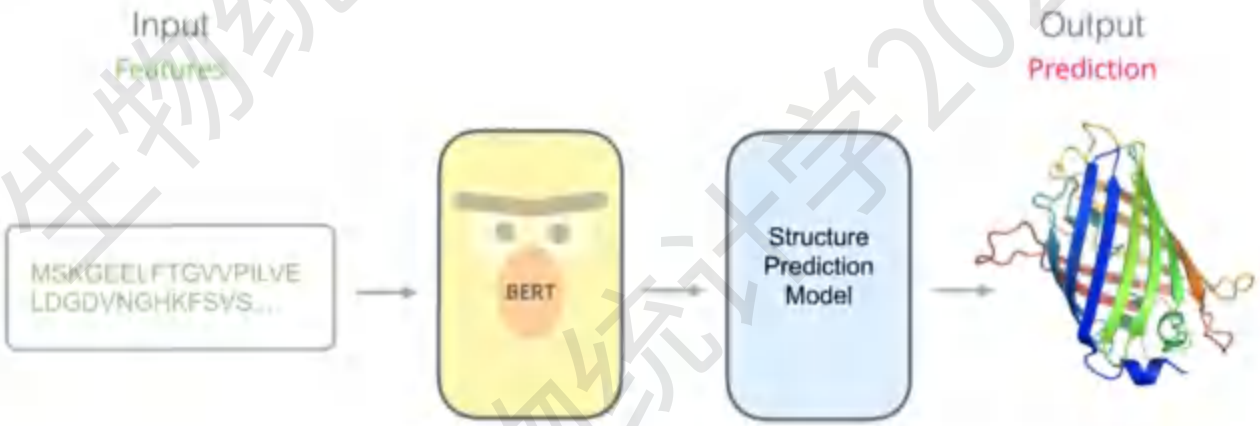
- PyTorch是基于Torch的衍生，支持Python语言，实现了机器学习框架 Torch 在 Python 语言环境的执行。

生物统计学2024

Bert

- 使用Transformer的结构将已经走向瓶颈期的Word2Vec带向了一个新的方向，并再一次炒火了《Attention is All you Need》这篇论文；
- 11个NLP任务的精度大幅提升足以震惊整个深度学习领域；
- 无私的开源了多种语言的源码和模型，具有非常高的商业价值；
- 迁移学习又一次胜利，而且这次是在NLP领域的大胜，狂胜。





Caffe

- Caffe由加州大学伯克利的PHD贾扬清开发，全称Convolutional Architecture for Fast Feature Embedding，是一个清晰而高效的开源深度学习框架，目前由伯克利视觉学中心（Berkeley Vision and Learning Center，BVLC）进行维护。（贾扬清曾就职于MSRA、NEC、Google Brain，他也是TensorFlow的作者之一，目前任职于Facebook FAIR实验室。）
- Caffe2脸书 (Facebook) 出品，为生产环境设计，提供在各种平台（包括移动设备）的运行。

Caffe

Caffe2

生物统计学2024

Theano

- 2008年诞生于蒙特利尔理工学院,Theano派生出了大量深度学习Python软件包，最著名的包括Blocks和Keras。Theano的核心是一个数学表达式的编译器，它知道如何获取你的结构。并使之成为一个使用numpy、高效本地库的高效代码，如BLAS和本地代码（C++）在CPU或GPU上尽可能快地运行。它是为深度学习中处理大型神经网络算法所需的计算而专门设计的，是这类库的首创之一（发展始于2007年），被认为是深度学习研究和开发的行业标准。

theano

生物统计学2024

Deeplearning4j

- Deeplearning4j是“for Java”的深度学习框架，也是首个商用级别的深度学习开源库。Deeplearning4j由创业公司Skymind于2014年6月发布，使用Deeplearning4j的不乏埃森哲、雪弗兰、博斯咨询和IBM等明星企业。DeepLearning4j是一个面向生产环境和商业应用的高成熟度深度学习开源库，可与Hadoop和Spark集成，即插即用，方便开发者在APP中快速集成深度学习功能。

DL4J Deep Learning for Java

生物统计学2024

MXNet

- 出自CXXNet、Minerva、Purine 等项目的开发者之手，主要用C++ 编写。MXNet 强调提高内存使用的效率，甚至能在智能手机上运行诸如图像识别等任务。

mxnet

生物统计学2024

CNTK

- CNTK（Computational Network Toolkit）是微软研究院（MSR）开源的深度学习框架。它最早由start the deep learning craze的演讲人创建，目前已经发展成一个通用的、跨平台的深度学习系统，在语音识别领域的使用尤其广泛。

Microsoft
CNTK

生物统计学2024

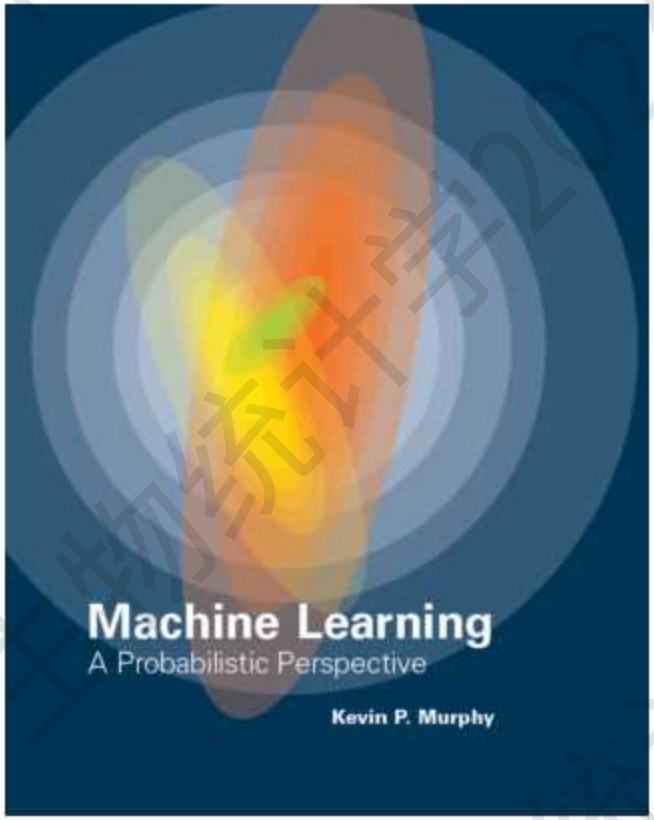
PaddlePaddle

- Visual DL是百度开发的，基于echar和PaddlePaddle，支持PaddlePaddle，PyTorch和MXNet等主流框架。

飞桨 VisualDL

生物统计学2024

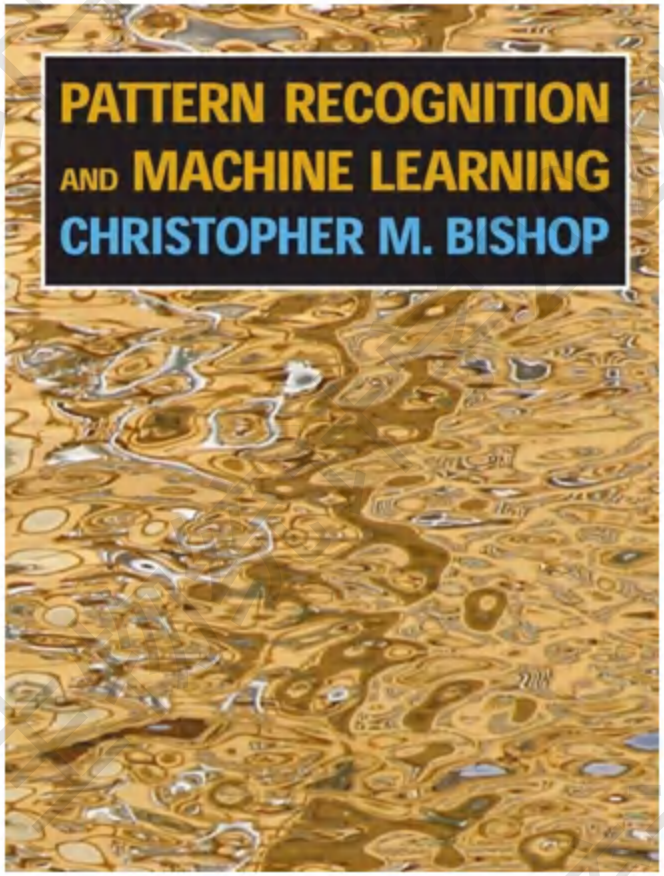
References



https://github.com/probml/pyprobml

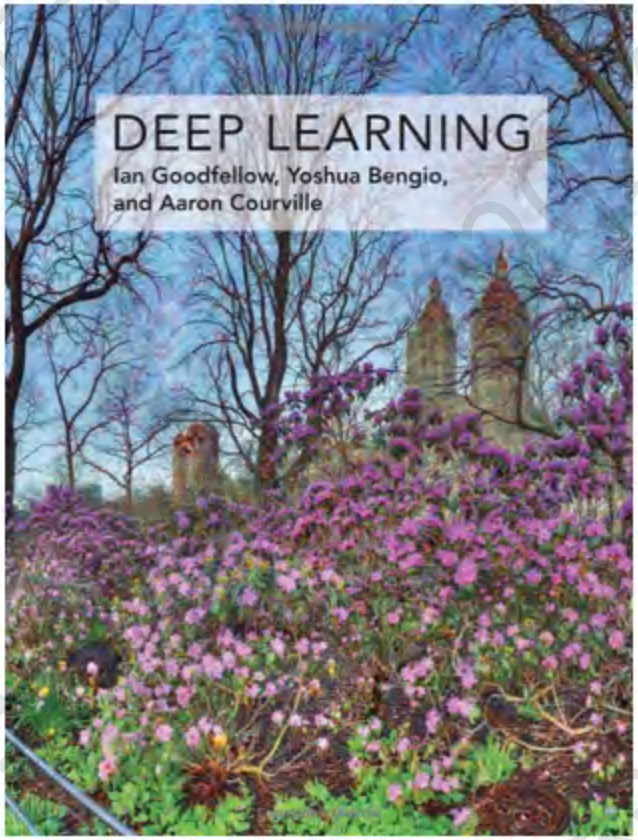
生物统计学2024

References



生物统计学2024

References



https://github.com/exacity/deeplearningbook-chinese

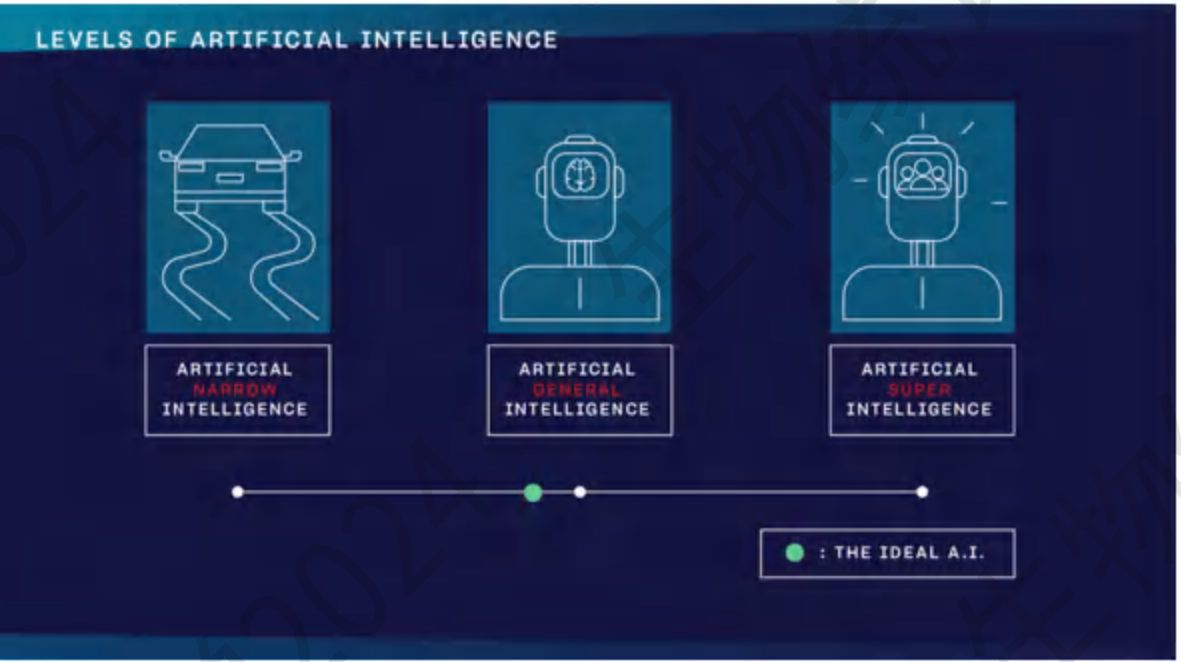
生物统计学2024

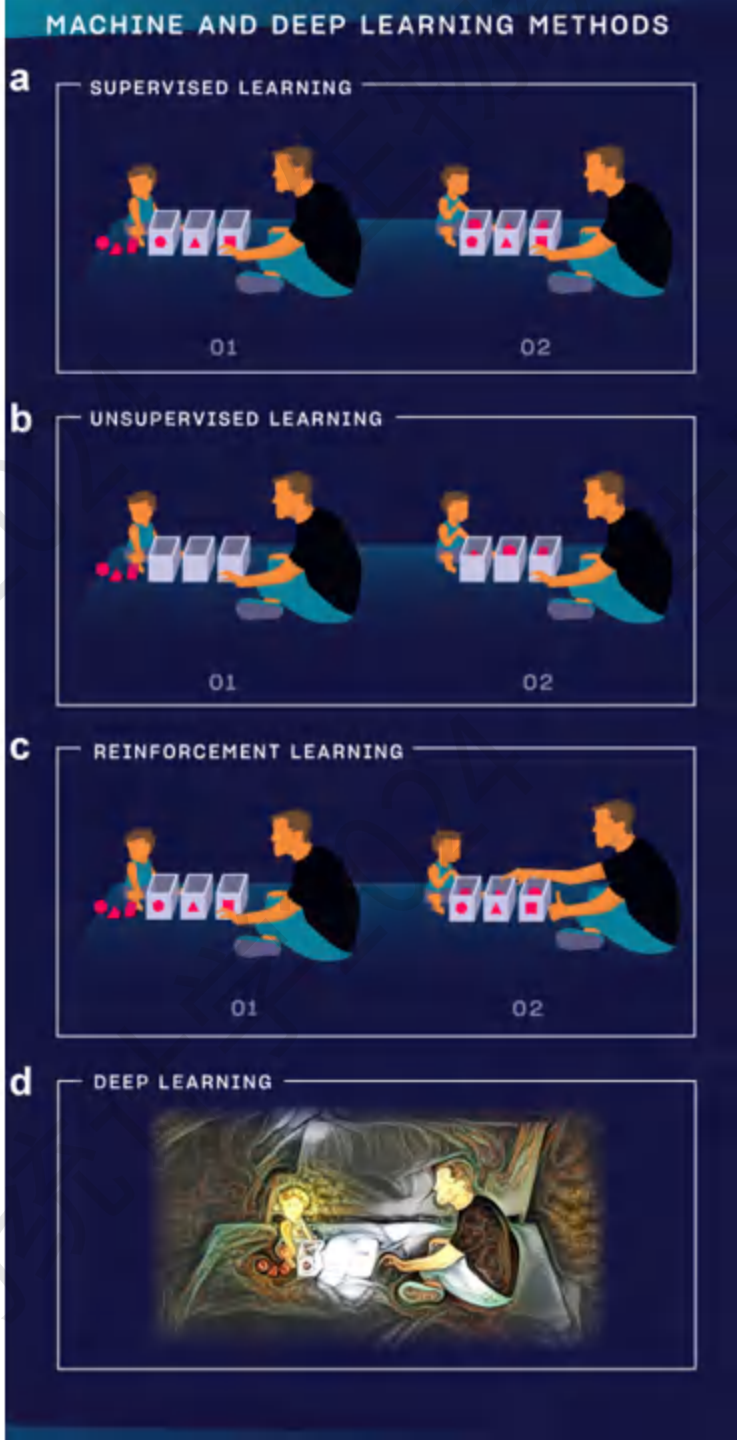
Part III

生物大数据的深度学习方法

生物统计学2024

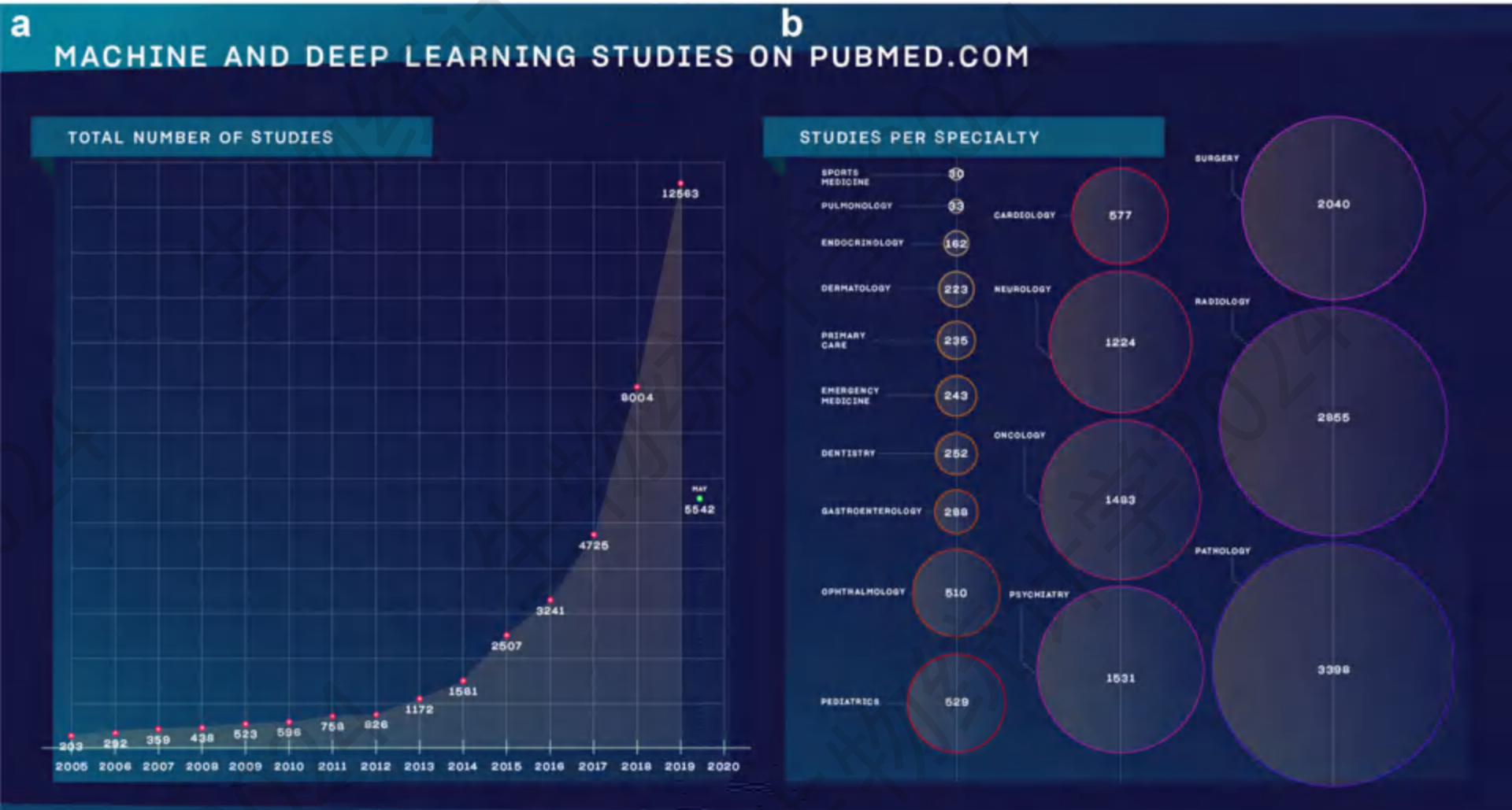
AI in a nutshell





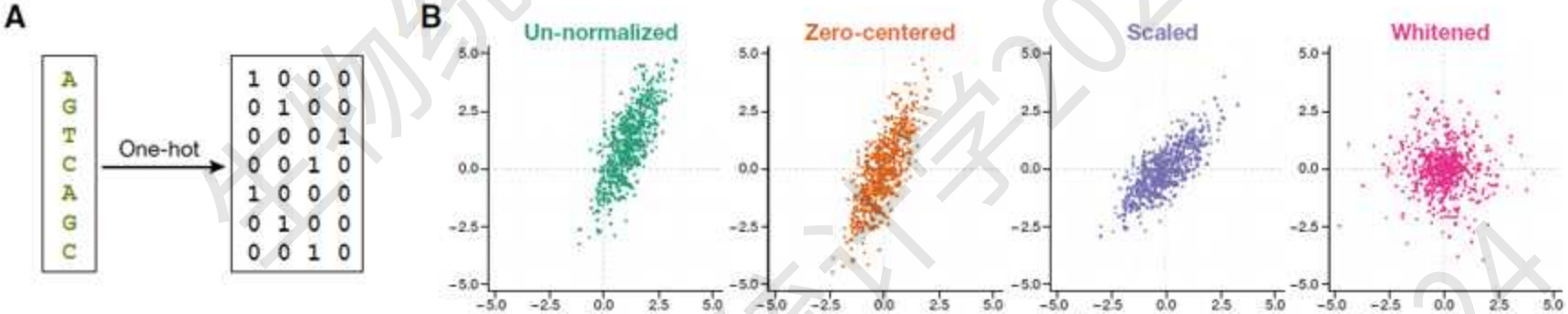
生物统计学2024

Machine Learning in biology



Feature Selection

Collection→Partition→Normalization
Data label

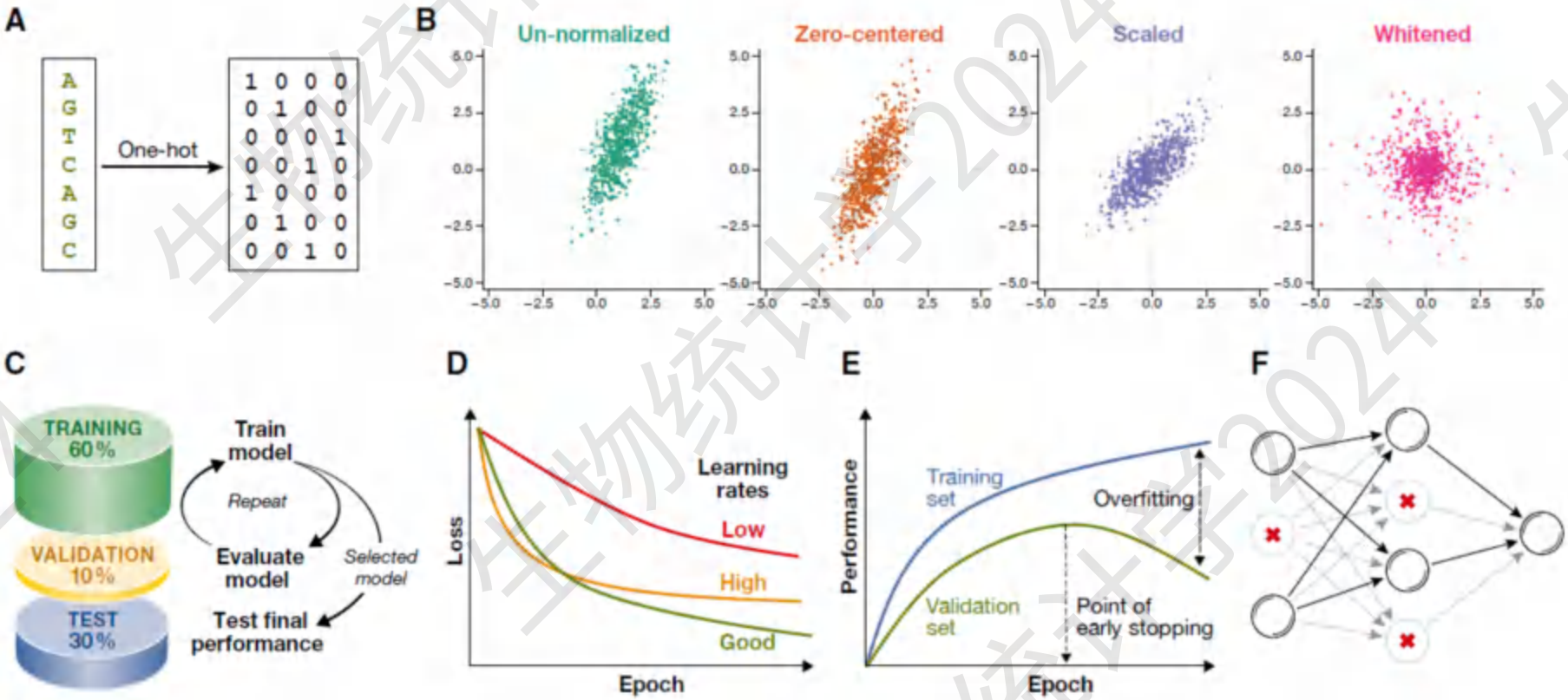


(A) Data One-hot Coding

(B) Different data normalization methods

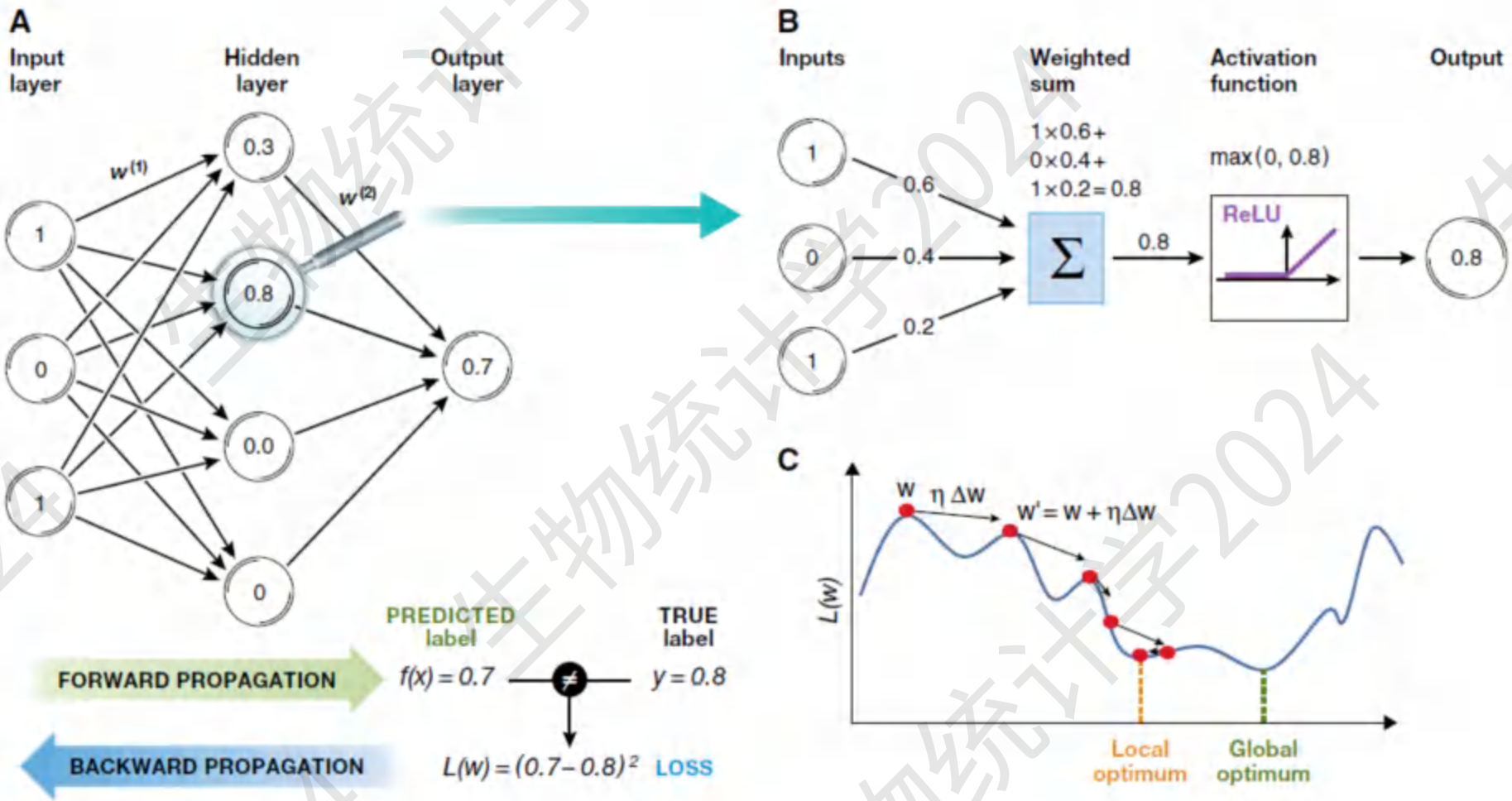
(C) Training data, validation data(adjust model structure),Test data

Data normalization



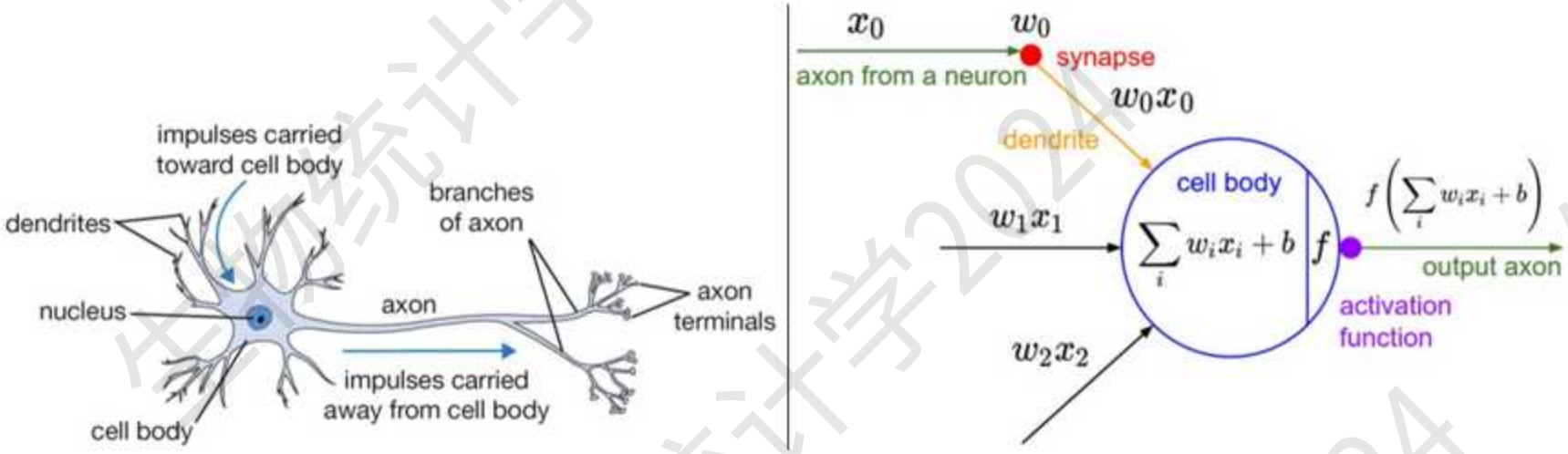
Adapted from "Deep learning for computational biology. Molecular Systems Biology, 2016."

Artificial Neural Network



Adapted from "Deep learning for computational biology. Molecular Systems Biology, 2016."

Model Construction



A Cartoon drawing of a biological neuron(left) and its mathematical model(right)

$$f\left(\sum_i w_i x_i + b\right)$$

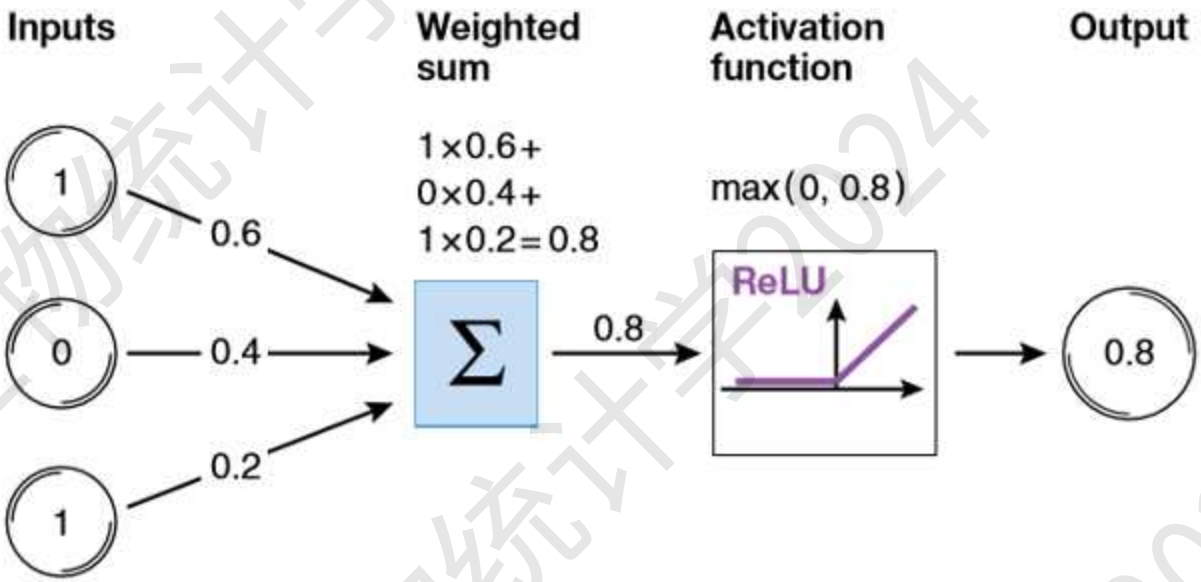
Activation function

Translatio n

Input x_i and corresponded weight w_i

生物统计学2024

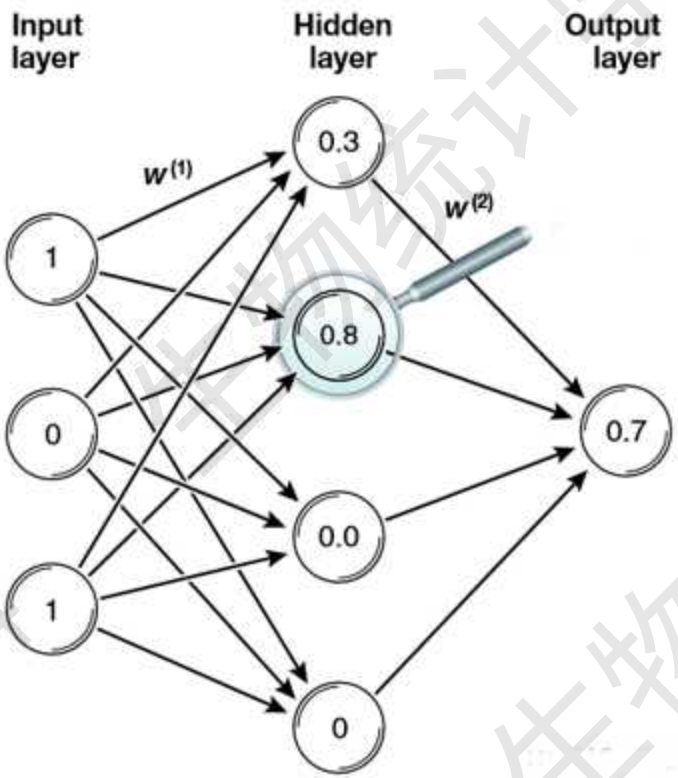
Model Construction



An example for a “transmission of neural signal”

生物统计学2024

Model Construction



Neuron network construction process

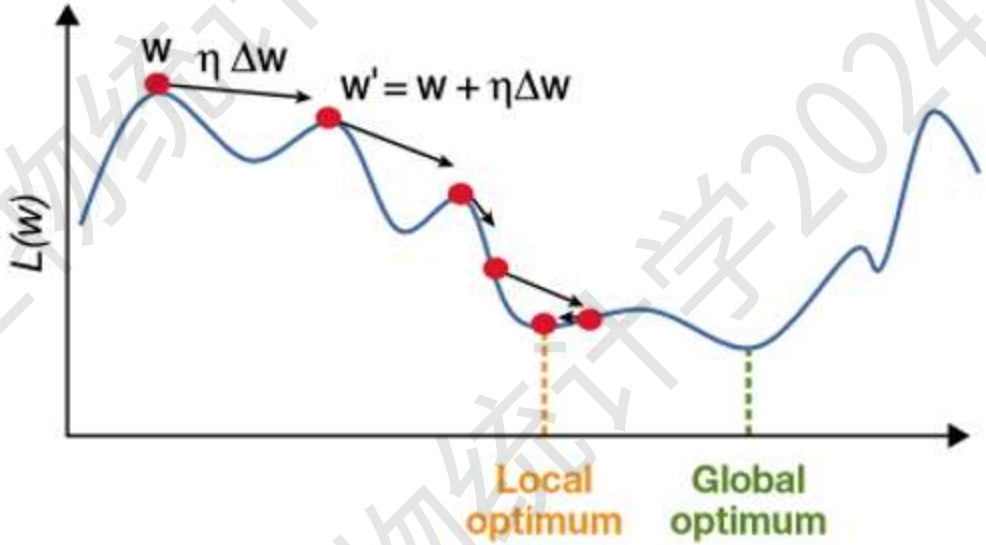
Model construction：
Input label:1,0,1
Predicted label: 0.7

Ture data label:0.8

生物统计学2024

Pick the best model

loss function: $L(w)=(\text{predicted label-true label})^2$



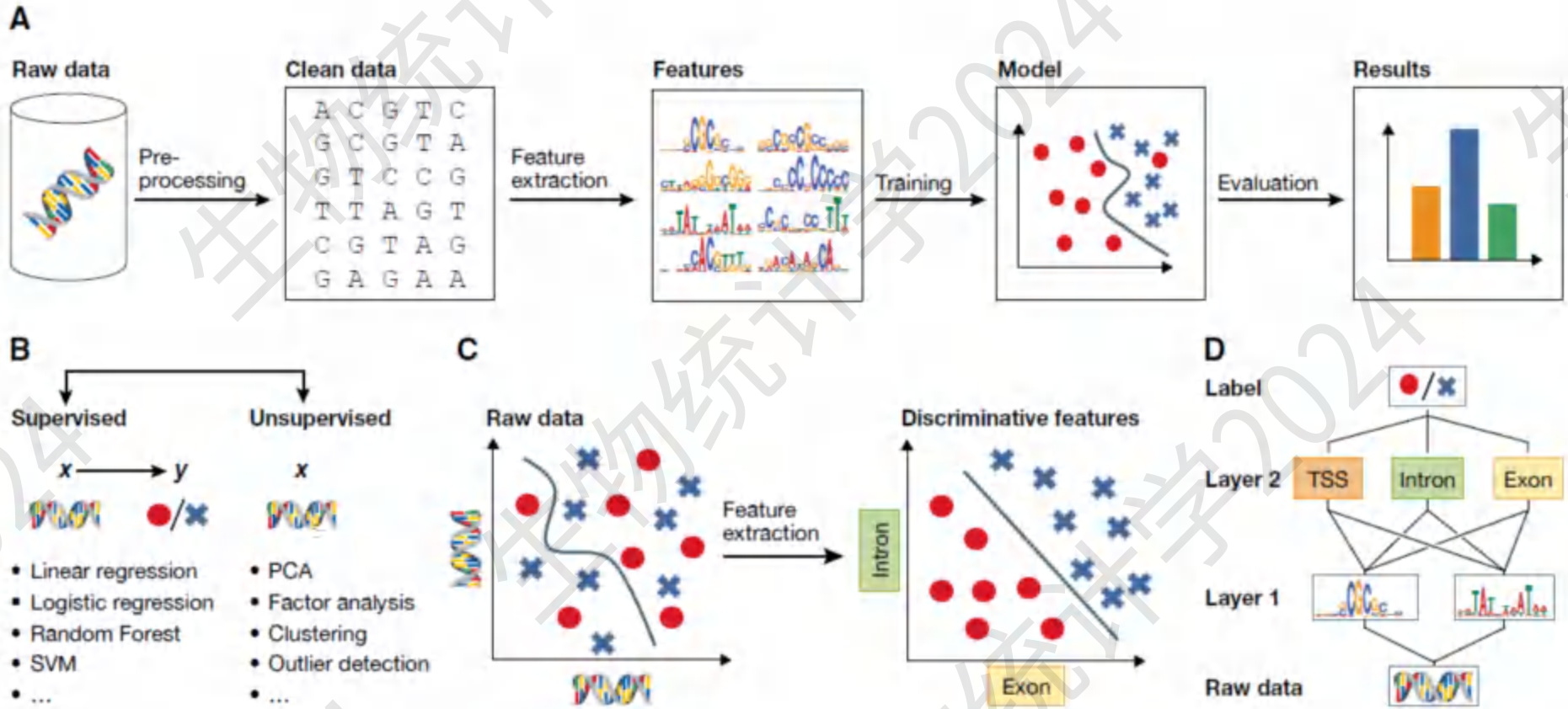
Many parameters are adjusted to find the global minimums then update the neuron weight

↓

Predicted label:0.8 → True label:0.8

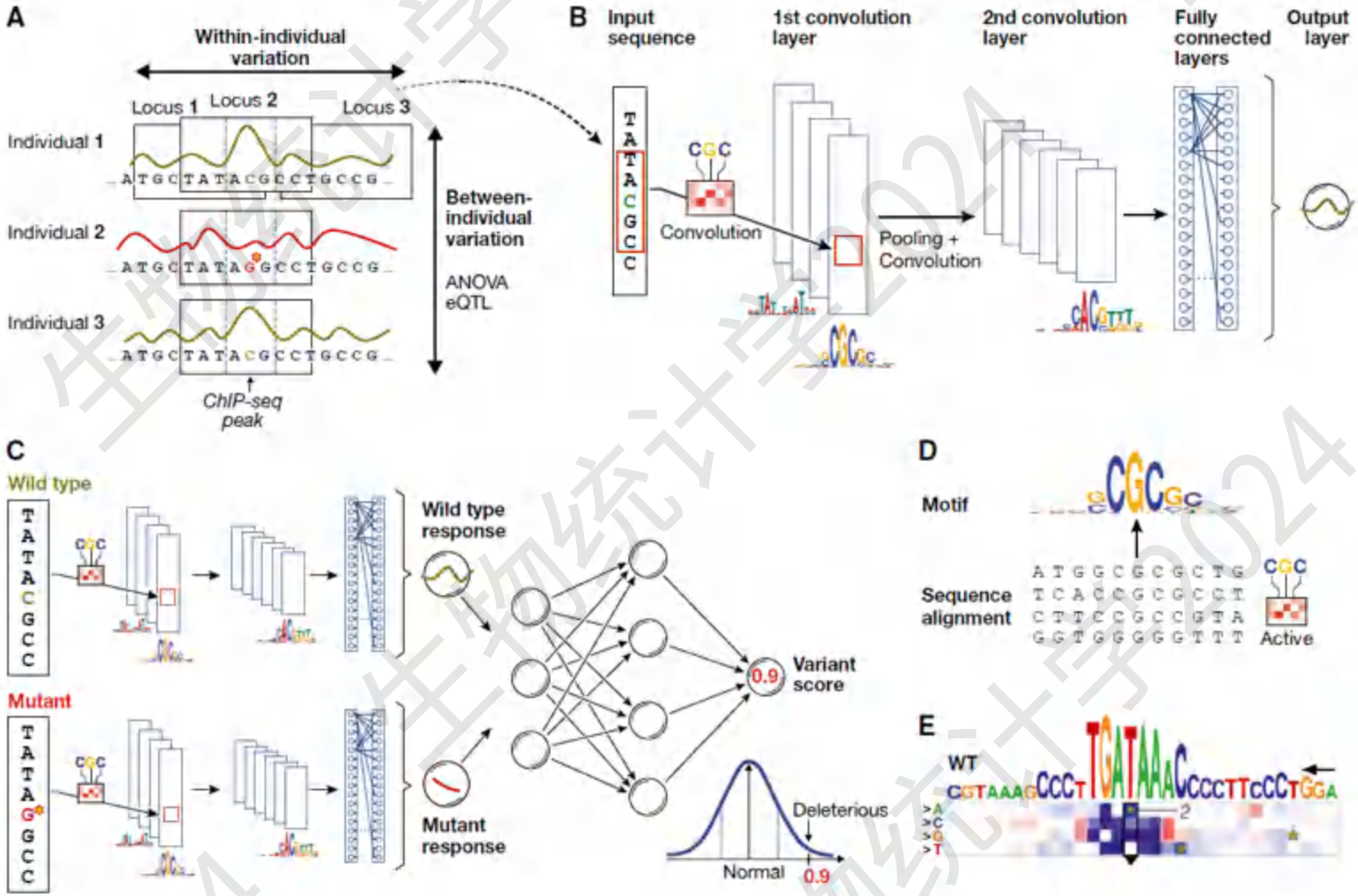
生物统计学2024

Machine learning and representation learning



Adapted from "Deep learning for computational biology. Molecular Systems Biology, 2016."

Neural Network for DNA sequences

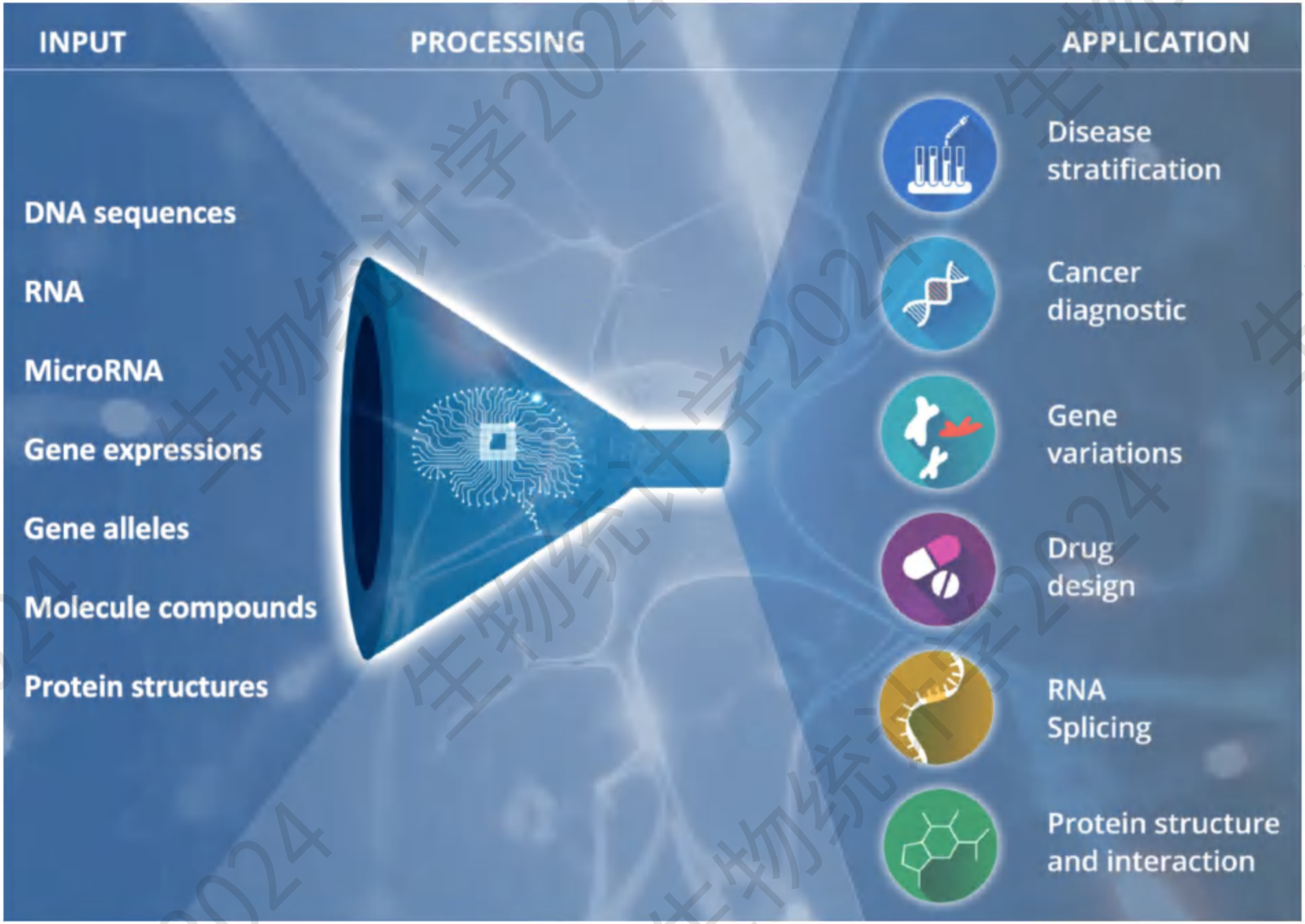


Adapted from "Deep learning for computational biology. Molecular Systems Biology, 2016."

Part IV

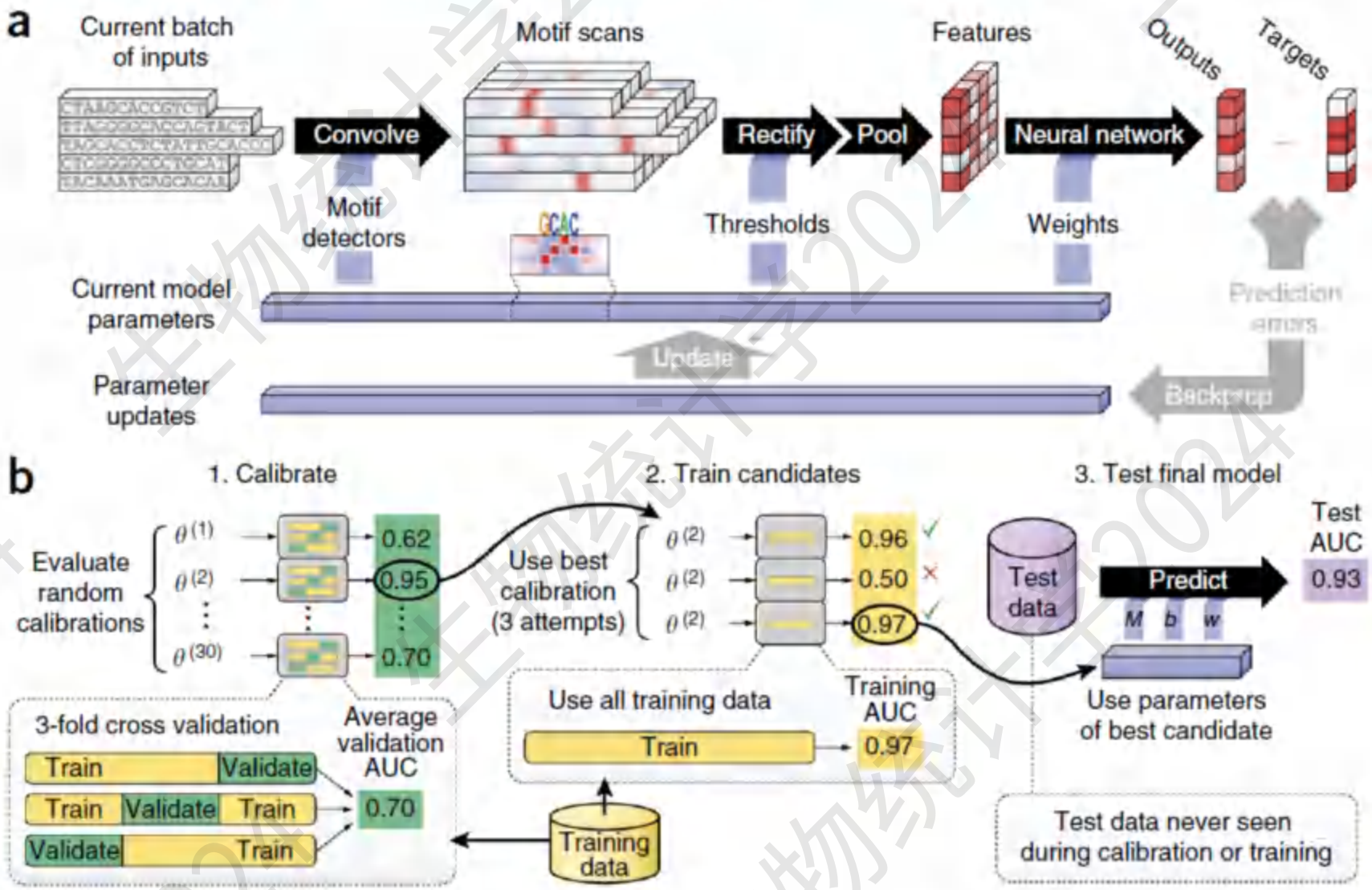
生物大数据深度学习的应用案例

生物统计学2024


INPUT
PROCESSING
APPLICATION
DNA sequences
RNA
MicroRNA
Gene expressions
Gene alleles
Molecule compounds
Protein structures
Disease stratification
Cancer diagnostic
Gene variations
Drug design
RNA Splicing
Protein structure and interaction

生物统计学2024

DNA- and RNA-binding protein prediction



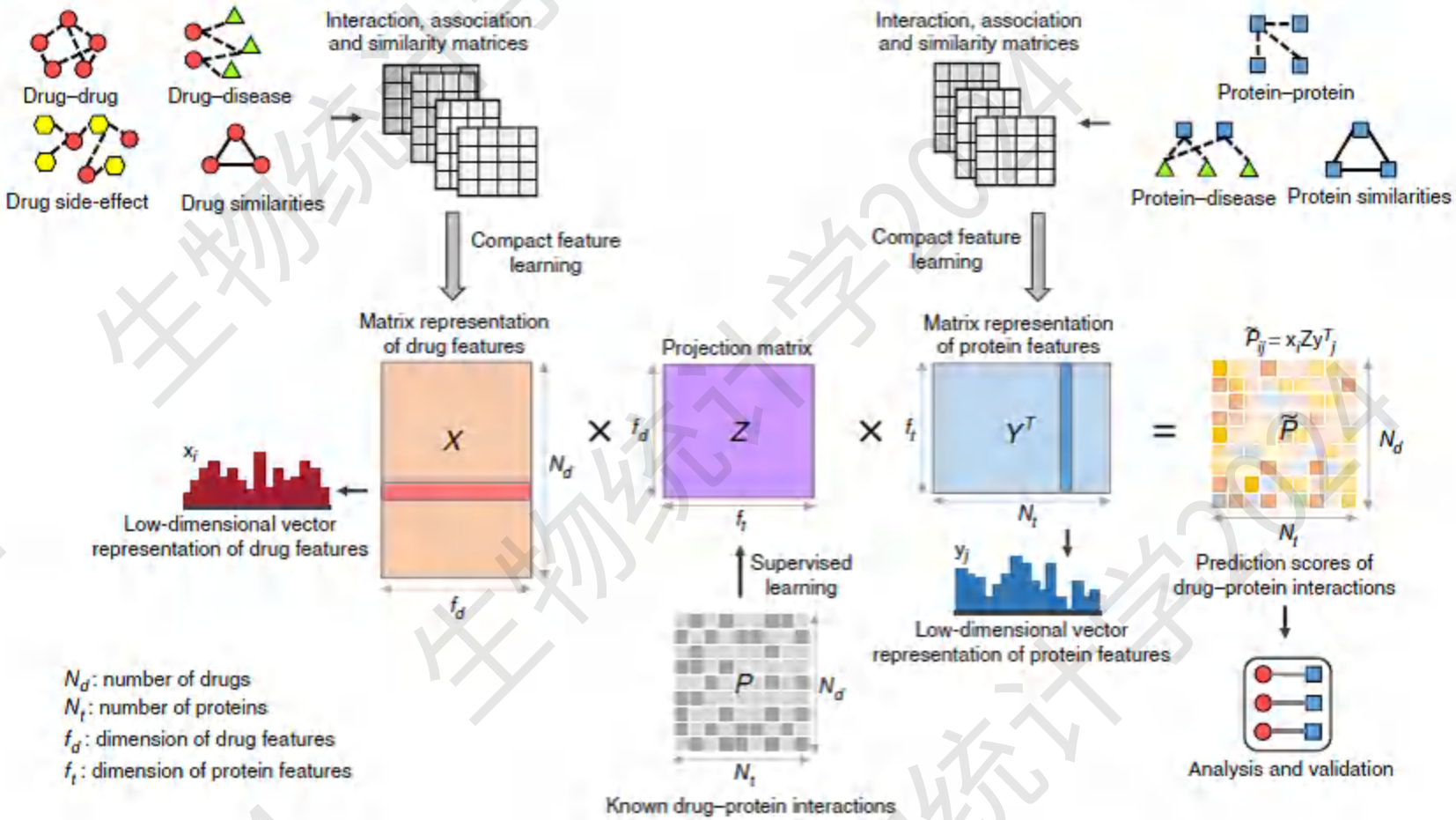
Adapted from "Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature Biotechnology, 2015."

DeepBind总结

- CNN算法自动学习了模版M的参数，这个模版就是能够决定蛋白质是否Binding的特征！
- 而CNN网络就是在前端自动总结出可能的模版，在后端根据这些模版在序列中出现与否而进行判断；
- CNN算法巧妙地将统计归纳和MLP的学习功能相结合，减少了MLP中要学习的参数；

生物统计学2024

Drug-target interaction prediction

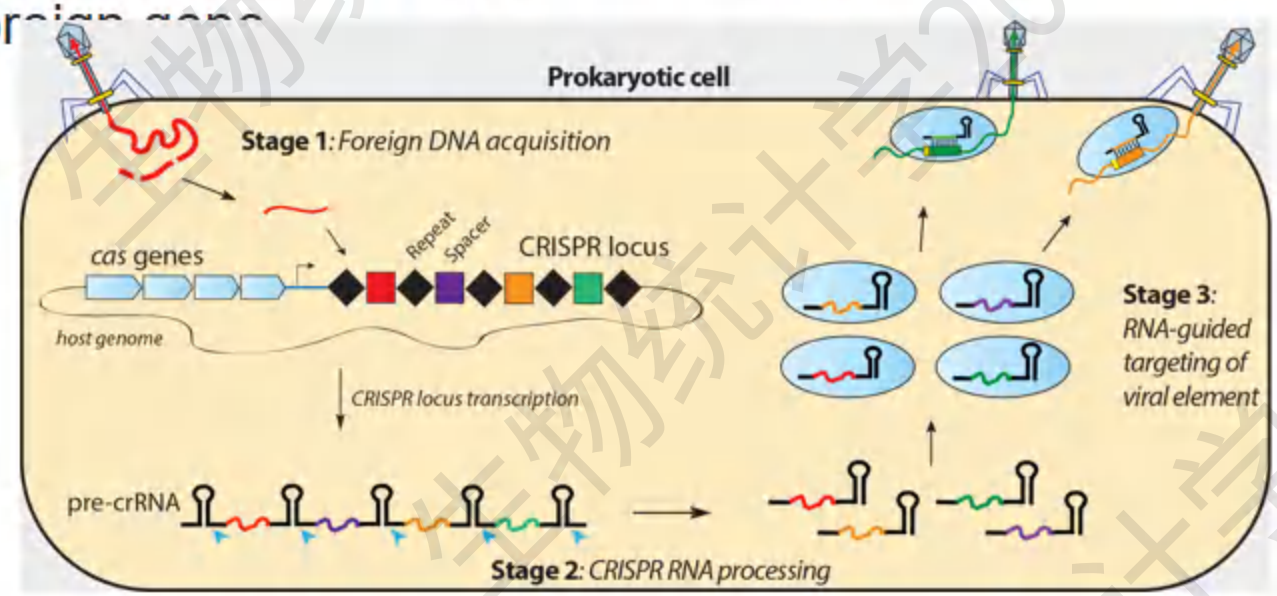


Adapted from "A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature Comunications, 2017"

What is Cas?

- Cas (CRISPR-associated system) genes

bacteria immune system: cleavage the foreign gene



CRISPR-Cas system workflow: foreign DNA acquisition→CRISPR RNA processing→RNA-guided targeting of viral element

生物统计学2024

The existing method for Cas genes prediction

- The existing method for Cas genes prediction

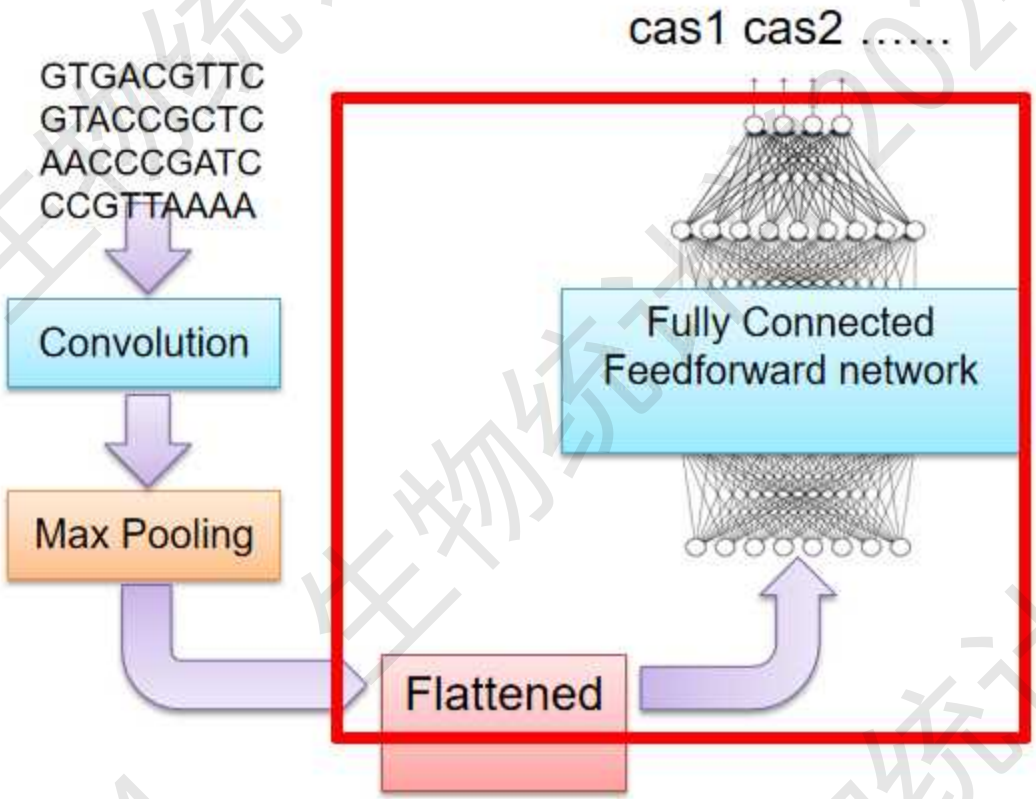
Existing tools	Description
HMMCAS	A web tool for Cas protein identification and domain annotation
CRISPRone	A website tool to predict CRISPR arrays of repeat-spacer units, and cas genes
BLAST	A website tool to find regions of local similarity between sequences
CD-search	A tool allows the conserved domain annotation for large sets of protein queries
Custom HMMs	HMMs deposited in the TIGRFAMs and Pfam protein family databases for known protein families

- Limitations
  - Scan for known Cas proteins.
  - Identify or predict based on the “best hits” of homology searches against existing databases.
  - Work well for detecting known and highly conserved types of Cas genes but may fail to detect novel Cas genes.

生物统计学2024

The whole neural network to predict Cas genes

Mainly based on Python Tensorflow package



生物统计学2024

Data preparation

10 core family,75 family,7500 Cas genes in total

Core family	Structural features
Cas1	N-terminal β stranded domain and catalytic C-terminal α-helical domain
Cas2	RuvC-like nuclease domains
Cas3	Helicase and HD domain
Cas4	RecB-like nuclease homolog with three-cysteine C-terminal cluster
Cas5	RRM (ferredoxin) fold, RAMP superfamily
Cas6	Double RRM (ferredoxin) fold, RAMP superfamily
Cas7	RRM (ferredoxin) fold with subdomains, RAMP superfamily
Cas8	Subunit of Cascade complex, involved in PAM recognition
Cas9	RuvC-like (RNAse H fold) and HNH nuclease domains
Cas10	Two domains homologous to Palm domain polymerases and cyclases

生物统计学2024

Data preparation

one-hot encoding, the 4 nucleotides as binary vectors A = [1, 0, 0, 0], T=[0, 1, 0, 0], G=[0, 0, 1, 0] and C=[0, 0, 0, 1].

- Raw data format:

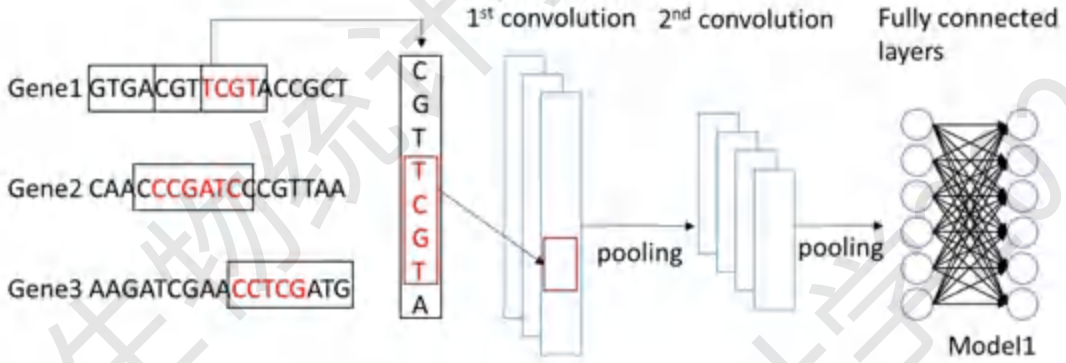
```
>CP001154.1_cds_ACO74068.1_1072
GTGGAGCTGCTGGTCGCCCCGGACCTCAAAACCATCCTGCATTCCGGGCGGGCCAGCCTTCATGGCCTGGAATACTGCCG
>CP010338.1_cds_AMC82843.1_3024
ATGCGTACCGAAGCCATGAACCCGCAACTCATCGAGGCCATAATCGGCTGCCTCTTGCACGACATTGGCAAACCGGTCCA
>CP002567.1_cds_AFI63574.1_1345
ATGAGCAAGATTCCCGGAGTAAAACCACCAGAATTGAAATCGTTGGTGCGGGTGCAGGATAGGCTGACGTTCCTGTATGC
>CP018217.1_cds_APG74590.1_777
ATGTCATGGAGATATGTATTTGTCACTAATCCGTGTAGGATAACGTATACCAACAATAGTATAGTTGTTCAAACTTTCGA
>CP012396.1_cds_AOC85297.1_899
GTGACGTTCGTACCGCTCAACCCGATCCCGTTAAAAGATCGAACCTCGATGATCTTCCTCCAGTACGGTCAAATTGACGT
```

- After change to one_hot:

```
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1
1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1
1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1
```

Sequence label，save in an individual file

Convolution Neural Networks model



- Design a neural network that will classify cas proteins.
- Given a cas gene sequence, we want our neural net to let us know which cas gene that it belong to.
- The sequence is be made up of A, T, C,G, and fasta format:
- the neural net will have xxx inputs, each one representing a particular sequence and a hidden layer consisting of a number of neurons (more on this later) all feeding their output into just one neuron in the output layer.

A convolutional layer

A CNN is a neural network with some convolutional layers (and some other layers). A convolutional layer has a number of filters that does convolutional operation.

1	0	0	0
0	1	0	0
0	0	0	1
1	0	0	0
⋮	⋮	⋮	⋮
0	0	1	0

80 x 4

These are the network parameters to be learned

1	-1	-1	-1
-1	1	-1	-1
-1	-1	-1	1
1	-1	-1	-1

Filter 1

-1	1	-1	-1
-1	-1	1	-1
-1	-1	-1	1
1	-1	-1	-1

Filter 2

.....

Each filter detects a small pattern (4 x 4).

生物统计学2024

A convolutional layer

stride=1

1	0	0	0
0	1	0	0
0	0	0	1
1	0	0	0
. .	. .	. .	. .
0	0	1	0

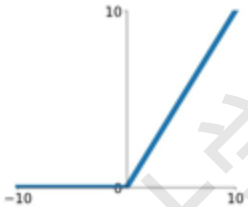
80 x 4

Dot product →

4

1	-1	-1	-1
-1	1	-1	-1
-1	-1	-1	1
1	-1	-1	-1

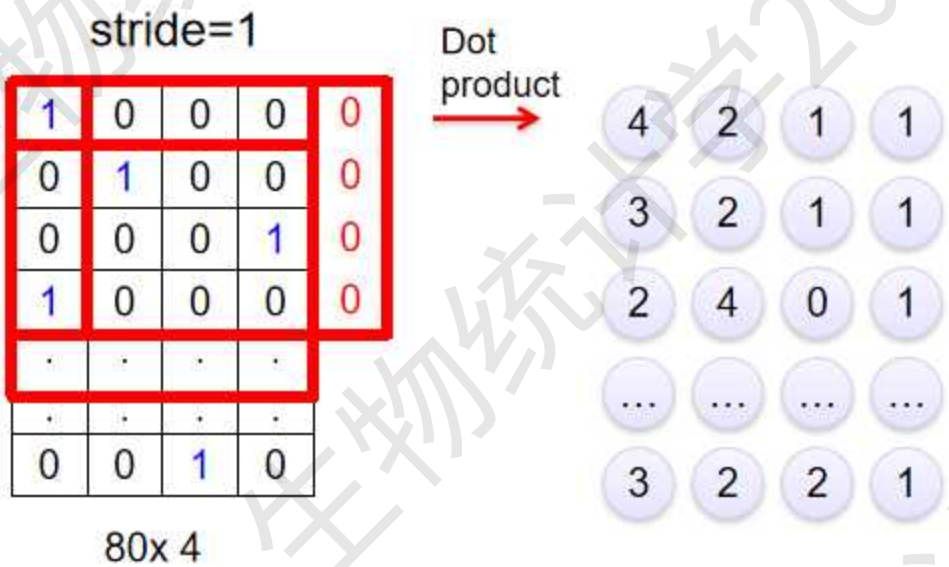
Filter 1



$\mathrm{ReLU}(x) = \max(0, \ x)$

生物统计学2024

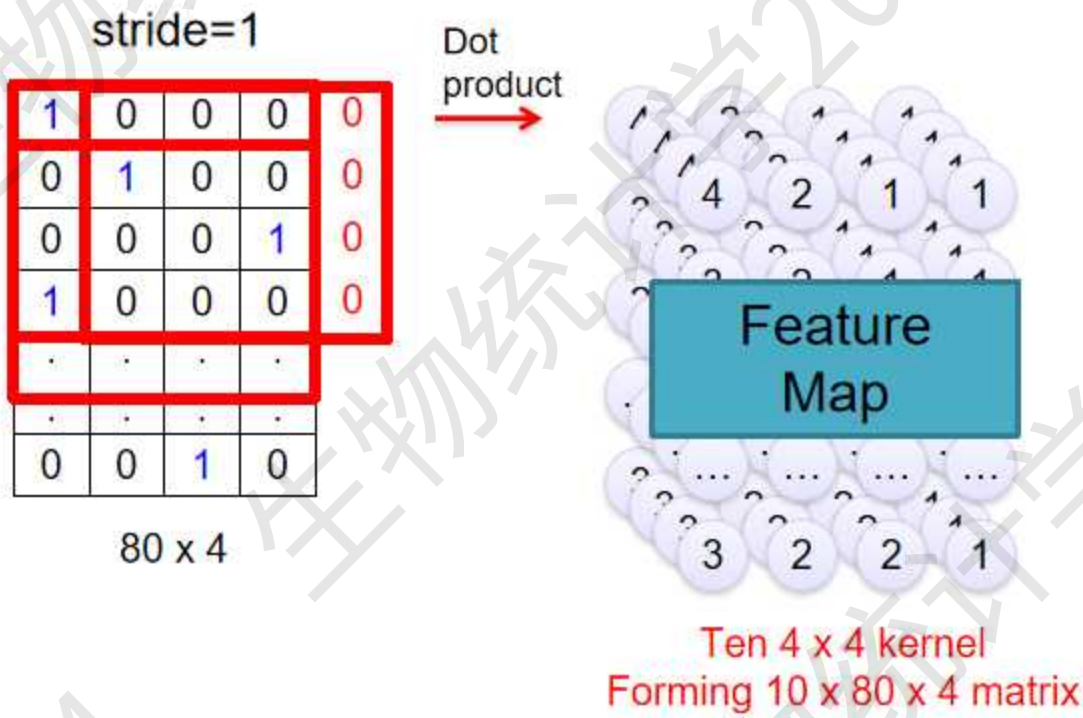
A convolutional layer



Repeat this for each filter

生物统计学2024

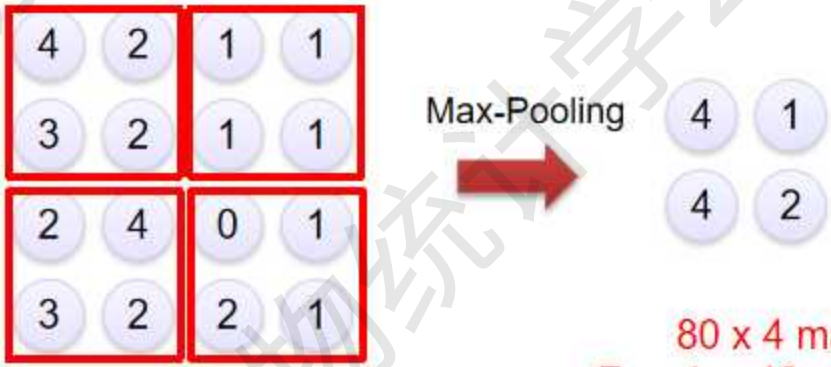
A convolutional layer



生物统计学2024

A convolutional layer

$$p_{nfi} = \max_{|k|<P/2}(a_{nf,i+k})$$



80 x 4 matrix
Forming 40 x 2 matrix

生物统计学2024

CNN model construction in Tensorflow

input

Convolution

```
weights = tf.get_variable('weights',
                          shape=[4, 4, 1, 320],
                          dtype=tf.float32,
                          initializer=tf.truncated_normal_initializer(stddev=0.1,
                                                                      dtype=tf.float32))
biases = tf.get_variable('biases',
                         shape=[320],
                         dtype=tf.float32,
                         initializer=tf.constant_initializer(0.1))
conv = tf.nn.conv2d(features, weights, strides=[1,1,1,1], padding='SAME')
```

1	-1	-1	-1	
-1	-1	-1	-1	-1
-1	1	-1	-1	-1
1	-1	-1	1	-1
	-1	-1	-1	-1

…

…

There are 320 4x4 filters.

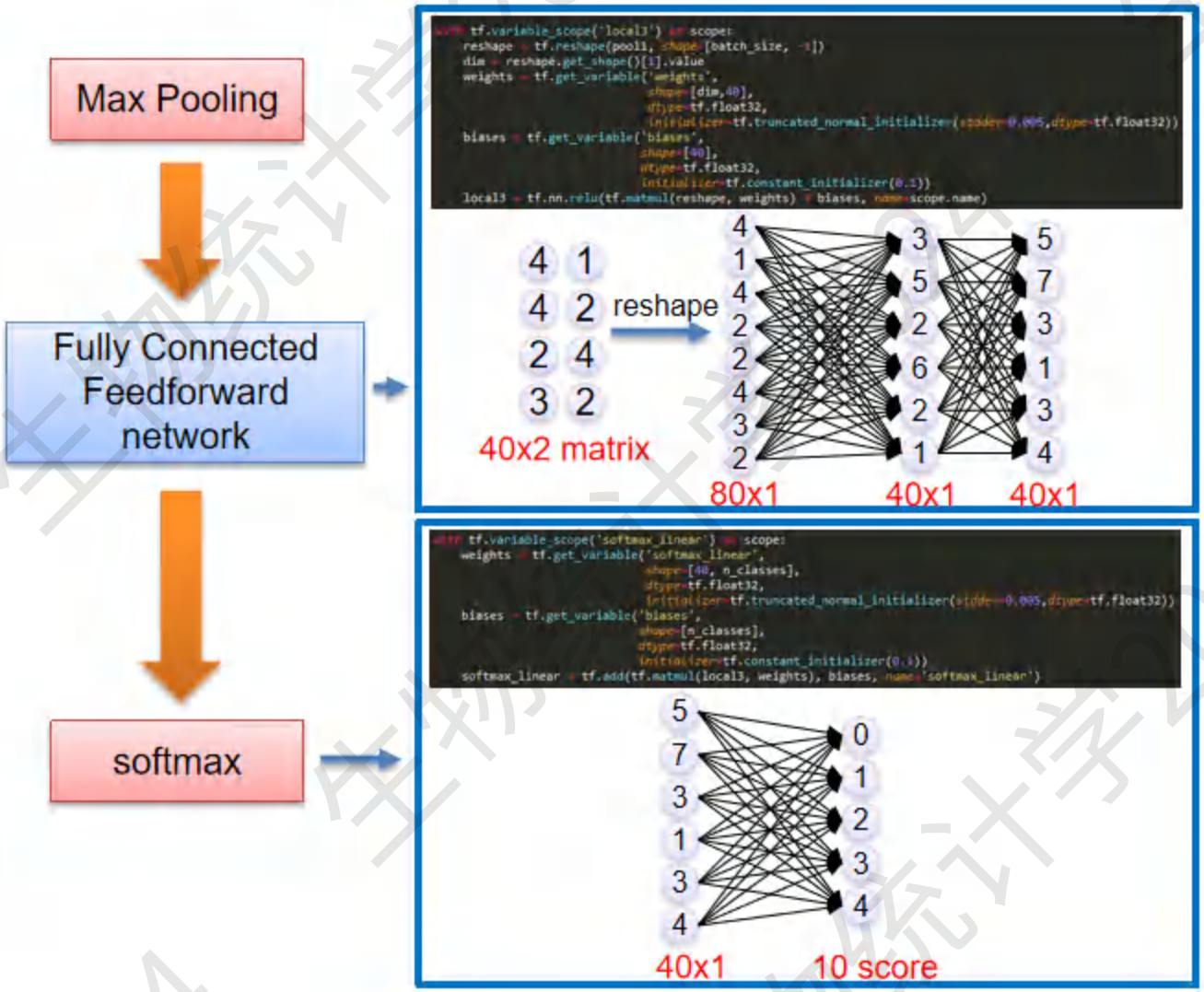
Input_shape = (80 , 4 , 1)

80 x 4 pixels

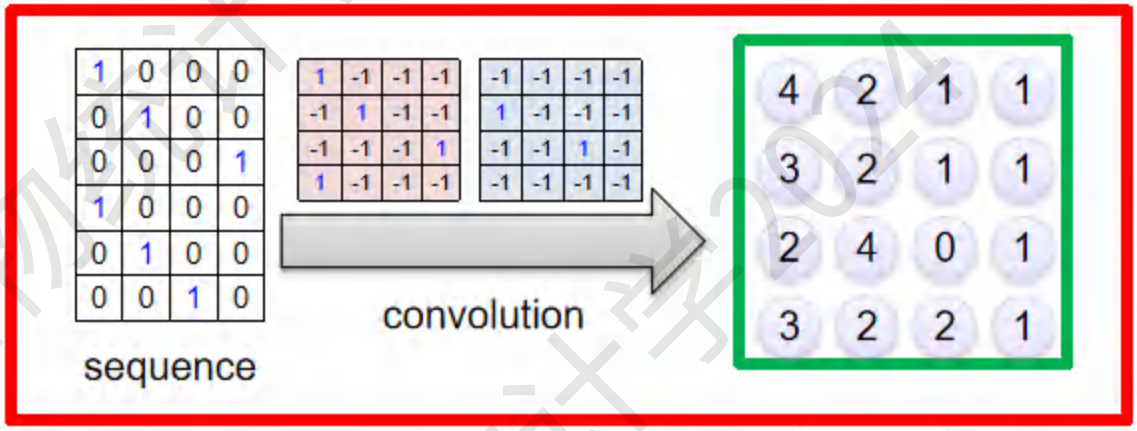
1: black/white, 3: RGB

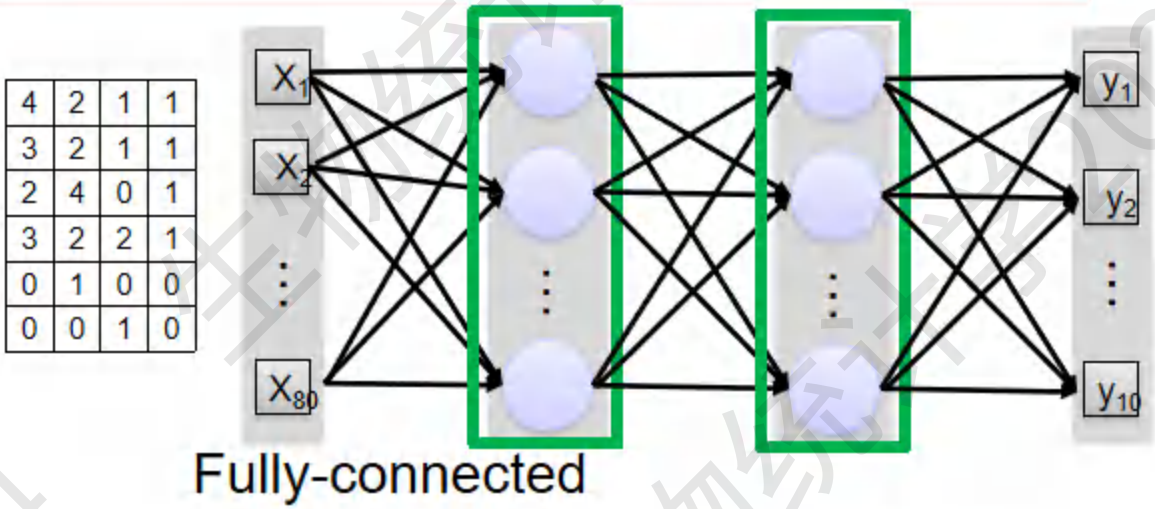
Max Pooling

```
pool1 = tf.nn.max_pool(conv1, ksize=[1,3,3,1],strides=[1,2,2,1],
                       padding='SAME', name='pooling1')
```


Max Pooling
Fully Connected Feedforward network
softmax
with tf.variable_scope('local3') as scope:
reshape = tf.reshape(pool1, shape=[batch_size, -1])
dim = reshape.get_shape()[1].value
weights = tf.get_variable('weights',
shape=[dim,40],
dtype=tf.float32,
initializer=tf.truncated_normal_initializer(stddev=0.005,dtype=tf.float32))
biases = tf.get_variable('biases',
shape=[40],
dtype=tf.float32,
initializer=tf.constant_initializer(0.1))
local3 = tf.nn.relu(tf.matmul(reshape, weights) + biases, name=scope.name)
4 1
4 2
2 4
3 2
reshape
40x2 matrix
4 1 4 2 2 4 3 2
3 5 2 6 2 1
5 7 3 1 3 4
80x1
40x1
40x1
with tf.variable_scope('softmax_linear') as scope:
weights = tf.get_variable('softmax_linear',
shape=[40, n_classes],
dtype=tf.float32,
initializer=tf.truncated_normal_initializer(stddev=0.005,dtype=tf.float32))
biases = tf.get_variable('biases',
shape=[n_classes],
dtype=tf.float32,
initializer=tf.constant_initializer(0.1))
softmax_linear = tf.add(tf.matmul(local3, weights), biases, name='softmax_linear')
5 7 3 1 3 4
0 1 2 3 4
40x1
10 score

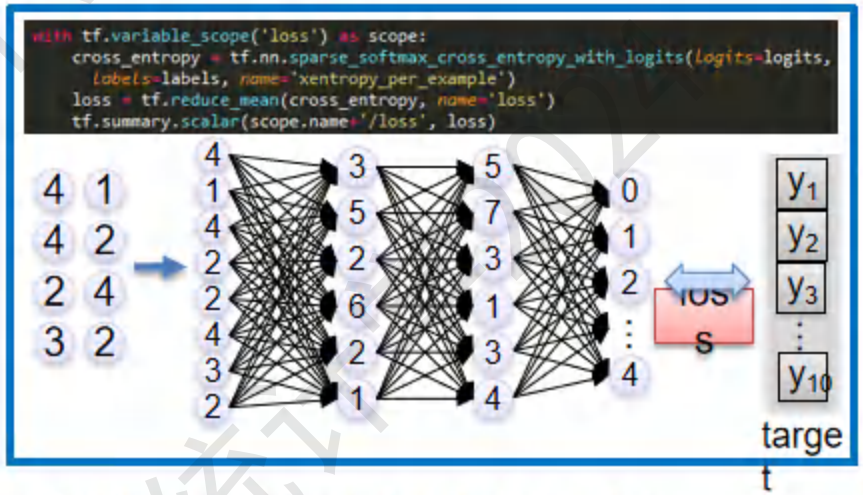
Convolution v.s. Fully Connected





Pick the best model

loss



Randomly initialize network parameters
Pick the 1st batch
$L'=l_1+l_2+\cdots$
Update parameters once
Pick the 2nd batch
$L''=l_1+l_2+\cdots$
Update parameters once
…
Until all mini-batches have been picked

training

```
N_CLASSES = 2
seq_W = 20  # reaize the seq, if the input seq is too large, training will be very slow.
seq_H = 80
BATCH_SIZE = 16
CAPACITY = 2000
MAX_STEP = 10000 # with current parameters, it is suggested to use MAX_STEP>10k
learning_rate = 0.0001 # with current parameters, it is suggested to use learning rate<0.0001
n = 1601
```

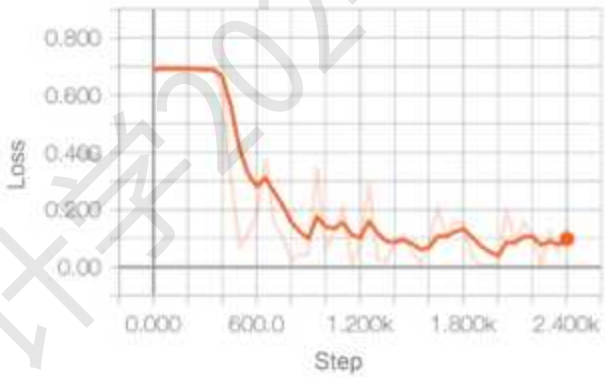
Output:

```
Step 0, train loss = 0.69, train accuracy = 62.50%
Step 50, train loss = 0.69, train accuracy = 56.25%
Step 100, train loss = 0.69, train accuracy = 43.75%
Step 150, train loss = 0.69, train accuracy = 56.25%
Step 200, train loss = 0.68, train accuracy = 93.75%
Step 250, train loss = 0.68, train accuracy = 43.75%
Step 300, train loss = 0.57, train accuracy = 68.75%
Step 350, train loss = 0.45, train accuracy = 68.75%
Step 400, train loss = 0.36, train accuracy = 75.00%
Step 450, train loss = 0.41, train accuracy = 87.50%
```

生物统计学2024

Training Module1

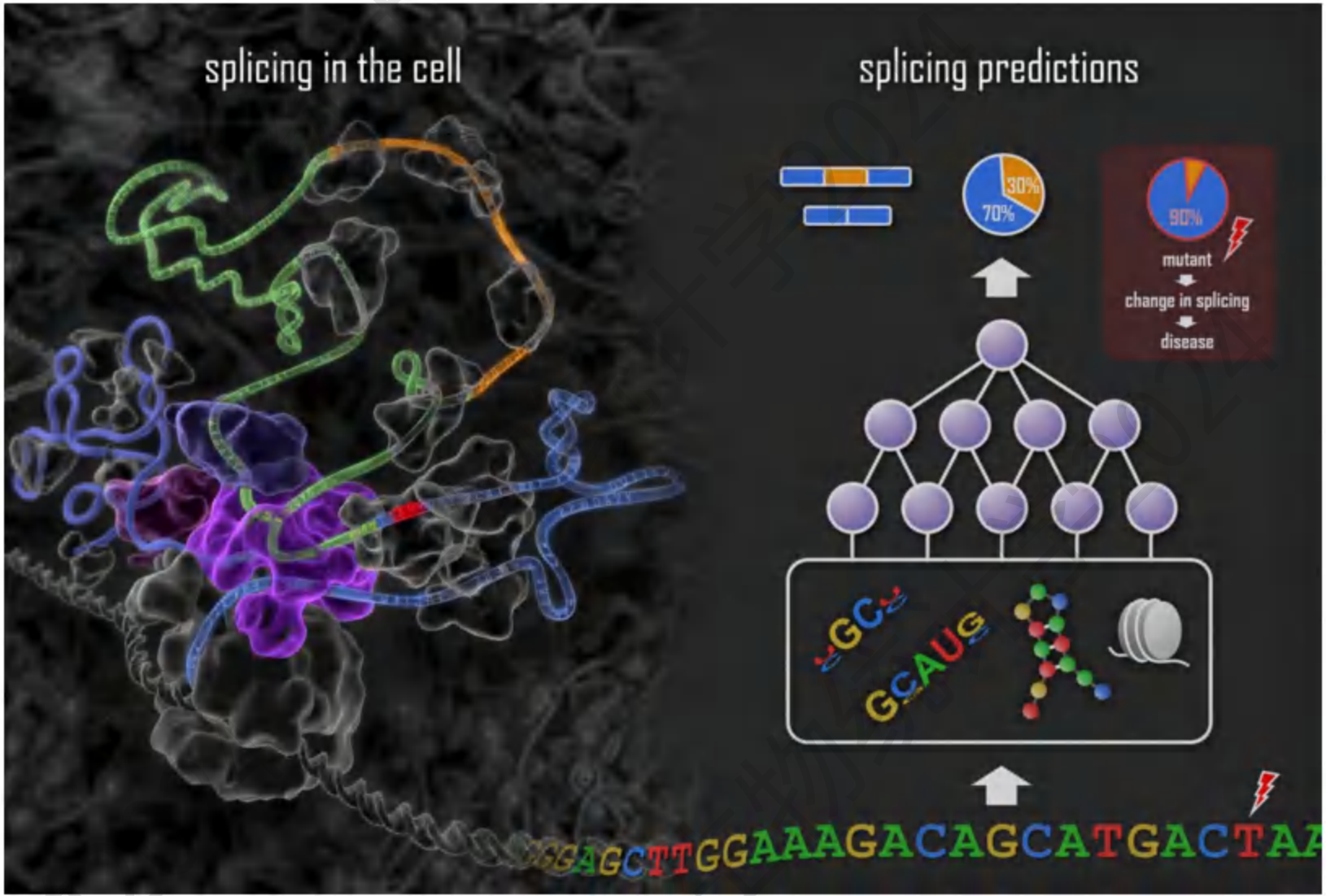
- Data: 5000 Cas9 sequences with domain sites and 5000 without domain sites



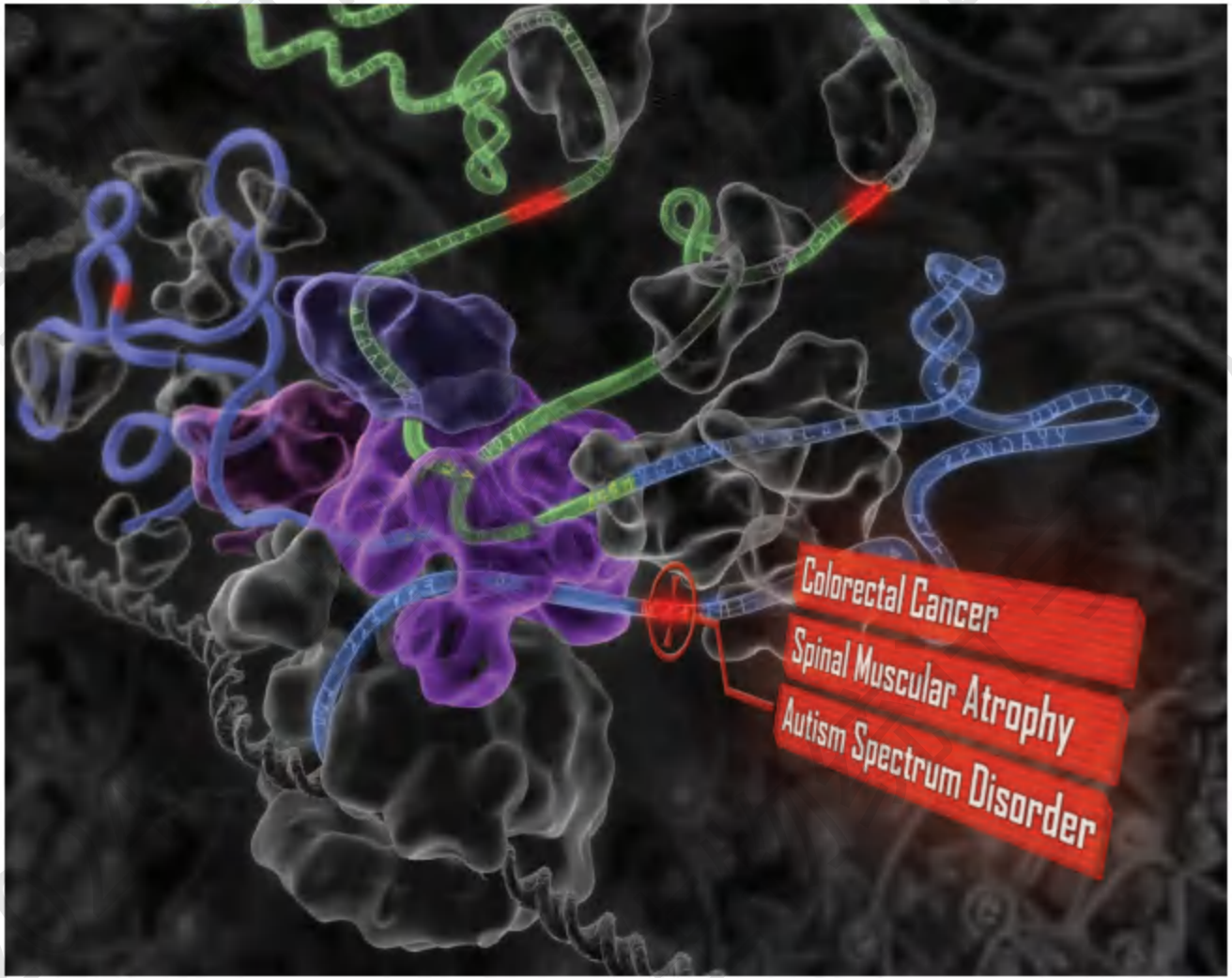
Accuracy：0.80225

生物统计学2024

New variant discovery

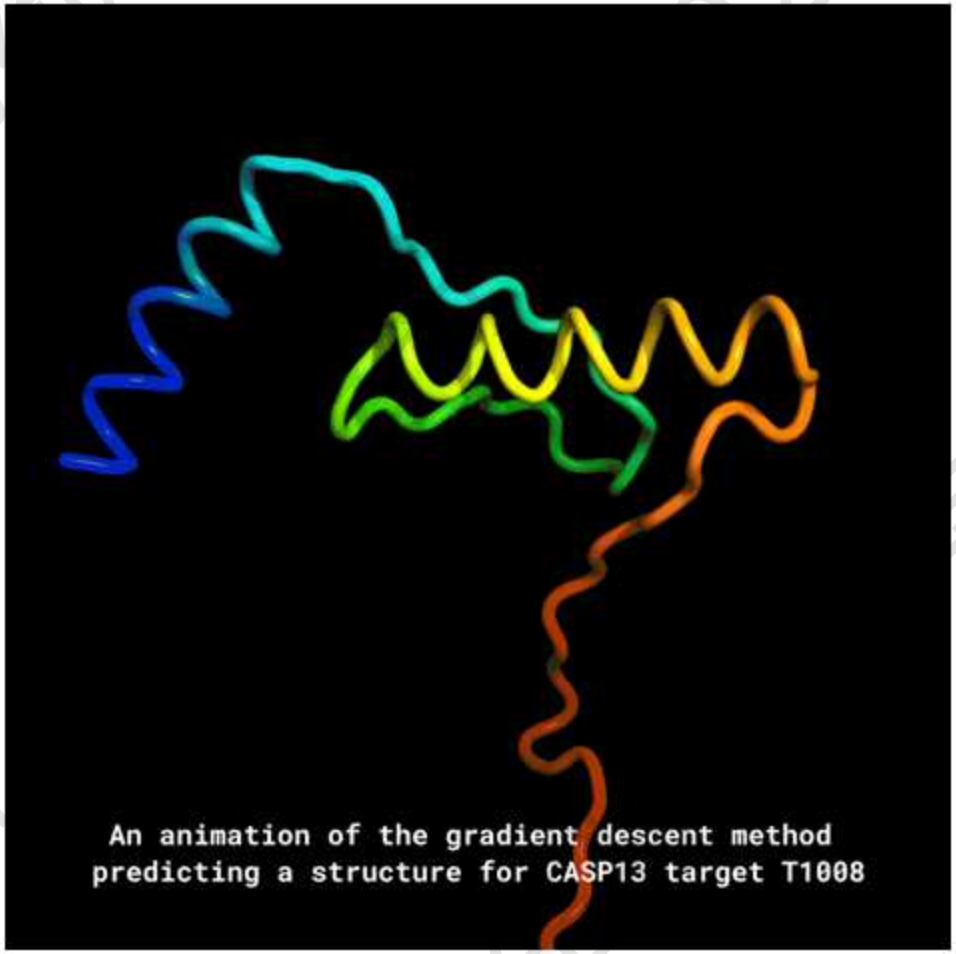


New variant discovery



CASP

(Critical Assessment of Techniques for Protein Structure Prediction)

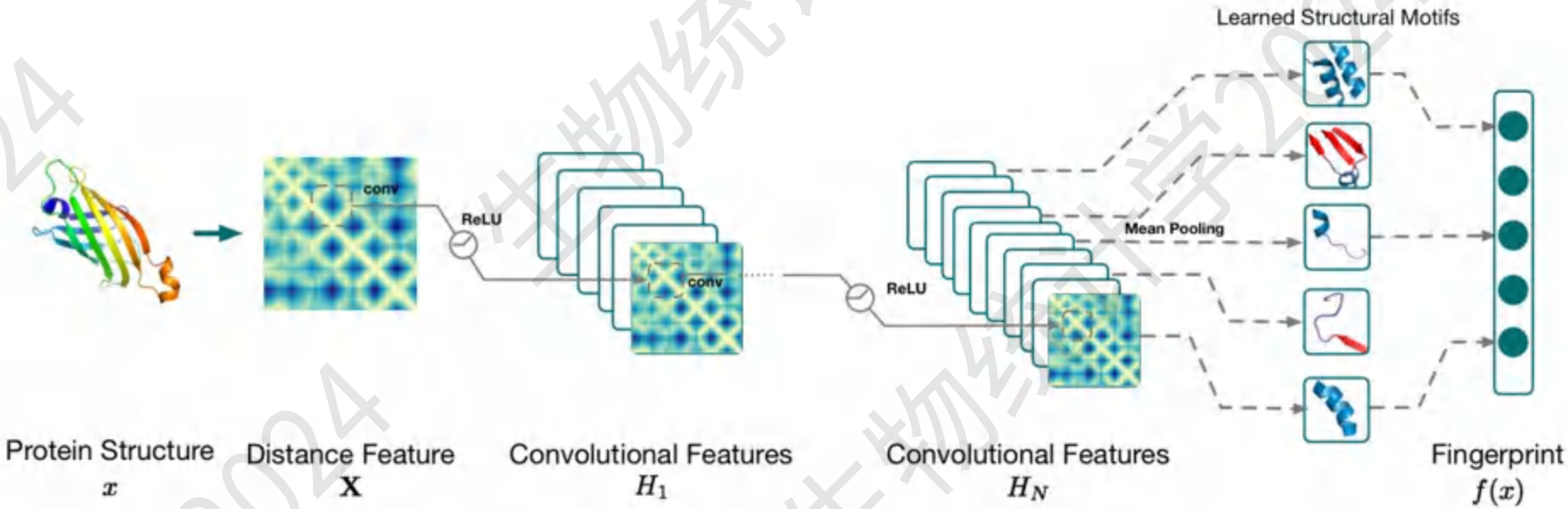


生物信息学2024

CASP13

(Critical Assessment of Techniques for Protein Structure Prediction)

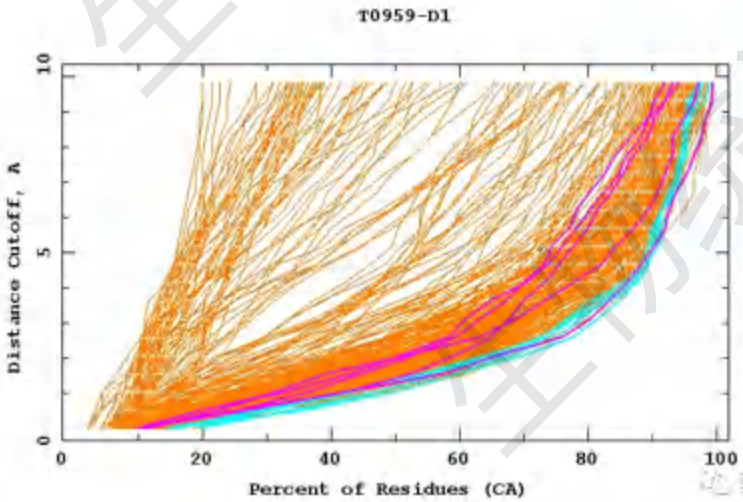
DeepFold

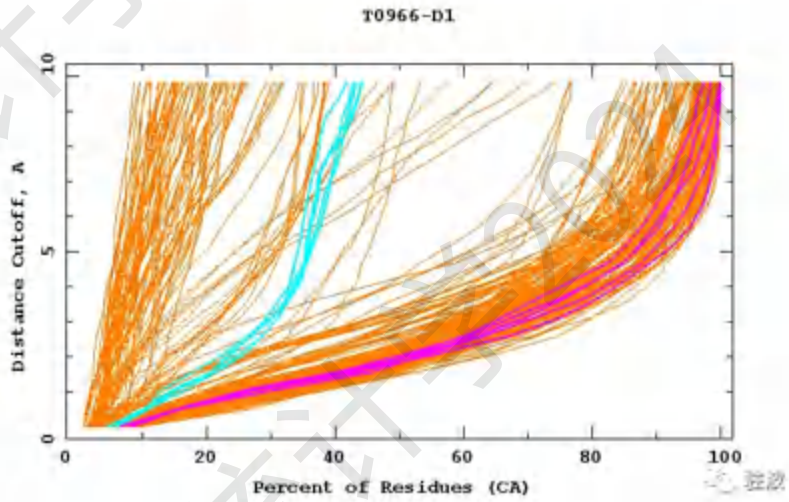


生物统计学2024

CASP13

(Critical Assessment of Techniques for Protein Structure Prediction)





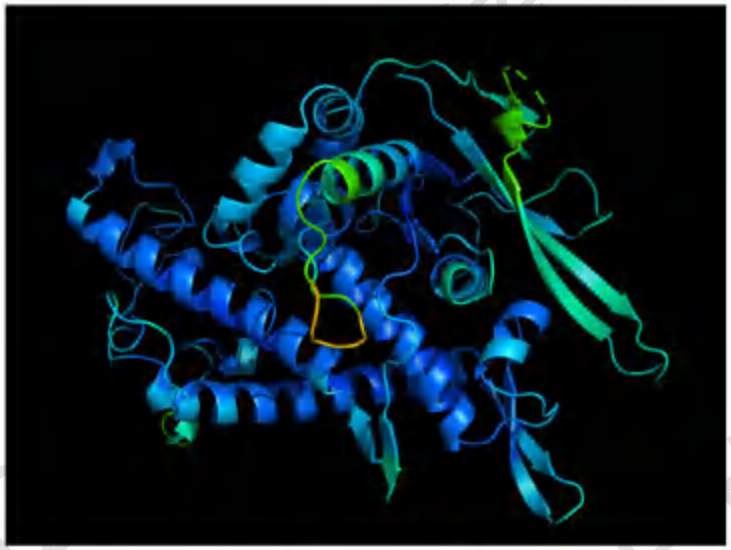
生物统计学2024

NEWS · 30 NOVEMBER 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway



A protein's function is determined by its 3D shape. Credit: DeepMind

RELATED ARTICLES

AI protein-folding algorithms solve structures faster than ever

The revolution will not be crystallized: a new method sweeps through structural biology

The computational protein designers

Revolutionary microscopy technique sees individual atoms for first time

SUBJECTS

生物统计学2024

CASP14

(Critical Assessment of Techniques for Protein Structure Prediction)

AlphaFold: a solution to a 50-year-old grand challenge in biology

DeepMind宣布，其新一代AlphaFold人工智能系统，在国际蛋白质结构预测竞赛（CASP）上击败了其余的参会选手，能够精确地基于氨基酸序列，预测蛋白质的3D结构。其准确性可以与使用冷冻电子显微镜（CryoEM）、核磁共振或 X 射线晶体学等实验技术解析的3D结构相媲美。

生物统计学2024

CASP14

(Critical Assessment of Techniques for Protein Structure Prediction)

Sundar Pichai @sundarpichai · 10h

.@DeepMind's incredible AI-powered protein folding breakthrough will help us better understand one of life's fundamental building blocks + enable researchers to tackle new and hard problems, from fighting diseases to environmental sustainability.

AlphaFold: a solution to a 50-year-old grand challenge in biology

In a major scientific advance, the latest version of our AI system AlphaFold has been recognised as a solution to this grand challenge ...

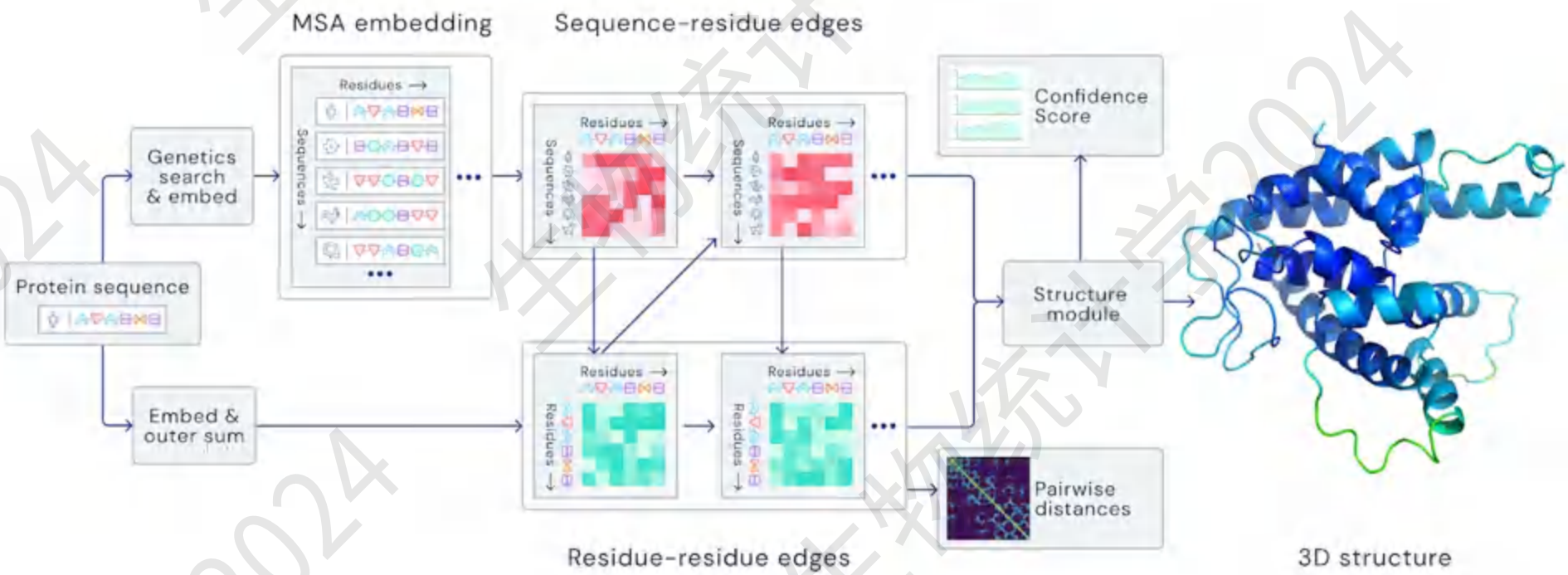
deepmind.com

生物统计学2024

CASP14

(Critical Assessment of Techniques for Protein Structure Prediction)

AlphaFold2



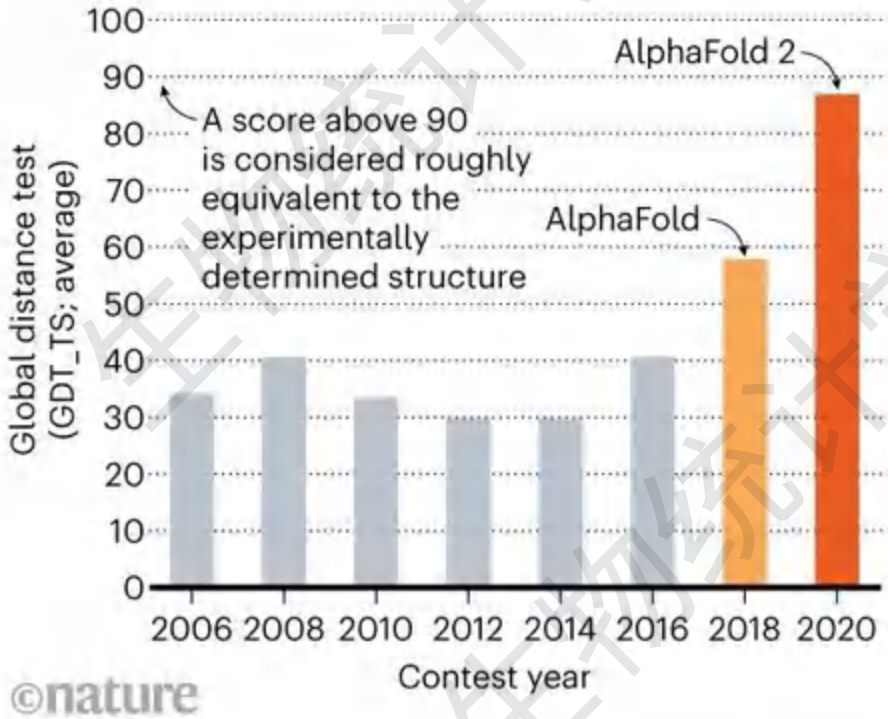
生物统计学2024

CASP14

(Critical Assessment of Techniques for Protein Structure Prediction)

STRUCTURE SOLVER

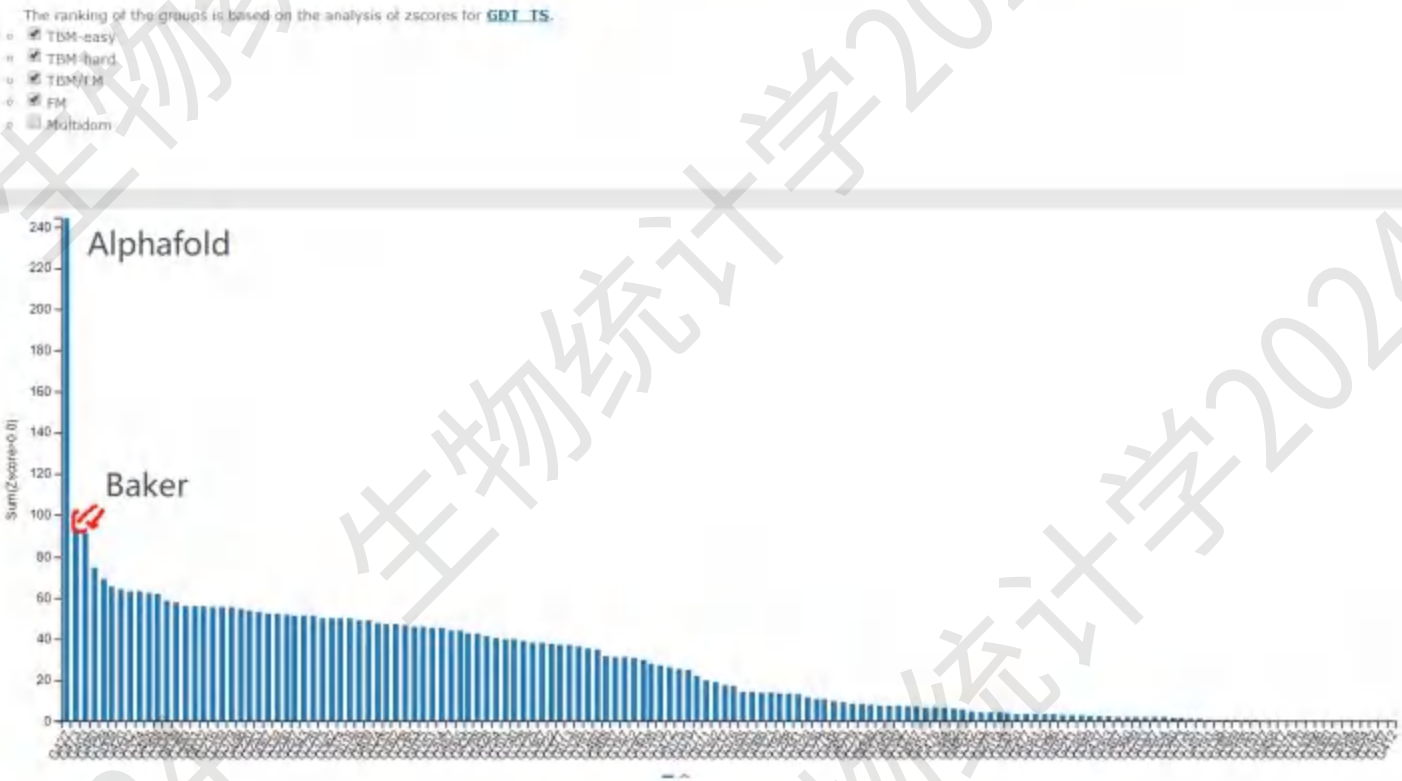
DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.



生物统计学2024

CASP14

(Critical Assessment of Techniques for Protein Structure Prediction)

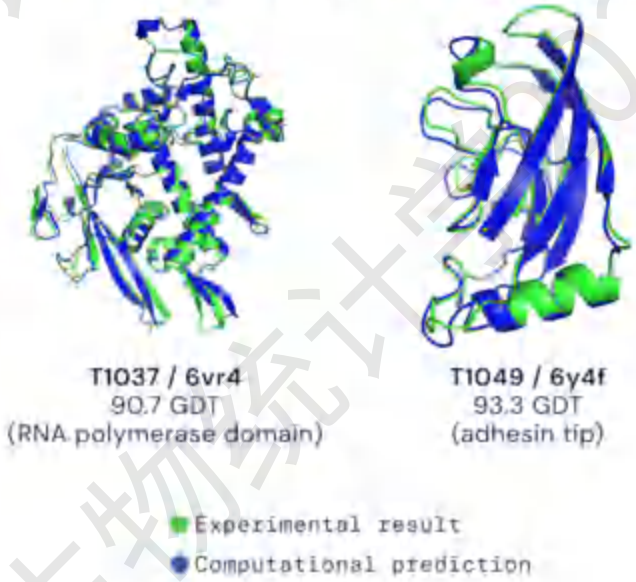


https://predictioncenter.org/casp14/zscores_final.cgi

生物统计学2024

CASP14

(Critical Assessment of Techniques for Protein Structure Prediction)



在2020的CASP中，AlphaFold系统对所有蛋白靶点3D结构预测的中位GDT评分为92.4分。即便是针对最难解析的蛋白靶点，AlphaFold的中位GDT评分也达到了87.0分。在接受检验的近100个蛋白靶点中，AlphaFold对三分之二的蛋白靶点给出的预测结构与实验手段获得的结构相差无几。CASP创始人Moult教授表示，在有些情况下，已经无法区分两者之间的区别是由于AlphaFold的预测出现错误，还是实验手段产生的假象。

生物统计学2024

CASP14
(Critical Assessment of Techniques for Protein Structure Prediction)

- 我们可以把蛋白质折叠看作一个「空间图」，节点表示残基（residue），边则将残基紧密连接起来。这个空间图对于理解蛋白质内部的物理交互及其演化史至关重要。对于在 CASP14 比赛中使用的最新版 AlphaFold，DeepMind 团队创建了一个基于注意力的神经网络系统，并用端到端的方式进行训练，以理解图结构，同时基于其构建的隐式图执行推理。该方法使用进化相关序列、多序列比对（MSA）和氨基酸残基对的表示来细化该图。

- DeepMind 团队在公开数据上训练这一系统，这些数据来自蛋白质结构数据库（PDB）和包含未知结构蛋白质序列的大型数据库，共包括约 170,000 个蛋白质结构。该系统使用约 128 个 TPUv3 内核（相当于 100-200 个 GPU）运行数周，与现今机器学习领域出现的大型 SOTA 模型相比，该系统所用算力相对较少。

- AlphaFold 还具备很多令人兴奋的技术潜力：探索数亿个目前还没有模型的数亿蛋白质，以及未知生物的广阔领域。由于 DNA 指定了构成蛋白质结构的氨基酸序列，基因组学革命使大规模阅读自然界的蛋白质序列成为可能——在通用蛋白质数据库（UniProt）中有 1.8 亿个蛋白质序列。相比之下，考虑到从序列到结构所需的实验工作，蛋白质数据库（PDB）中只有大约 170000 个蛋白质结构。在未确定的蛋白质中可能有一些新的和未确定的功能——就像望远镜帮助人类更深入的观察未知宇宙一样，像 AlphaFold 这样的技术可以帮助找到未确定的蛋白质结构。

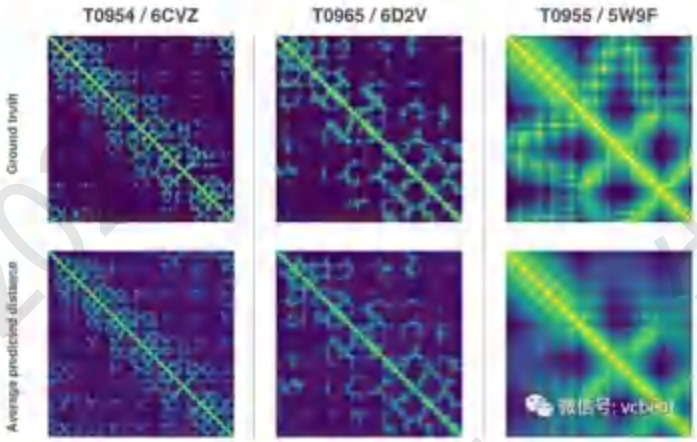
CASP14
(Critical Assessment of Techniques for Protein Structure Prediction)

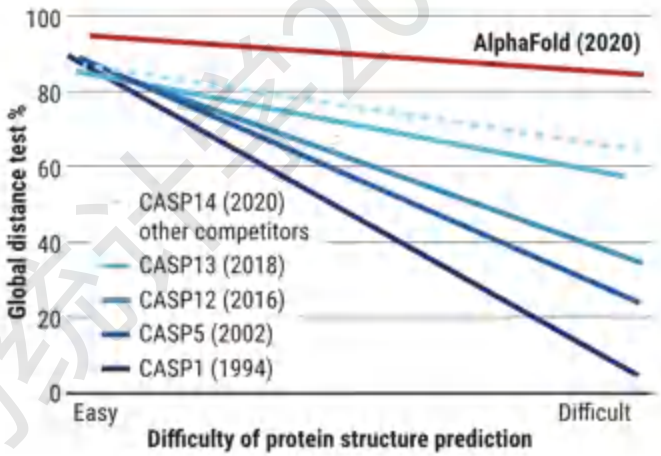
- 在今年早些时候，DeepMind已经利用这一系统预测了多种新冠病毒蛋白的结构。后续的实验显示，AlphaFold预测的新冠病毒Orf3a蛋白结构与冷冻电镜解析的结构非常相似。

- 虽然，AlphaFold不见得会取代冷冻电子显微镜等其它实验手段，但是DeepMind的研究人员表示，这一令人兴奋的结果表明，生物学家们可以使用计算结构预测作为科学研究的核心工具之一。这一手段对于特定类型的蛋白来说可能尤为便利，例如膜蛋白一直非常难于结晶，因此很难用实验手段获得它们的结构。

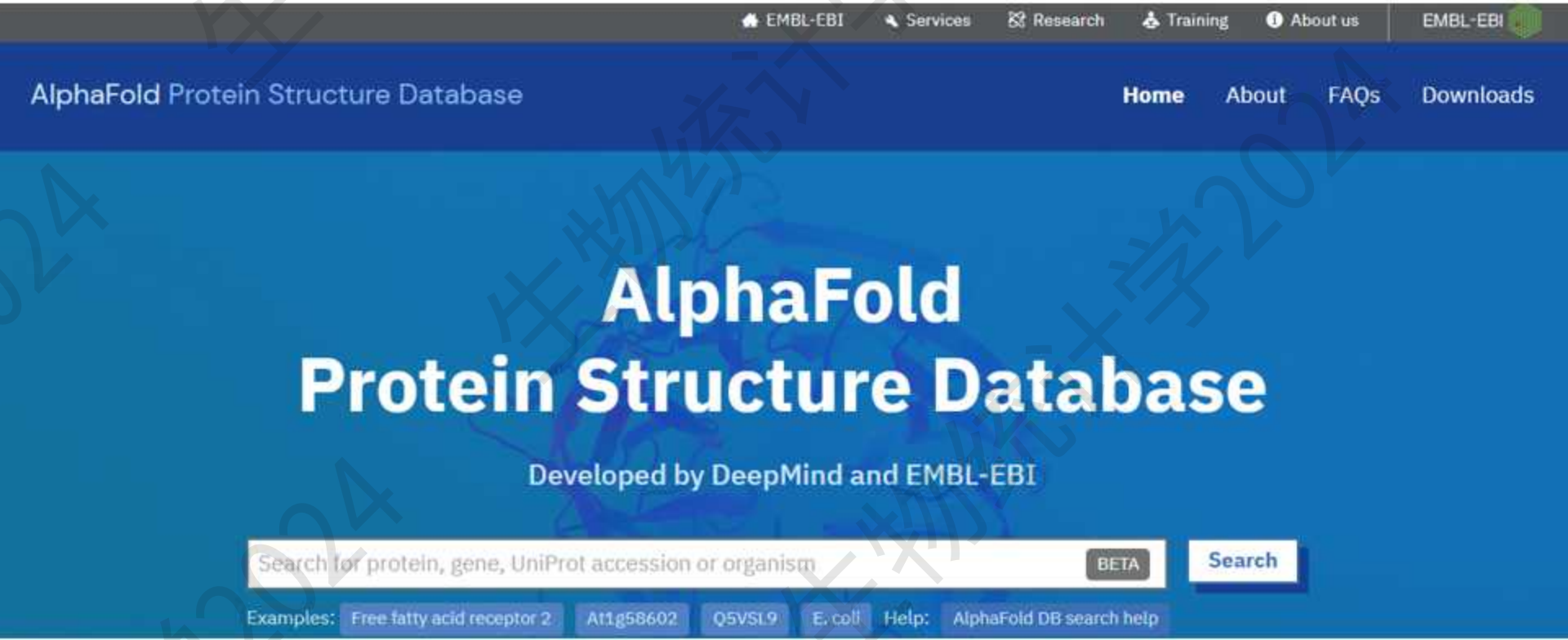
- 而对于从事计算和机器学习研究的DeepMind团队来说，AlphaFold的表现证明了AI在辅助基础科学发现方面惊人的潜力。该团队在公司发布的博文中表示，他们相信，AI将成为人类拓展科学知识前沿最有力的工具之一！

AI + 蛋白质3D结构


T0954 / 6CVZ
T0965 / 6D2V
T0955 / 5W9F
Ground truth
Average predicted distance


100
80
60
40
20
0
Global distance test %
AlphaFold (2020)
CASP14 (2020)
other competitors
CASP13 (2018)
CASP12 (2016)
CASP5 (2002)
CASP1 (1994)
Easy
Difficult
Difficulty of protein structure prediction

AI + 蛋白质3D结构

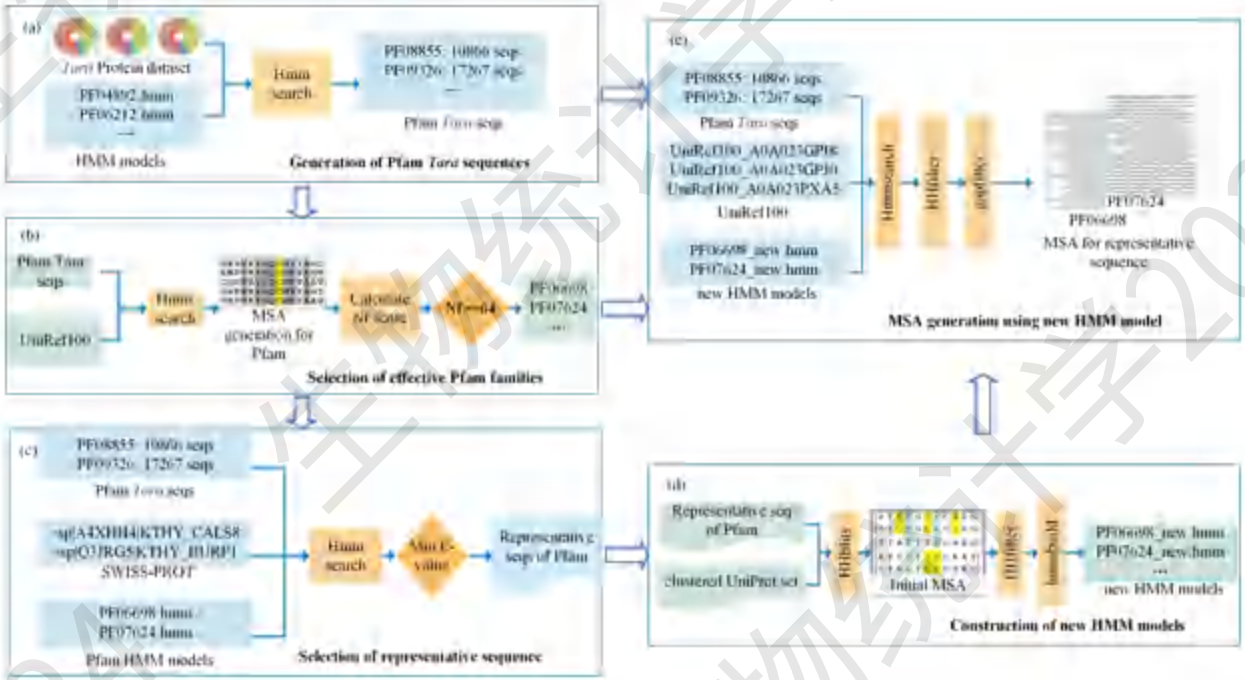


生物统计学2024

微生物组大数据 + 蛋白质3D结构

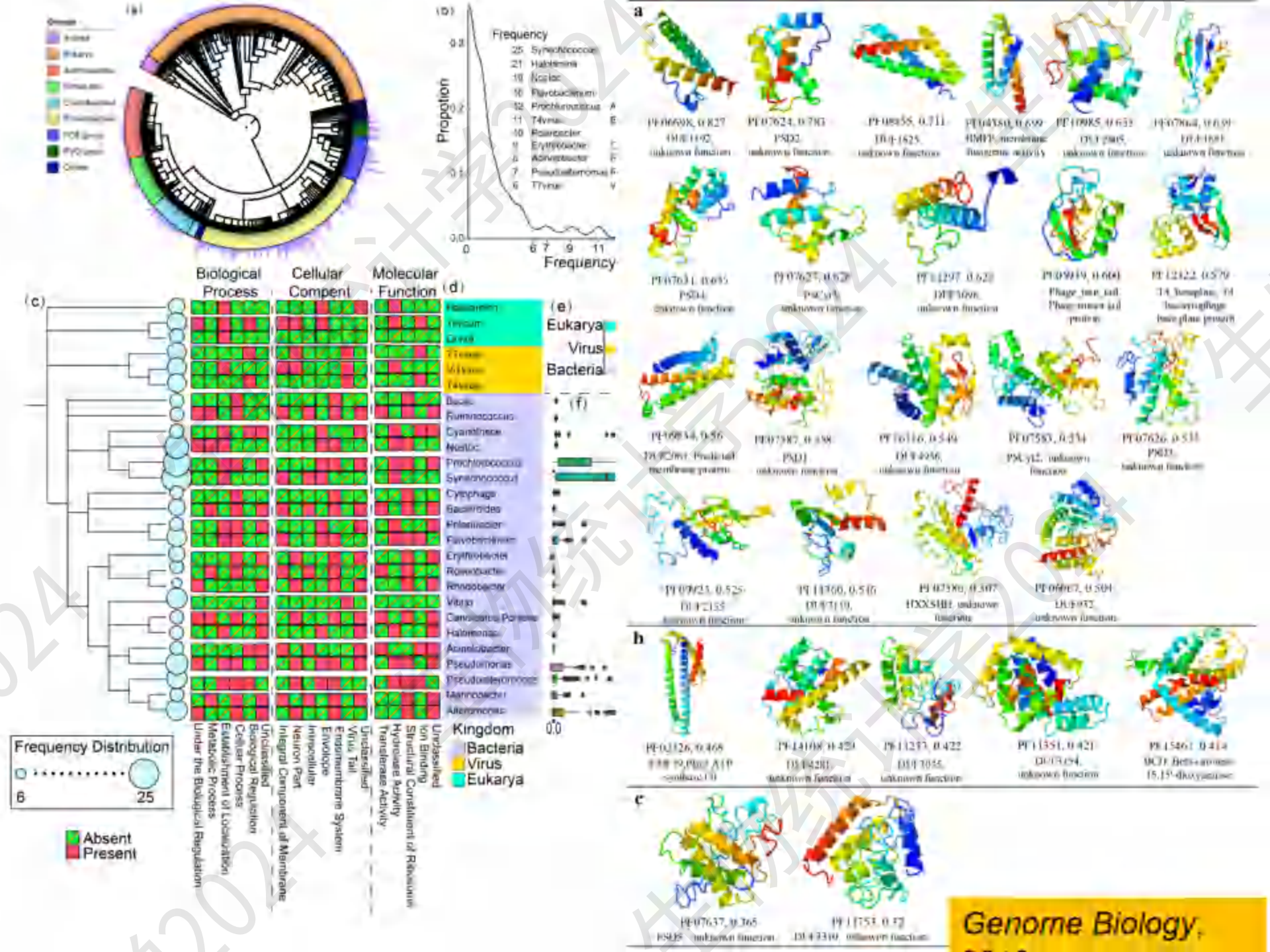
基于大数据挖掘的蛋白质结构预测和功能解析

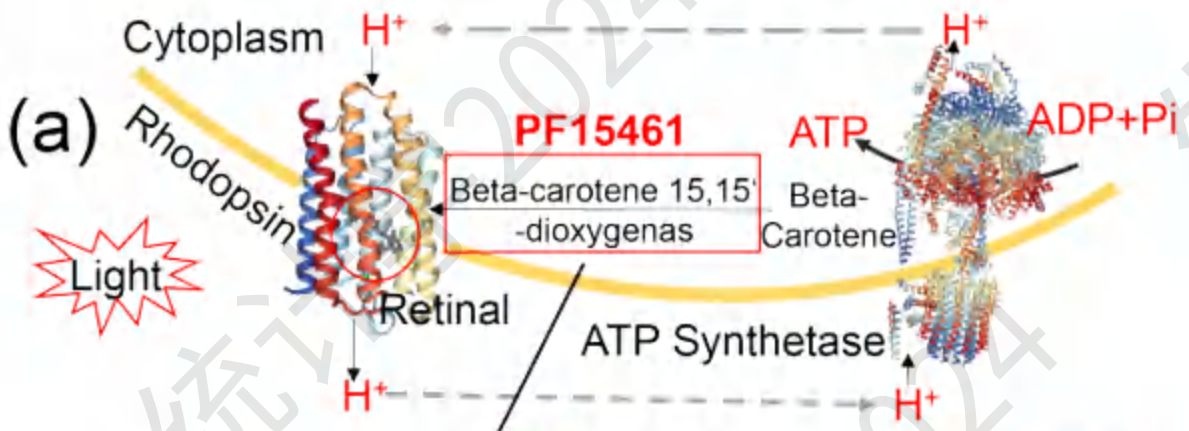
- 2TB的海洋微生物组大数据-->鉴定出了超过9千万的非冗余基因和超过3万个微生物物种
- 预测出了之前没有任何结构信息的27个蛋白质结构
- 利用人工神经网络挖掘蛋白结构和生境的关系

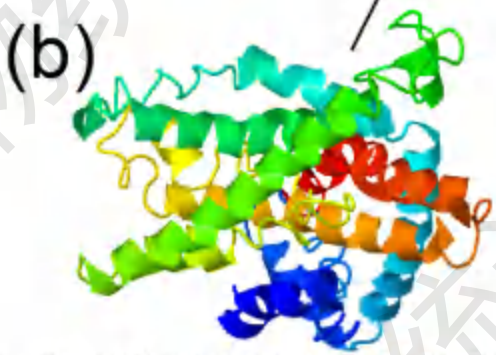


● 宁康团队

Genome Biology,


(a)
(b)
Frequency
Propotion
Frequency
(c)
Biological Process
Cellular Compent
Molecular Function
(d)
(e)
Eukarya
Virus
Bacteria
(f)
Kingdom
Bacteria
Virus
Eukarya
Frequency Distribution
6
25
Absent
Present
Under the Biological Regulation
Metabolic Process
Establishment of Localization
Cellular Process
Biological Regulation
Unclassified
Integral Component of Membrane
Neuron Part
Intracellular
Envelope
Endomembrane System
Virus, Tail
Unclassified
Transferase Activity
Hydrolase Activity
Structural Constituent of Ribosome
Ion Binding
Unclassified
a
b
c
Genome Biology,

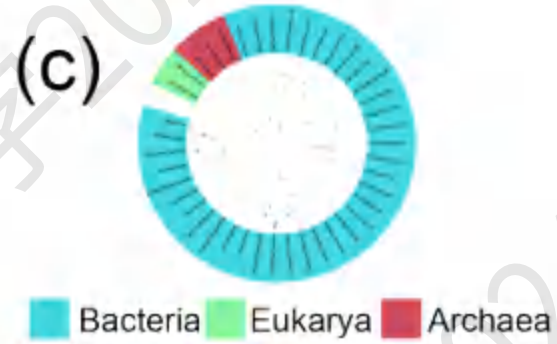


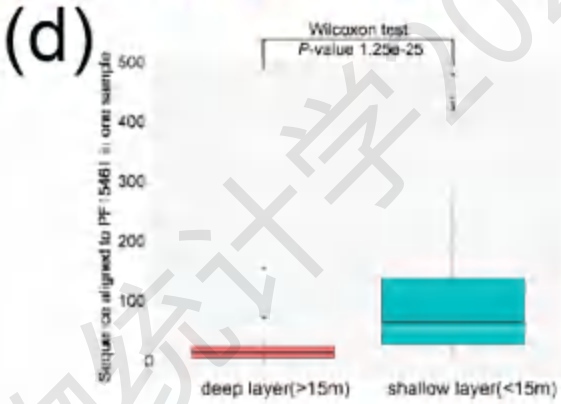


Predicted structure of PF15461

318 Amino Acid
Beta-carotene 15,15'-dioxygenase
369 sequence→14,353 sequence

Predicted Function:
Cellular Component: Respiratory Chain
Biological Process:
Single-organism Metabolic Process
Molecular Function:Oxidoreductase Activity





Genome Biology,

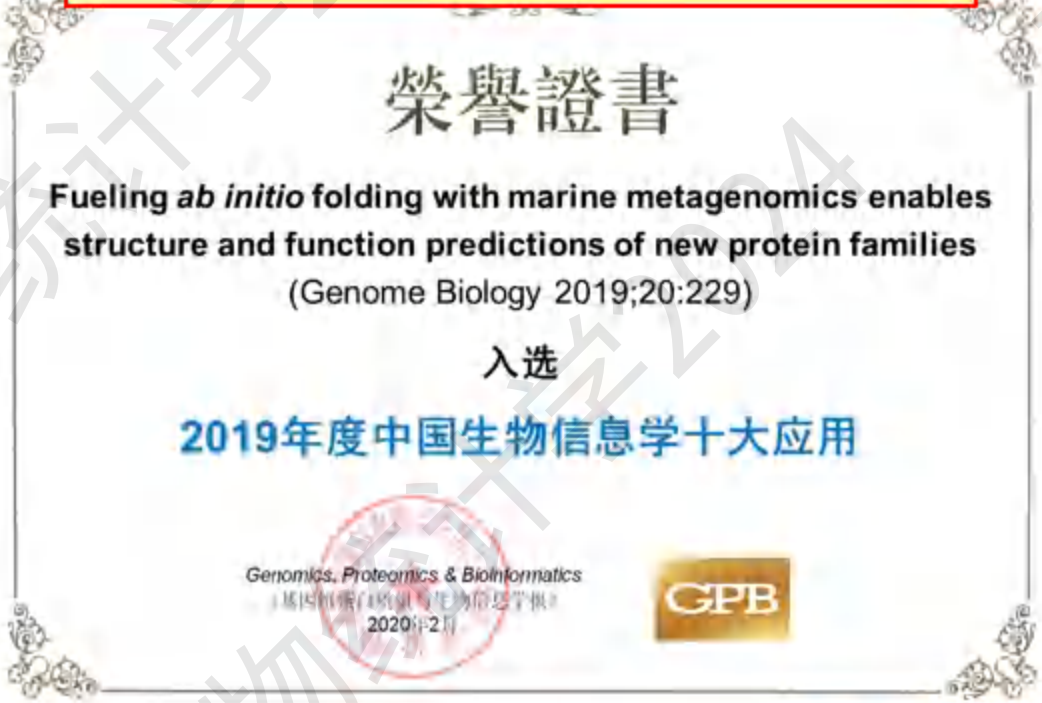
生物统计学2024

微生物组大数据 + 蛋白质3D结构

潜在应用途径

- 非定向/定向方法可以发掘大量未知功能基因
- 为合成生物学提供功能模块
- 药物发掘和药物设计
  - 环境菌群功能基因
  - 环境菌群代谢物

2019年度中国在生物信息学十大应用



荣譽證書

Fueling *ab initio* folding with marine metagenomics enables structure and function predictions of new protein families

(Genome Biology 2019;20:229)

入选

2019年度中国生物信息学十大应用

Genomics, Proteomics & Bioinformatics

2020年2月

GPB

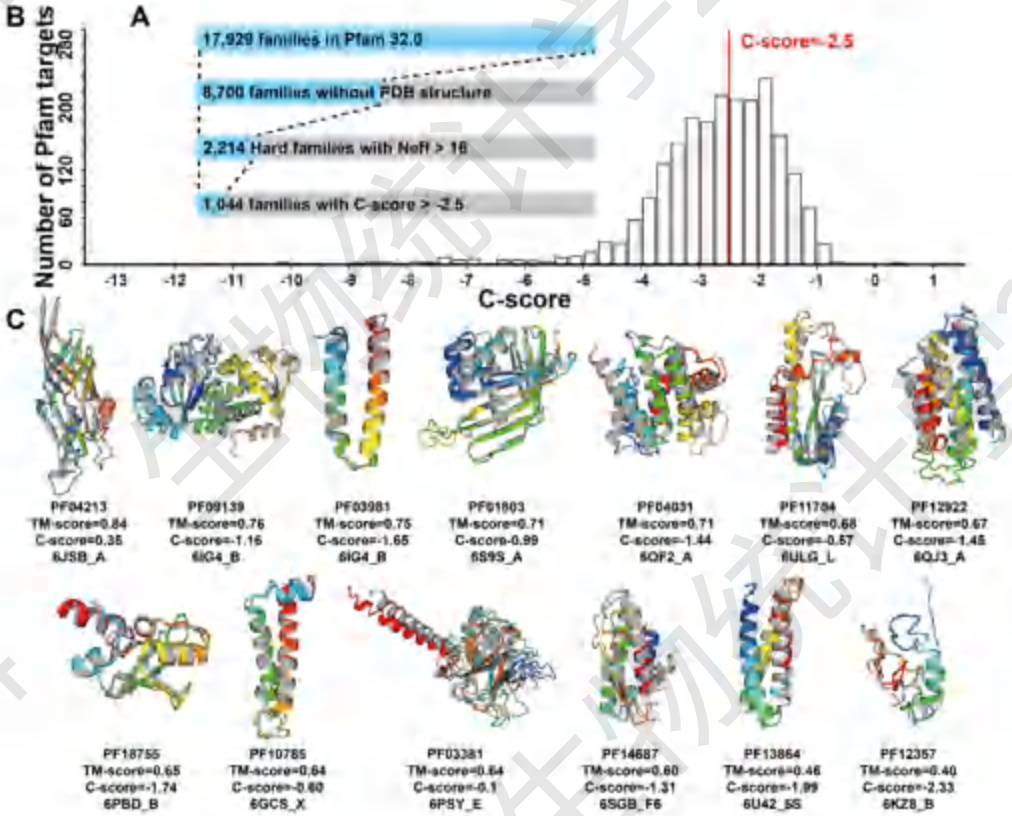
● 宁康团队

Genome Biology,

微生物组大数据 + 蛋白质3D结构

基于大数据挖掘的蛋白质结构预测和功能解析

- 42.5亿微生物组来源序列大数据
- 预测出了之前没有任何结构信息的1,044个蛋白质结构
- 利用人工神经网络挖掘蛋白结构和生境以及进化的关系



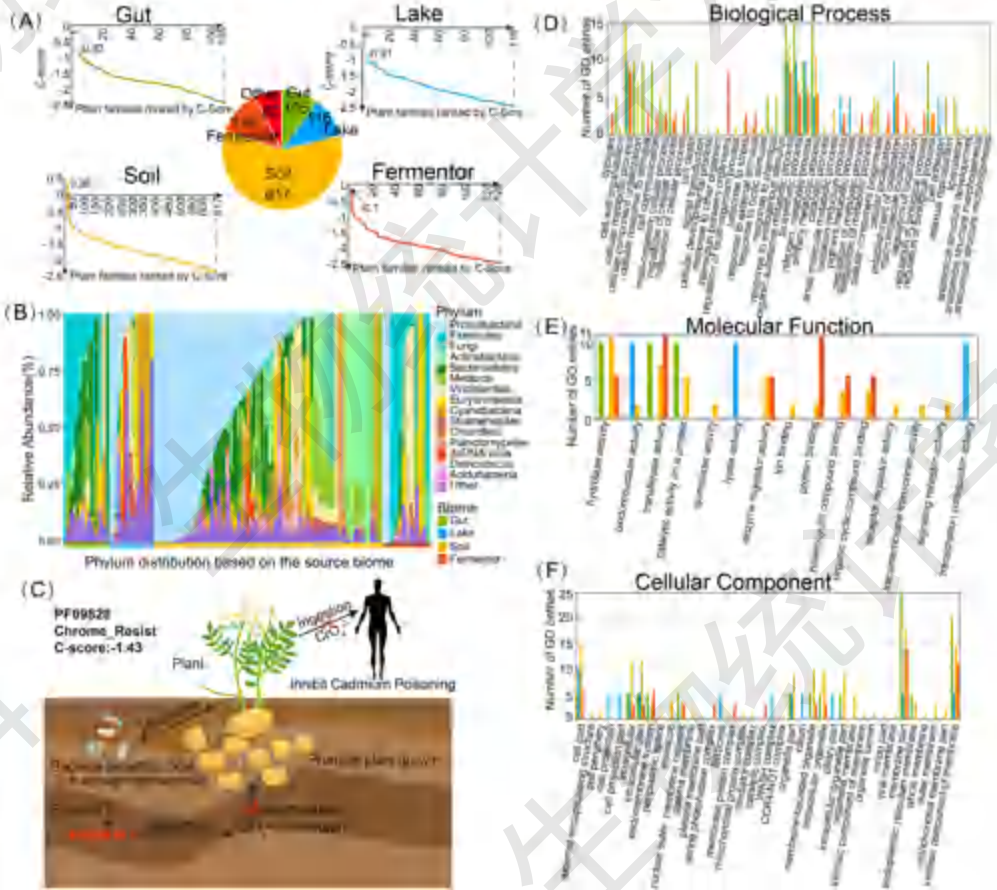
● 宁康团队

PNAS,

微生物组大数据 + 蛋白质3D结构

基于大数据挖掘的蛋白质结构预测和功能解析

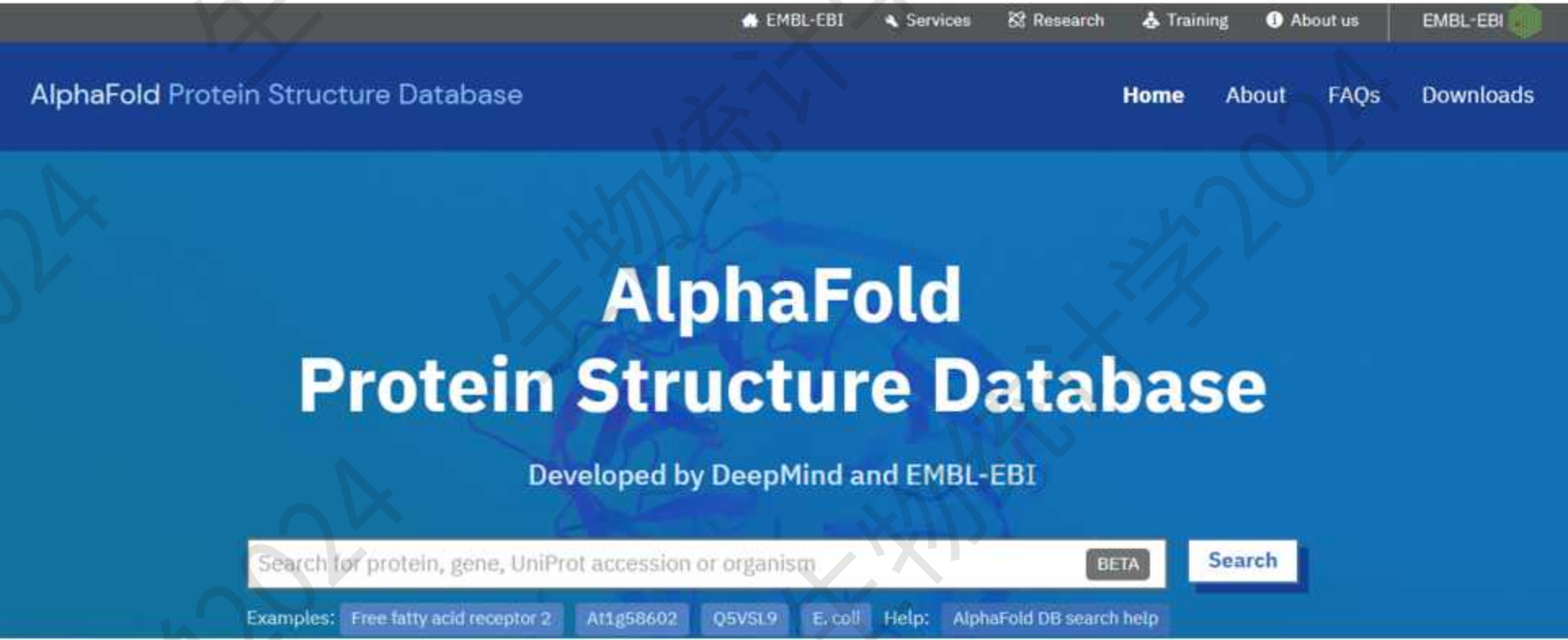
- 42.5亿微生物组来源序列大数据
- 预测出了之前没有任何结构信息的1,044个蛋白质结构
- 利用人工神经网络挖掘蛋白结构和生境以及进化的关系



● 宁康团队

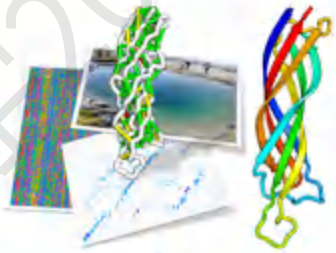
PNAS,

AI + 蛋白质3D结构



生物统计学2024

AI + 蛋白质3D结构



Protein sequences from:

- Public databases (~100M)
- Metagenome (>1,000M)

Google DeepMind

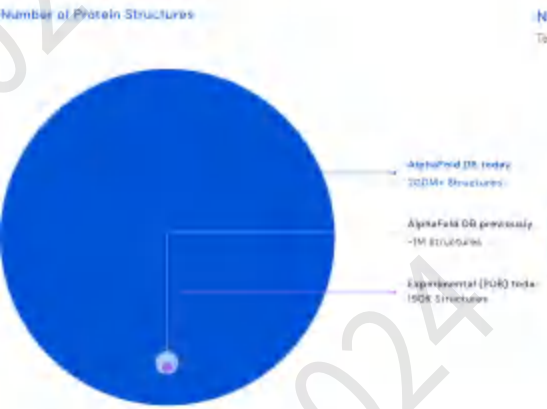
Meta

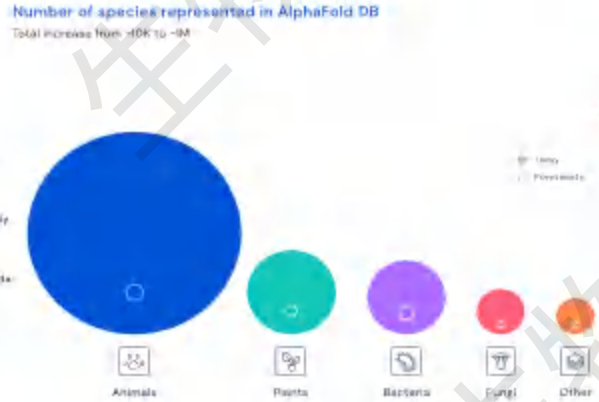
Tencent 腾讯

AlphaFold2
200M+

ESMFold
600M+

tFold
“SRD5A2”





nature communications

Explore our content ⌄ Journal information ⌄

nature > nature communications > articles > article

Article | Open Access | Published: 27 October 2020

Structure of human steroid 5α-reductase 2 with the anti-androgen drug finasteride

Qingpin Xiao, Lei Wang, Shreyas Supekar, Tao Shen, Heng Liu, Fei Ye, Junzhou Huang, Hao Fan, Zhiyi Wei & Cheng Zhang

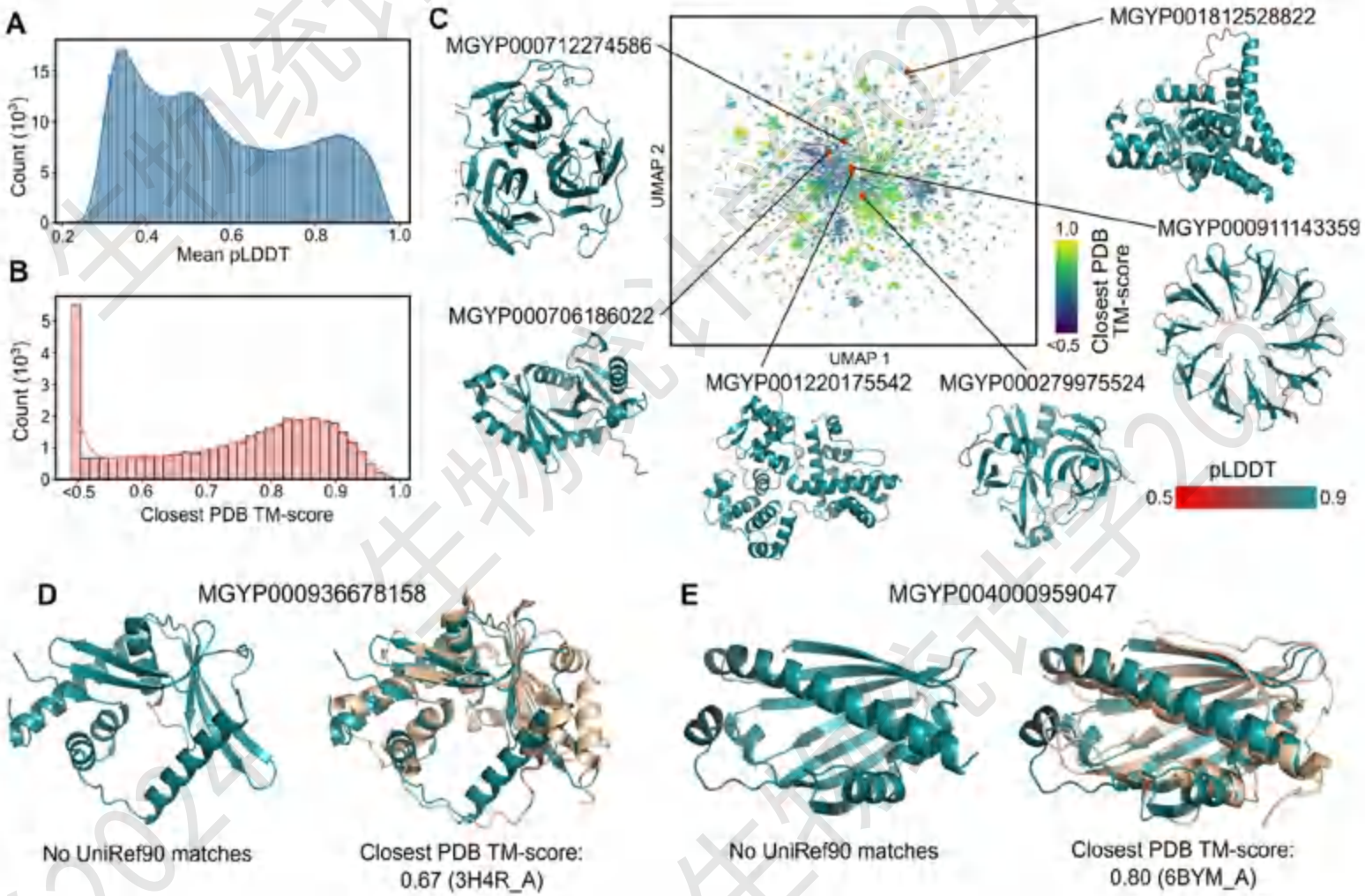
Nature Communications 11, Article number: 5430 (2020) | Cite this article

生物统计学2024

AI + 蛋白质3D结构

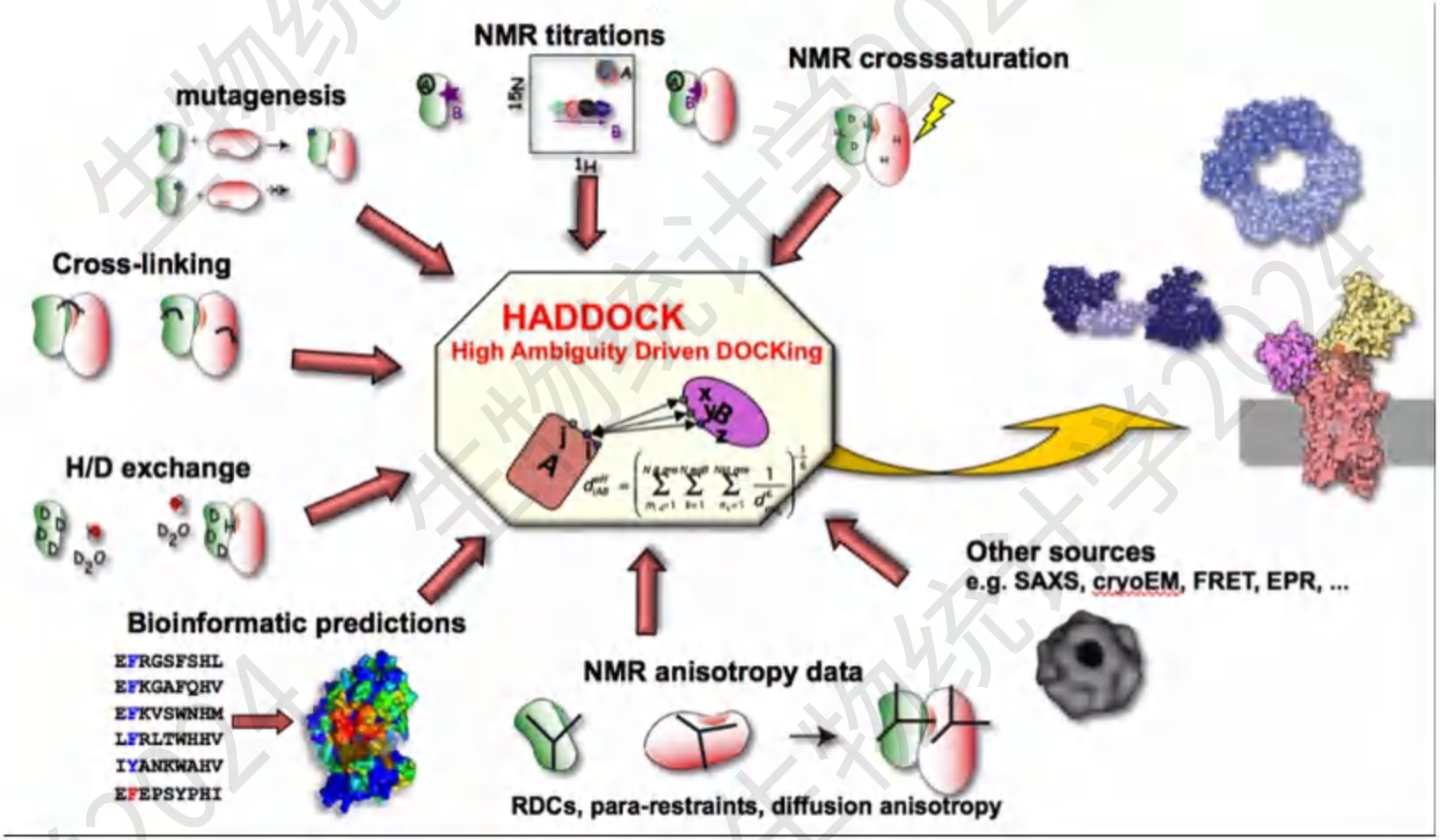
Meta

ESMFold
600M+

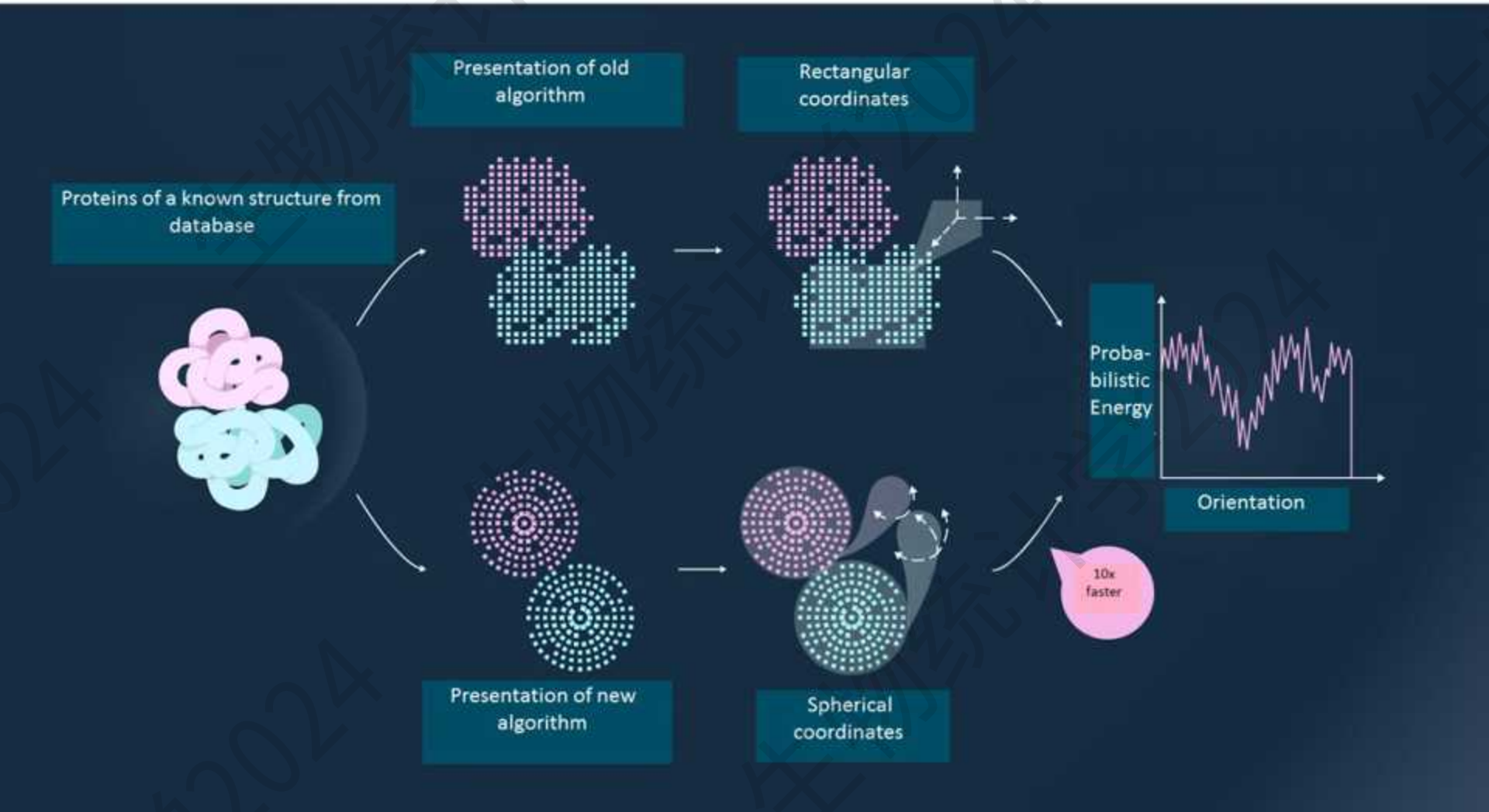


CAPRI

(Critical Assessment of PRediction of Interactions)



CAPRI



生物统计学2024

Protein-portein interaction by 3D modeling

- 通过模拟的方式预测蛋白的结合方式
  - 晶体结构（PDB）？
  - 同源建模（已有模板）？
  - 蛋白质结构预测 -- 蛋白对接 -- 蛋白复合物预测（Pymol）

生物统计学2024

Protein-portein interaction by 3D modeling

Science

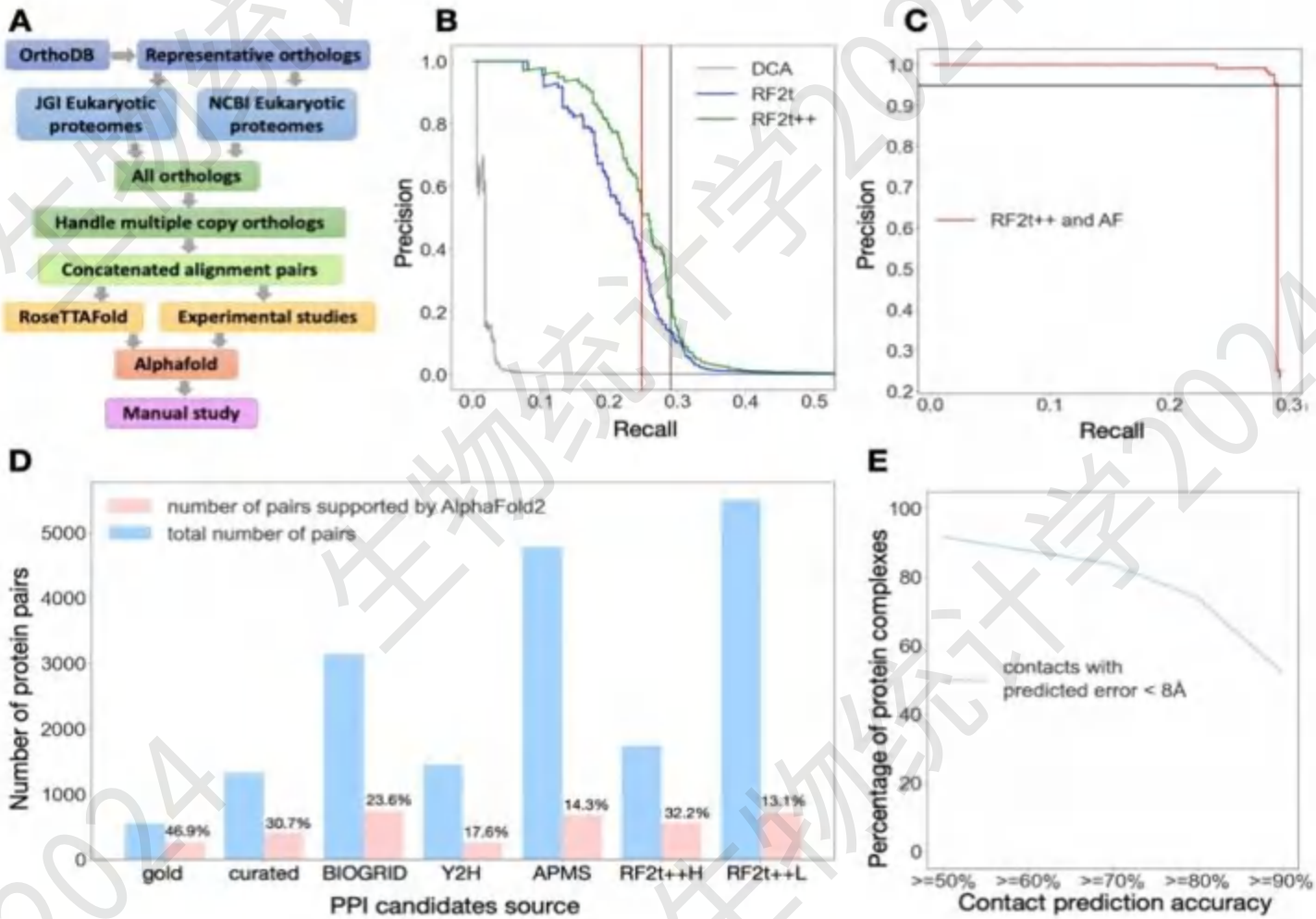
RESEARCH ARTICLES

Cite as: I. R. Humphreys *et al.*, *Science*
10.1126/science.abm4805

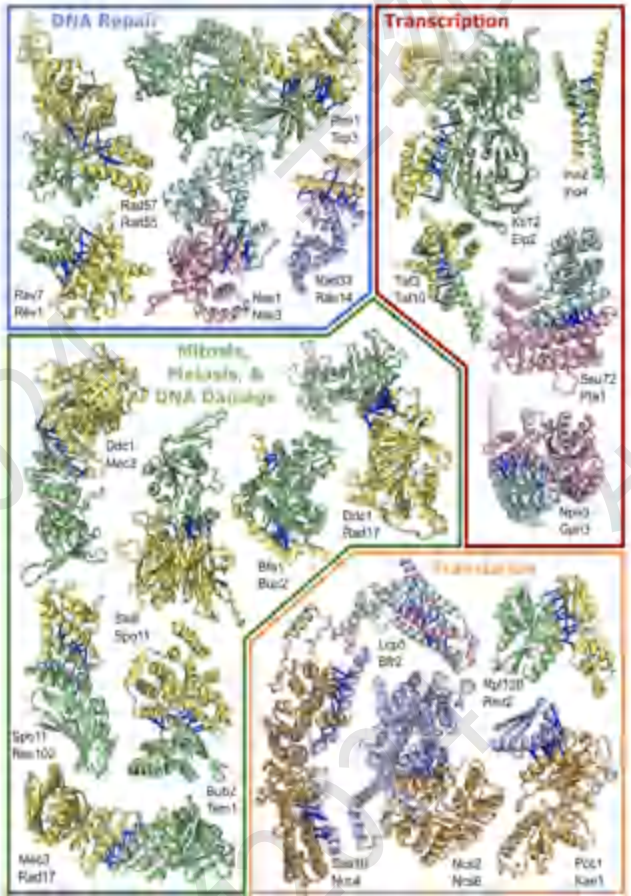
Computed structures of core eukaryotic protein complexes

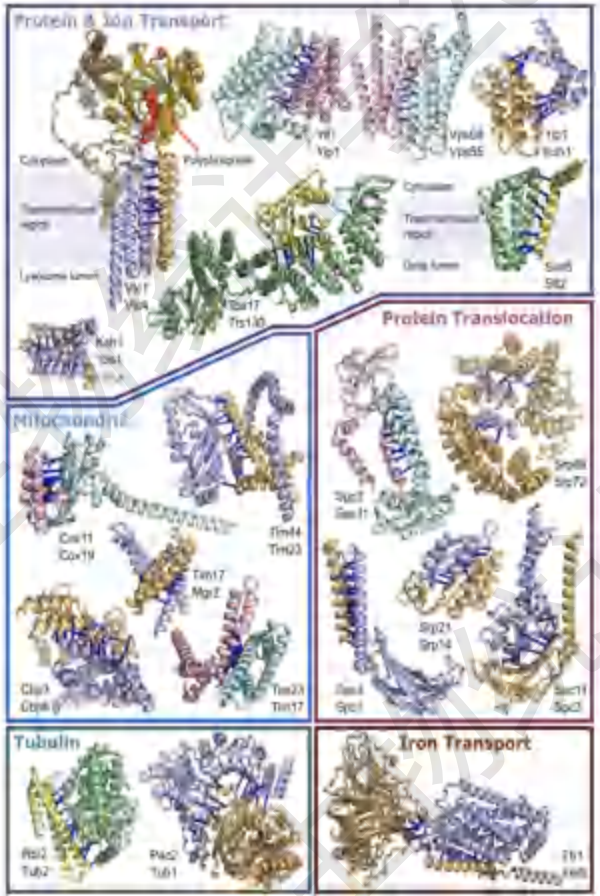
Ian R. Humphreys[1,2]†, Jimin Pei[3,4]†, Minkyung Baek[1,2]†, Aditya Krishnakumar[1,2]†, Ivan Anishchenko[1,2], Sergey Ovchinnikov[5,6], Jing Zhang[3,4], Travis J. Ness[7]‡, Sudeep Banjade[8], Saket R. Bagde[8], Viktoriya G. Stancheva[9], Xiao-Han Li[9], Kaixian Liu[10], Zhi Zheng[10,11], Daniel J. Barrero[12], Upasana Roy[13], Jochen Kuper[14], Israel S. Fernández[15], Barnabas Szakal[16], Dana Branzei[16,17], Josep Rizo[4,18,19], Caroline Kisker[14], Eric C. Greene[13], Sue Biggins[12], Scott Keeney[10,11,20], Elizabeth A. Miller[9], J. Christopher Fromme[8], Tamara L. Hendrickson[7], Qian Cong[3,4]§*, David Baker[1,2,21]§*

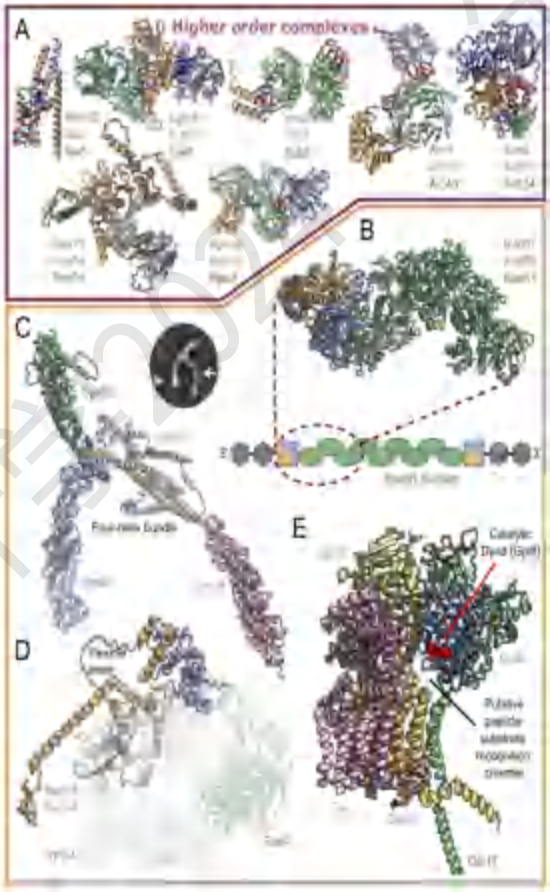
Protein-portein interaction by 3D modeling



Protein-portein interaction by 3D modeling

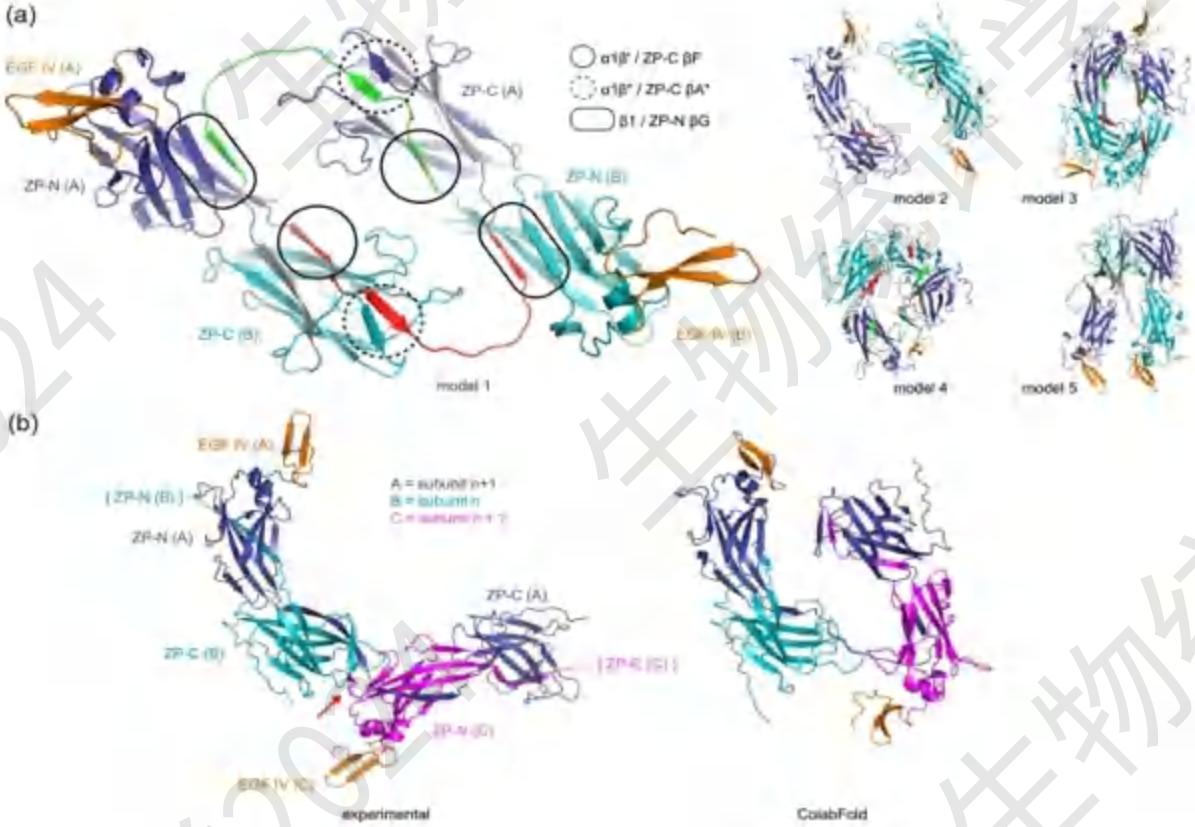






Protein-portein interaction by 3D modeling

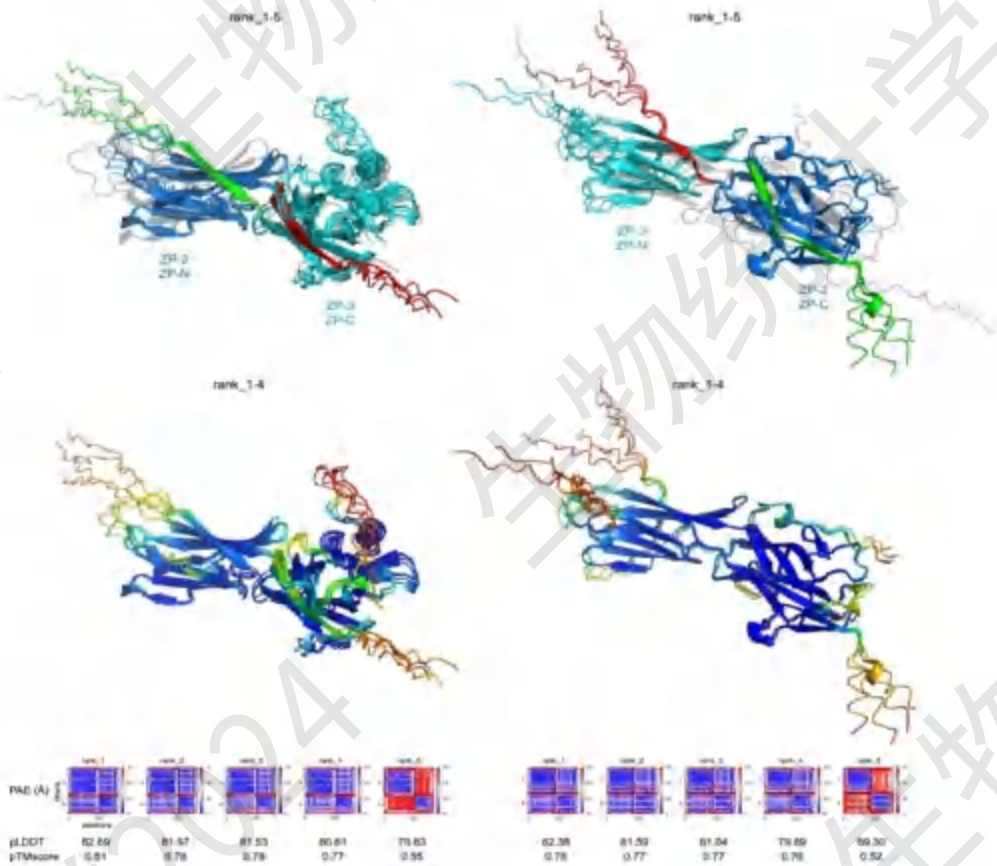
- ColabFold



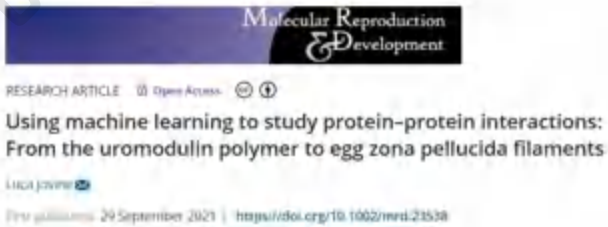
多亚基系统的 ColabFold 建模再现了尿调制素（UMOD）细丝中的透明带（ZP）-N/ZP-C 接头产生的三个主要相互作用。

Protein-portein interaction by 3D modeling

- ColabFold



ZP2/ZP3 蛋白复合物的 ColabFold 建模，由 ZP-N 结构域加一个亚基的域间接头和接头加另一个亚基的 ZP-C 结构域组成。



生物统计学2024

Protein design by AI

RESEARCH ARTICLE

Hallucinating symmetric protein assemblies

B. I. M. WICKY, L. F. MILLES, A. COURBET, R. J. RAGOTTE, J. DAUPARAS, E. KINFU, S. TIPPS, R. D. KIBLER, M. BAEK, F. DIMAIO, X. LI, L. CARTER, A. KANG, H. NGUYEN, A. K. BERA, AND D. BAKER fewer Authors Info & Affiliations

SCIENCE • 15 Sep 2022 • First Release • DOI: 10.1126/science.add1964

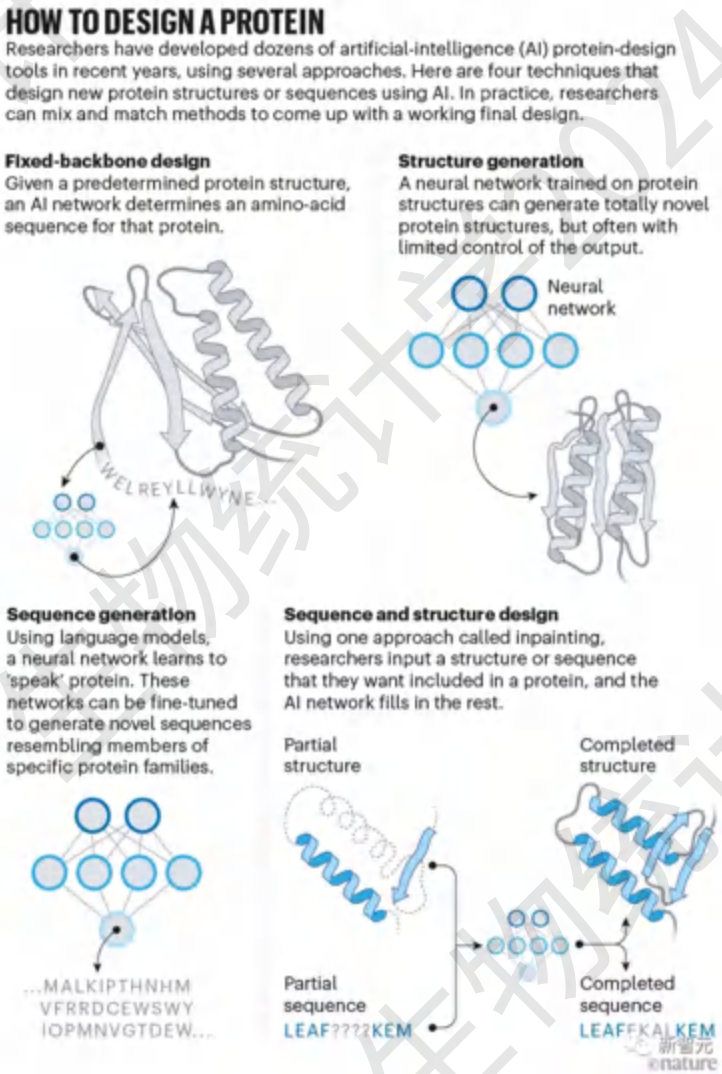
RESEARCH ARTICLE

Robust deep learning–based protein sequence design using ProteinMPNN

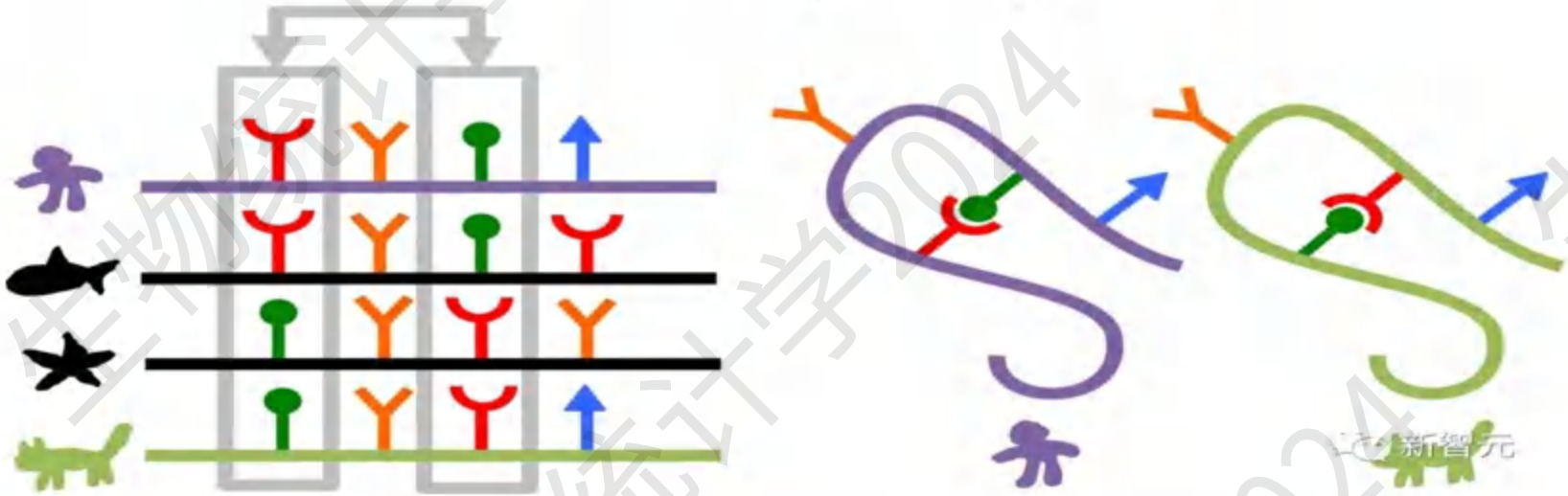
J. DAUPARAS, I. ANISHCHENKO, N. BENNETT, H. BAI, R. J. RAGOTTE, L. F. MILLES, B. I. M. WICKY, A. COURBET, R. J. DE HAAS, N. BETHEL, P. J. Y. LEUNG, T. F. HUDDY, S. PELLOCK, D. TISCHER, F. CHAN, B. KOEPNICK, H. NGUYEN, A. KANG, B. SANKARANARAYANAN, A. K. BERA, N. P. KING, AND D. BAKER fewer Authors Info & Affiliations

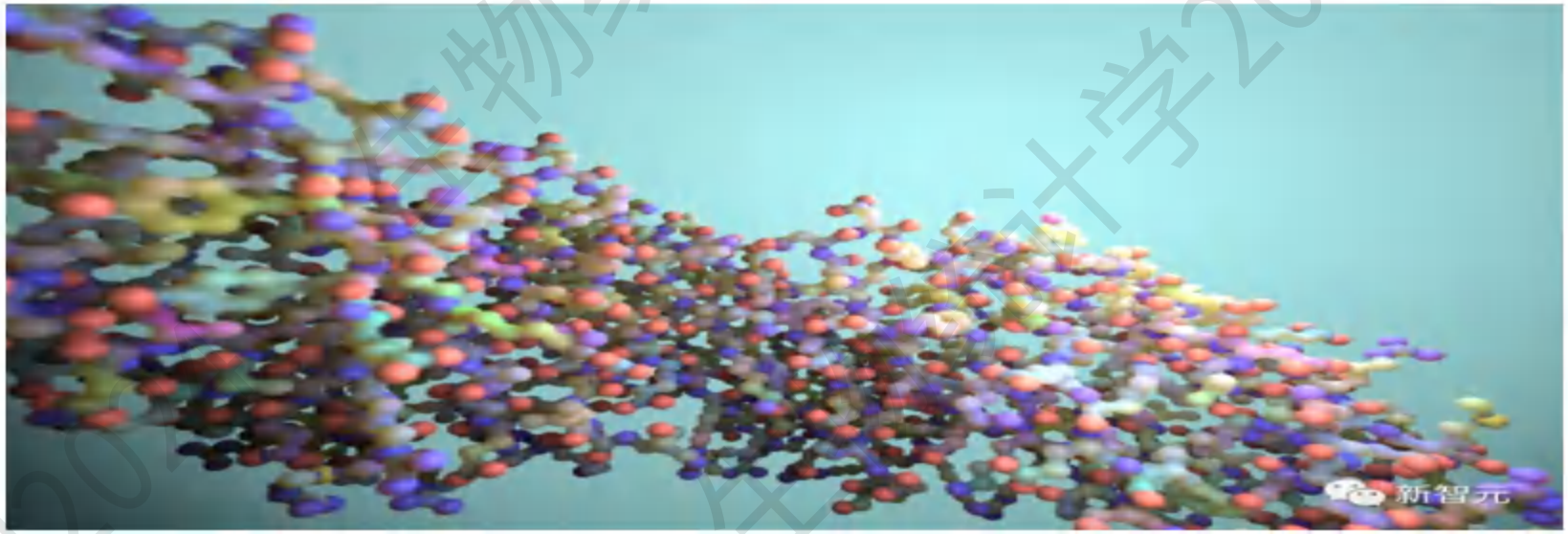
生物统计学2024

Protein design by AI



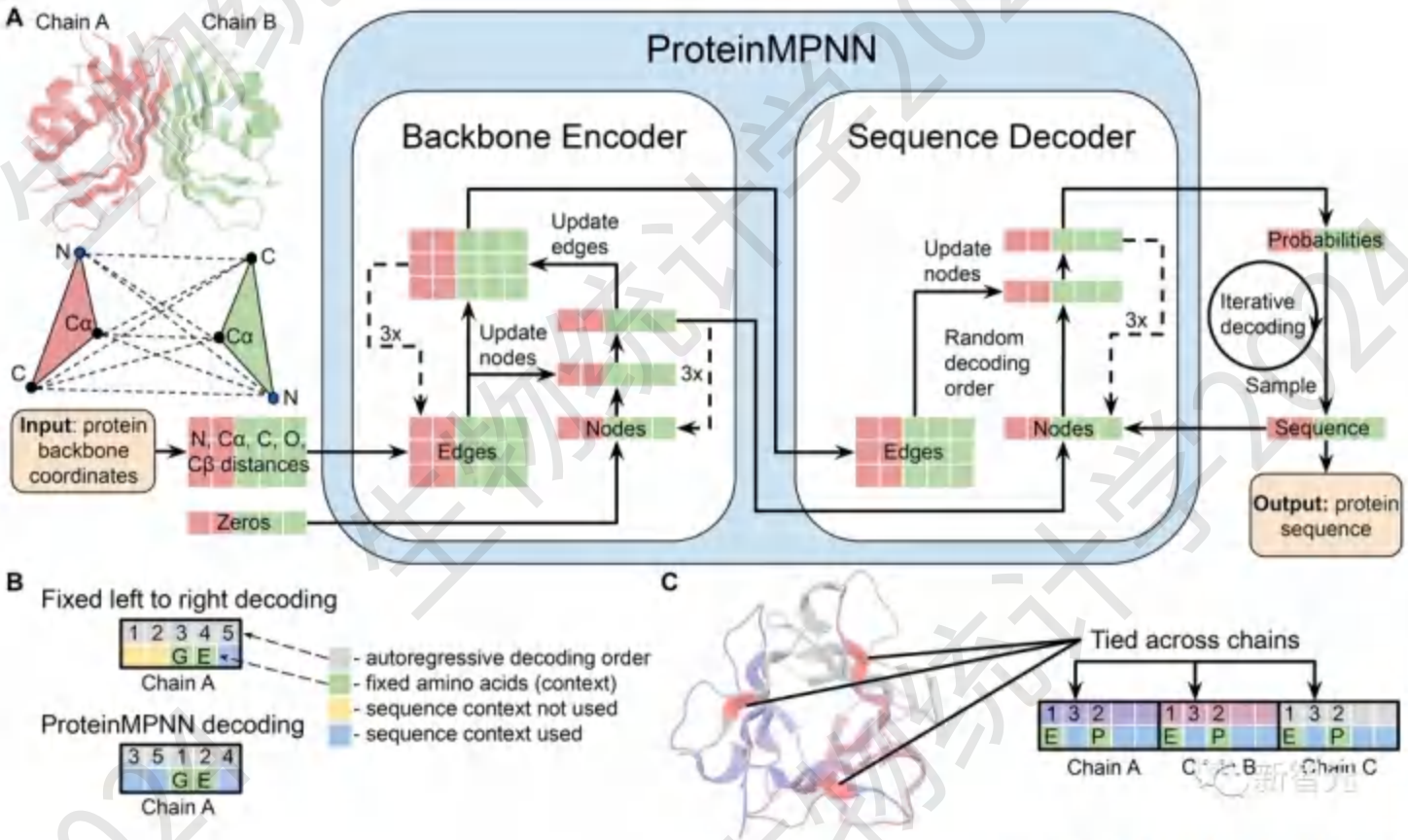
Protein design by AI





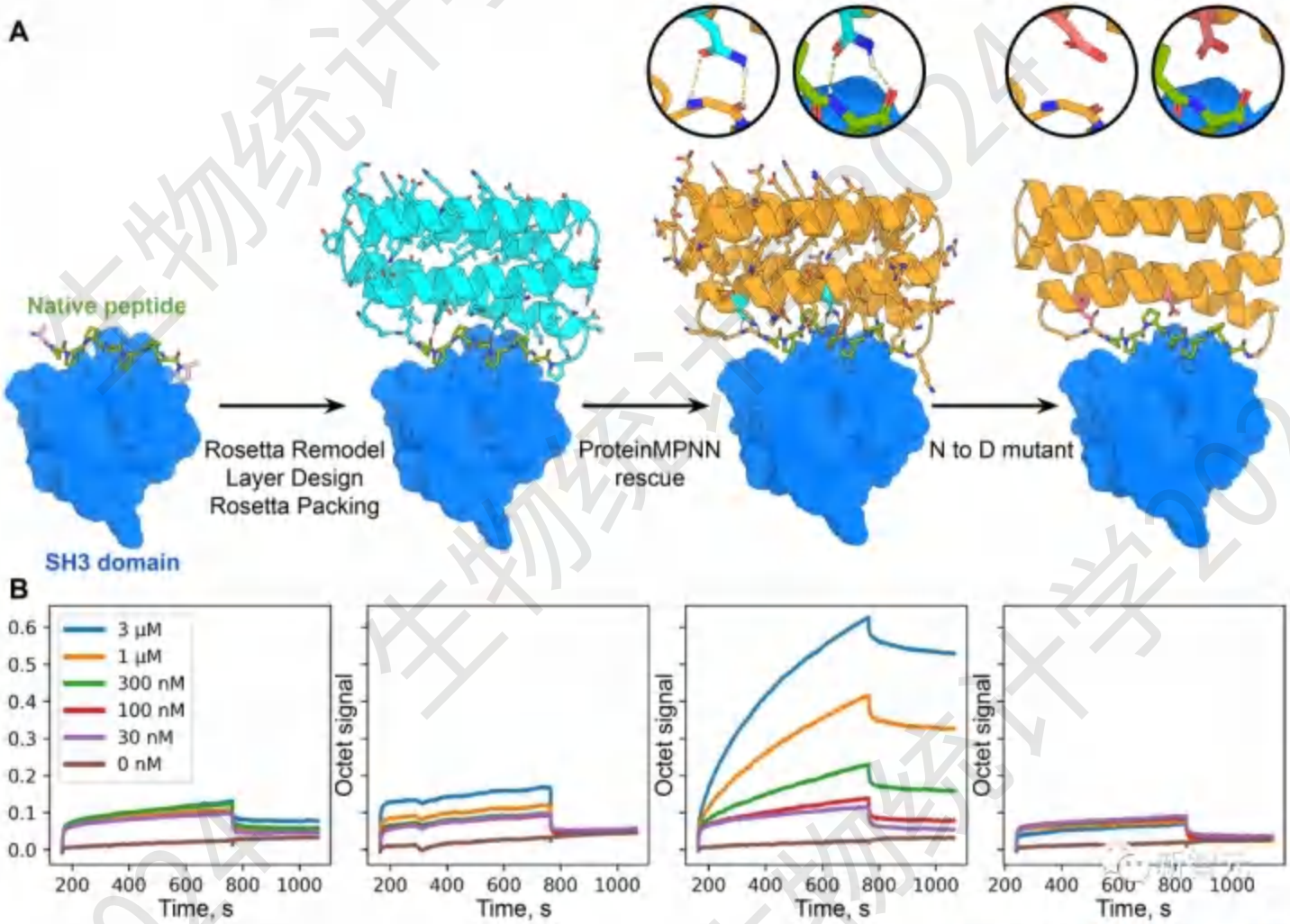
生物统计学2024

Protein design by AI



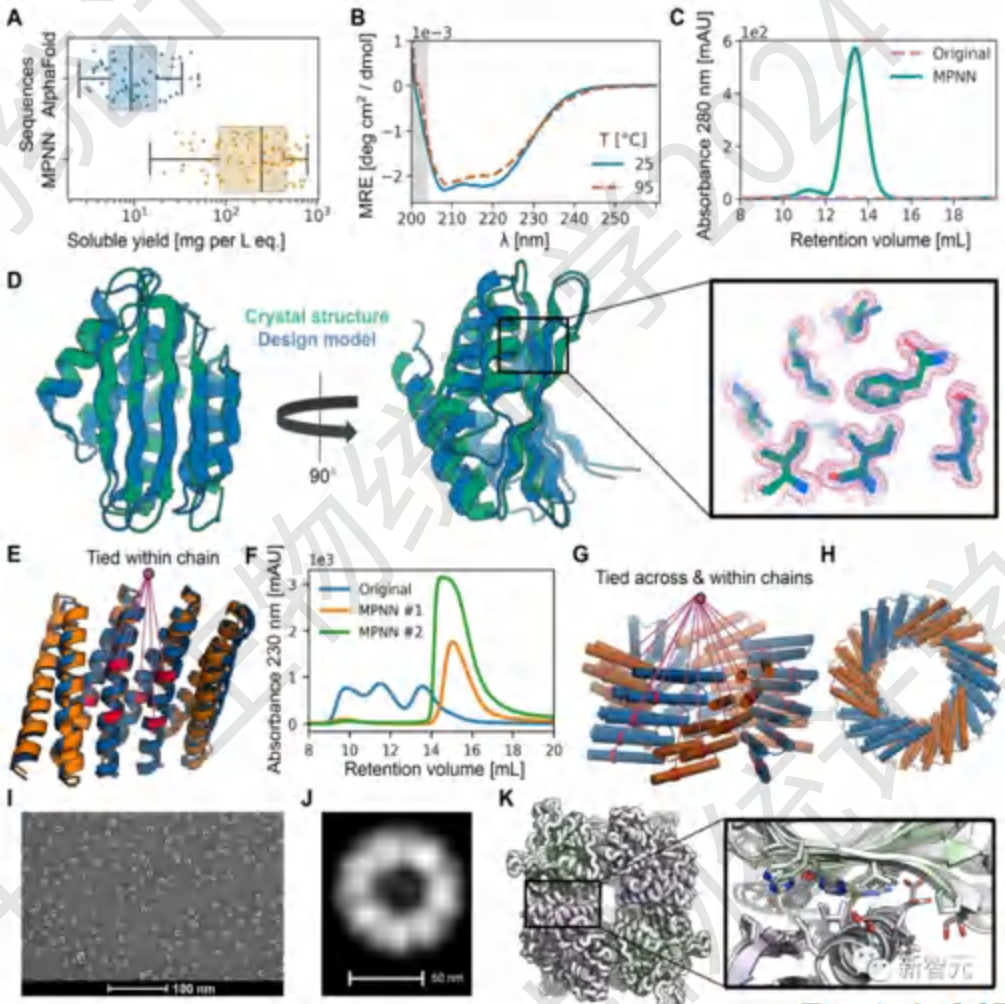
ProteinMPNN, Science, 2022

Protein design by AI



ProteinMPNN, Science, 2022

Protein design by AI

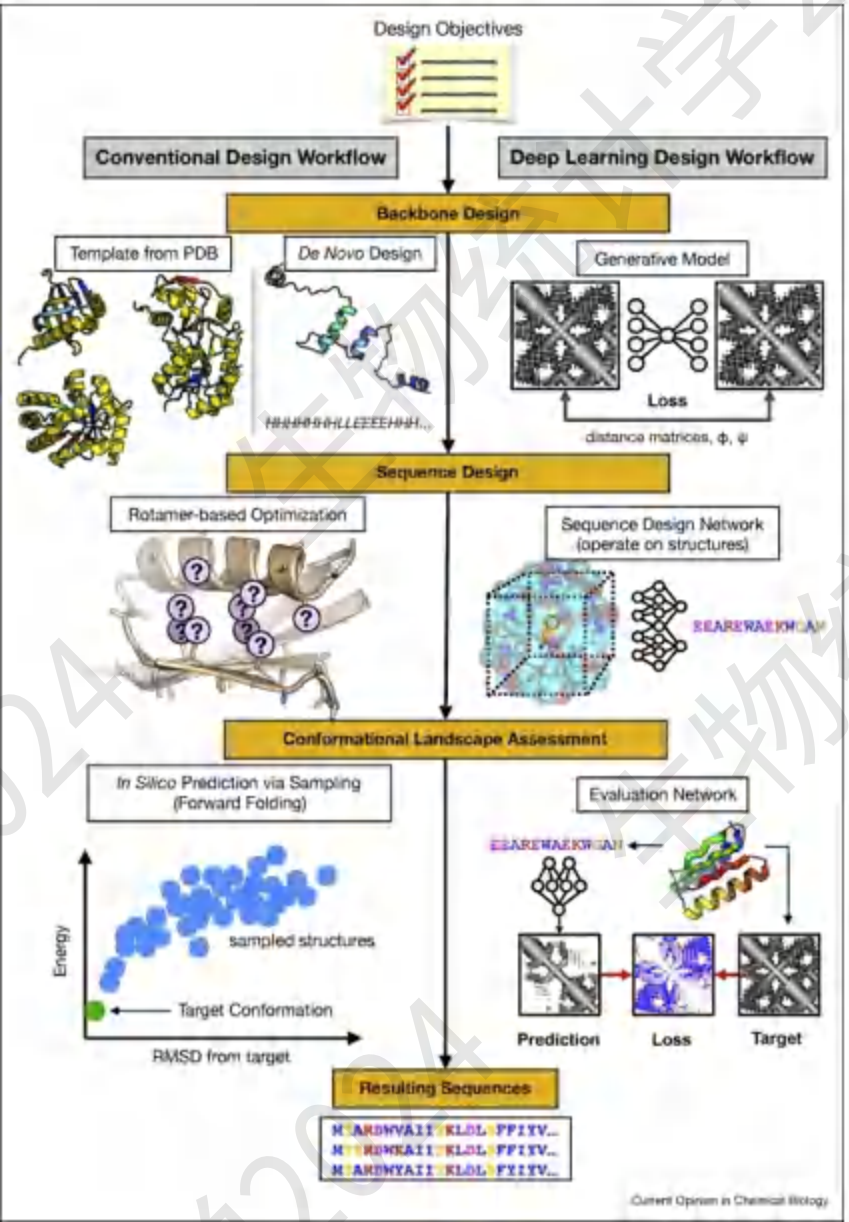


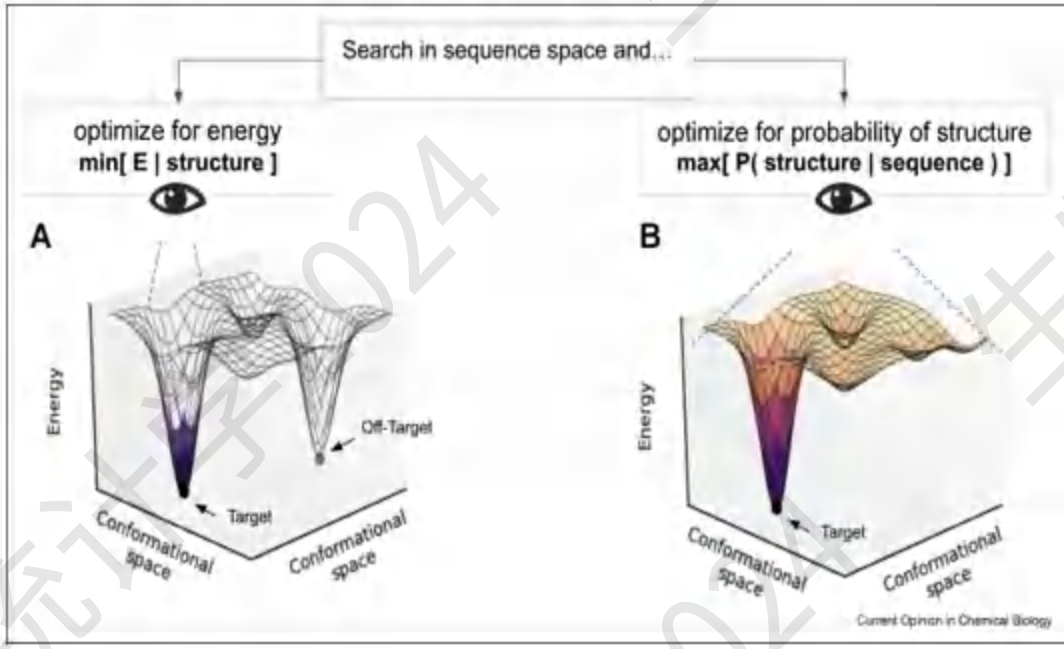
ProteinMPNN, Science, 2022

Protein design by AI

生物统计学2024

Protein 3D structure

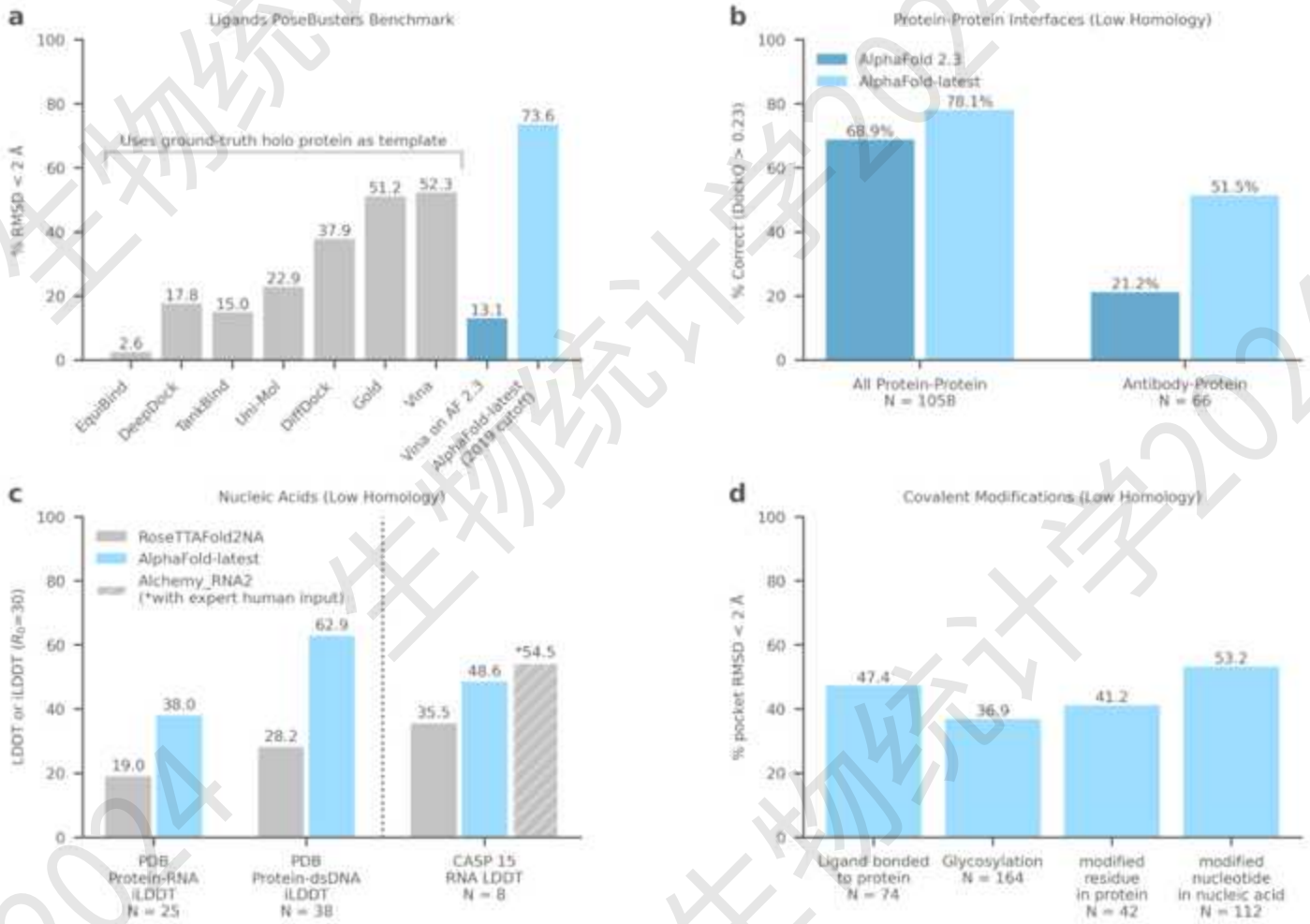




传统设计方法vs深度学习方法

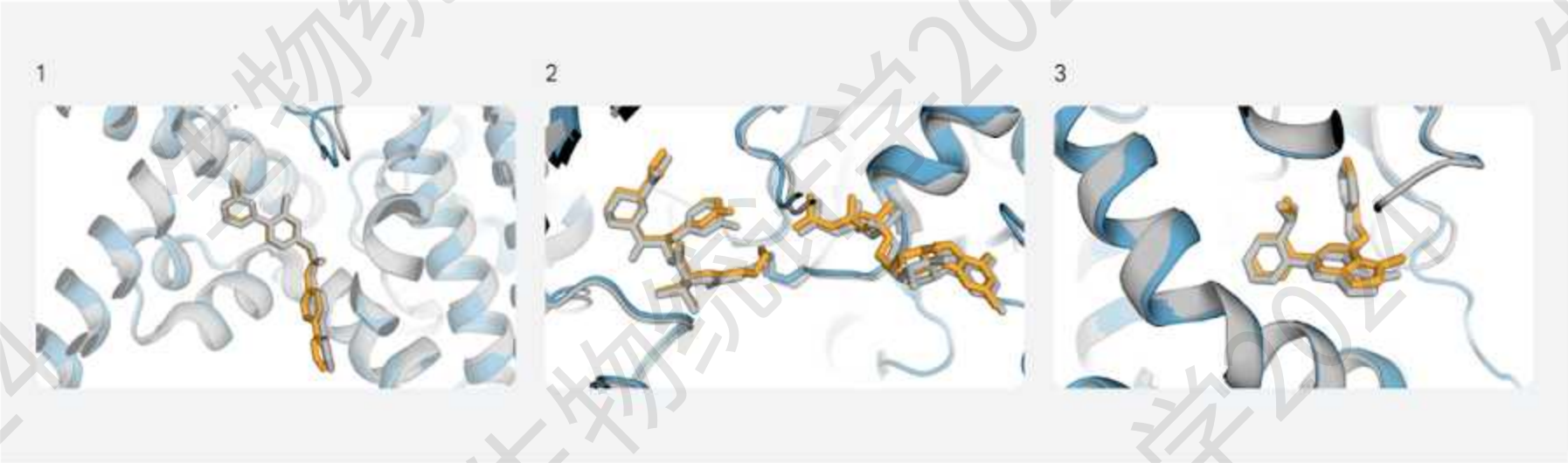
Current Opinion in Chemical Biology, 2022

AlphaFold3



生物统计学2024

AlphaFold3



Predictions for PORCN (1), KRAS (2), and PI5P4Kγ (3)

生物统计学2024

AI + 蛋白质3D结构

David Baker

"for computational protein design"

David Baker. Ill. Niklas Elmehed © Nobel Prize Outreach

Demis Hassabis

"for protein structure prediction"

Demis Hassabis. Ill. Niklas Elmehed © Nobel Prize Outreach

John Jumper

"for protein structure prediction"

John Jumper. Ill. Niklas Elmehed © Nobel Prize Outreach

- 2024年诺贝尔化学奖颁给人工智能模型AlphaFold，和蛋白质结构预测，是对人工智能与科学研究结合的一次里程碑式的认可。
- 今年的诺贝尔化学奖一半授予了David Baker，以表彰他在计算蛋白质设计方面的杰出贡献；另一半则共同授予了Demis Hassabis和John M. Jumper，以表彰他们在蛋白质结构预测方面的成就，特别是他们开发的AlphaFold人工智能模型。

生物统计学2024

AI + 蛋白质3D结构

nature

Article | Published: 08 May 2024

Accurate structure prediction of biomolecular interactions with AlphaFold 3

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg ✉, Demis Hassabis ✉ & John M. Jumper ✉ — Show fewer authors

生物统计学2024

AI + 蛋白质3D结构

alphafold3 Public

Watch 26 | Fork 121 | Star 1.4k

main | 1 Branch | 1 Tag | Go to file | Add file | Code

Augustin-Zidek Make fetch_databases compatible with Python 3.6 e2afc41 · 7 hours ago 4 Commits

docker	Initial release of AlphaFold 3	14 hours ago
docs	Small documentation fixes	10 hours ago
src/alphafold3	Initial release of AlphaFold 3	14 hours ago
CMakeLists.txt	Initial release of AlphaFold 3	14 hours ago
LICENSE	Initial release of AlphaFold 3	14 hours ago
OUTPUT_TERMS_OF_USE.md	Initial release of AlphaFold 3	14 hours ago
README.md	Remove incorrect trailing comma in README	7 hours ago
WEIGHTS_PROHIBITED_USE_POLICY.md	Initial release of AlphaFold 3	14 hours ago
WEIGHTS_TERMS_OF_USE.md	Initial release of AlphaFold 3	14 hours ago
dev-requirements.txt	Initial release of AlphaFold 3	14 hours ago
fetch_databases.py	Make fetch_databases compatible with Python 3.6	7 hours ago
pyproject.toml	Initial release of AlphaFold 3	14 hours ago
requirements.txt	Initial release of AlphaFold 3	14 hours ago
run_alphafold.py	Initial release of AlphaFold 3	14 hours ago
run_alphafold_test.py	Initial release of AlphaFold 3	14 hours ago

About

AlphaFold 3 inference pipeline.

Readme
View license
Activity
Custom properties
1.4k stars
26 watching
121 forks
Report repository

Releases 1

AlphaFold v3.0.0 Latest
14 hours ago

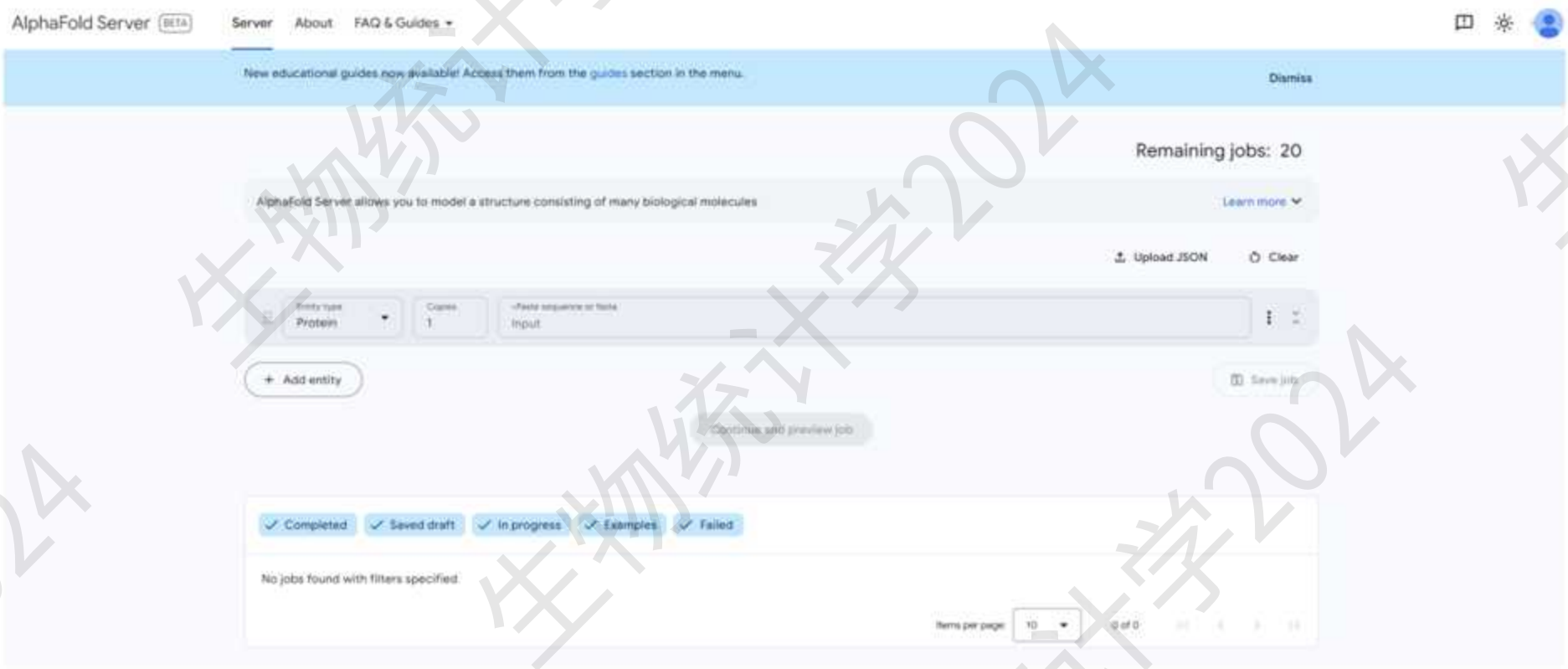
Languages

Python 85.8% C++ 13.8% Other 0.4%

https://github.com/google-deepmind/alphafold3

生物统计学2024

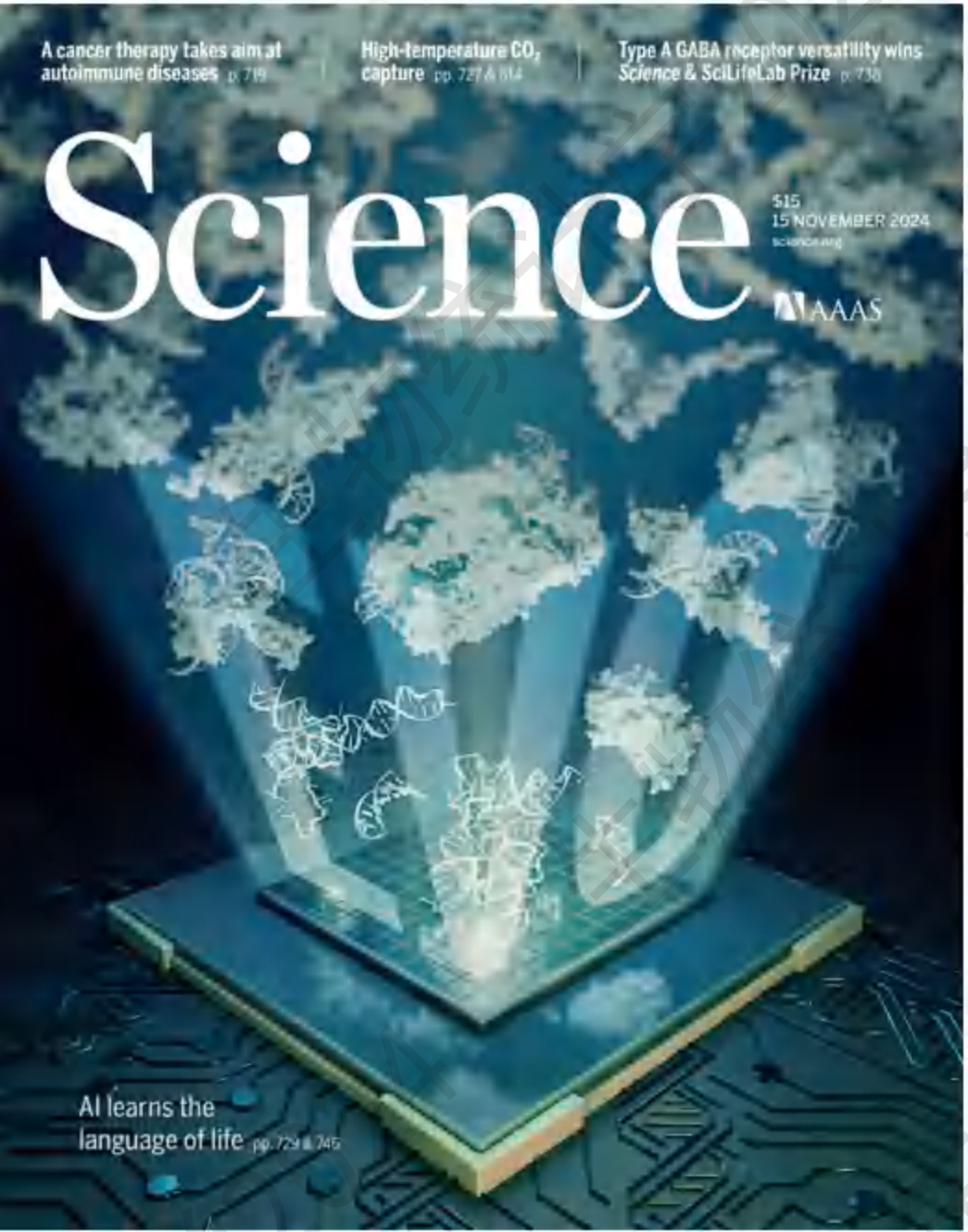
AI + 蛋白质3D结构



https://alphafoldserver.com/

生物统计学2024

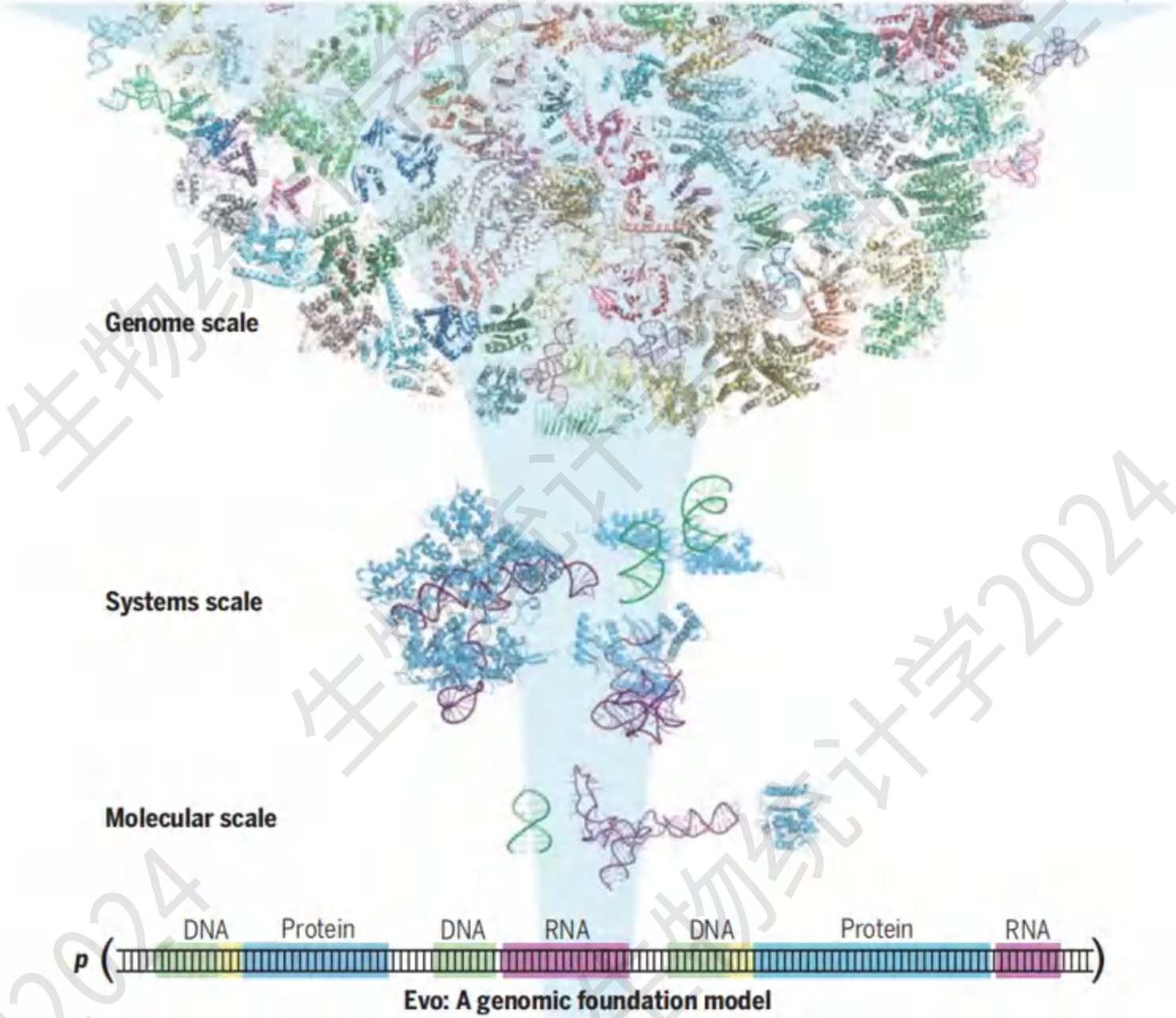
AI + 蛋白质3D结构



蛋白质结构研究与诺贝尔化学奖

年度	获奖人	获奖成果
1958	桑格（F. Sanger）	胰岛素一级结构
1962	佩鲁茨（M. F. Perutz）	球蛋白（血红蛋白和肌红蛋白）三维结构
	肯德鲁（J. C. Kendrew）	
1972	安芬森（C. B. Anfinsen）	一级结构决定高级结构
	摩尔（S. Moore）	核糖核酸酶一级结构
	斯坦（W. H. Stein）	
1982	克卢格（A. Klug）	晶体电子显微镜发展
1988	戴森霍弗（J. Deisenhofer）	光合反应中心的三维结构
	胡贝尔（R. Huber）	
	米歇尔（H. Michel）	
1997	沃克（J. E. Walker）	ATP 酶三维结构
2002	芬恩（J. B. Fenn）	质谱法测蛋白质一级结构
	田中耕一（Koichi Tanaka）	
	维特里希（K. Wüthrich）	磁共振波谱研究蛋白质高级结构
2003	麦金农（R. MacKinnon）	钾离子通道三维结构
2006	科恩伯格（R. D. Kornberg）	真核生物的转录结构
2009	拉马克里斯南（V. Ramakrishnan）	核糖体三维结构
	施泰茨（T. A. Steitz）	
	尤纳斯（A. Yonath）	
2012	科比尔卡（B. Kobilka）	G- 蛋白偶联受体结构
2017	杜博歇（J. Dubochet）	研究蛋白质高级结构的冷冻电镜开发
	弗兰克（J. Frank）	
	亨德森（R. Henderson）	

AI + 蛋白质3D结构



生物统计学2024

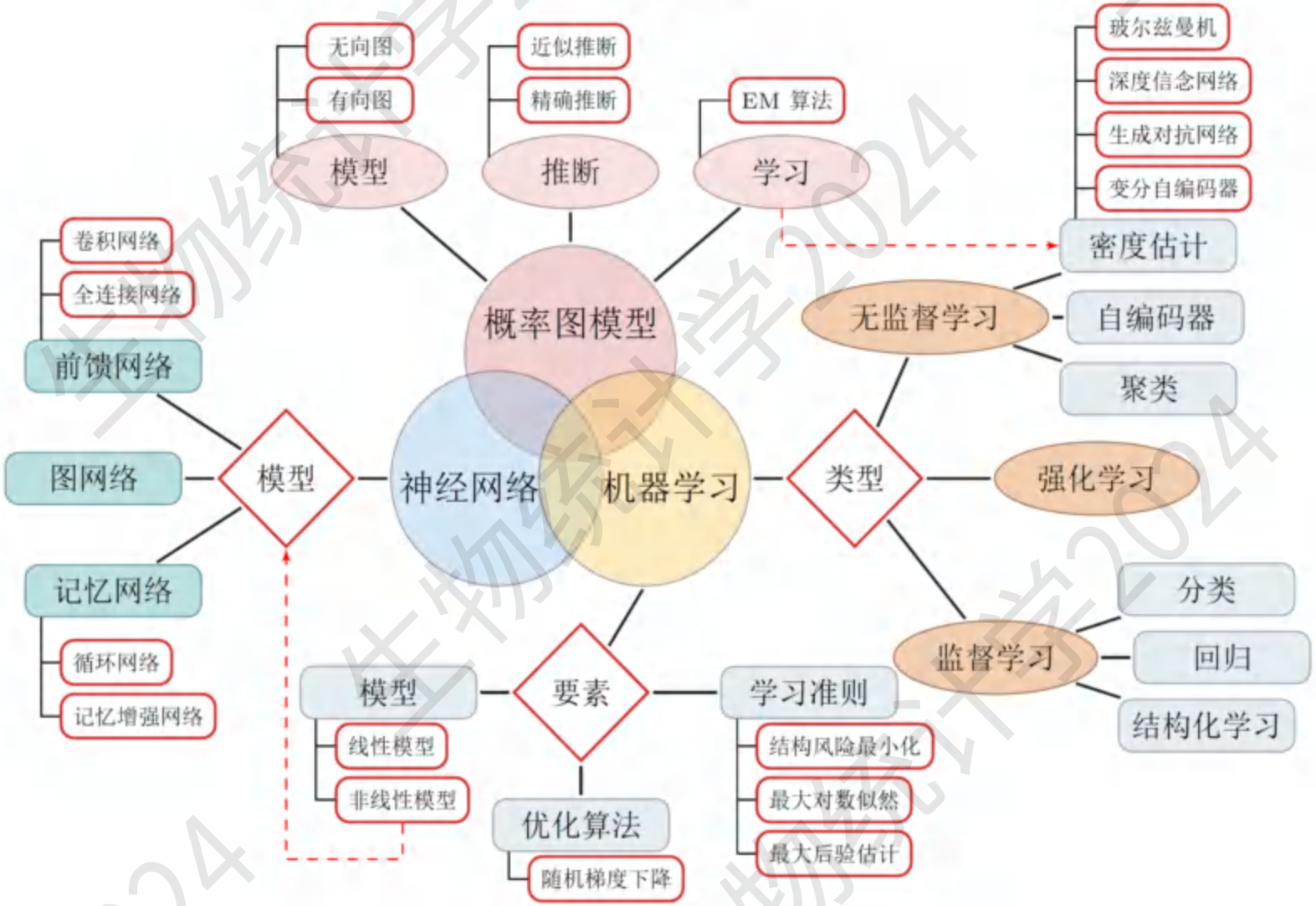
Recap (知识点总结)

- 深度学习：
  - 概念和区别
  - 模型和方法
  - 应用领域

- 生物大数据的深度学习：
  - 序列和特征
  - 特征学习和模型构建
  - 应用案例

生物统计学2024

汇总



生物统计学2024

推荐课程

- 斯坦福大学CS224d: Deep Learning for Natural Language Processing
  - http://cs224d.stanford.edu/
  - Richard Socher 主要讲解自然语言处理领域的各种深度学习模型
- 斯坦福大学CS231n: Convolutional Neural Networks for Visual Recognition
  - http://cs231n.stanford.edu/
  - Fei-Fei Li Andrej Karpathy 主要讲解CNN、RNN在图像领域的应用
- 加州大学伯克利分校 CS 294: Deep Reinforcement Learning
  - http://rail.eecs.berkeley.edu/deeprlcourse/

生物统计学2024

推荐材料

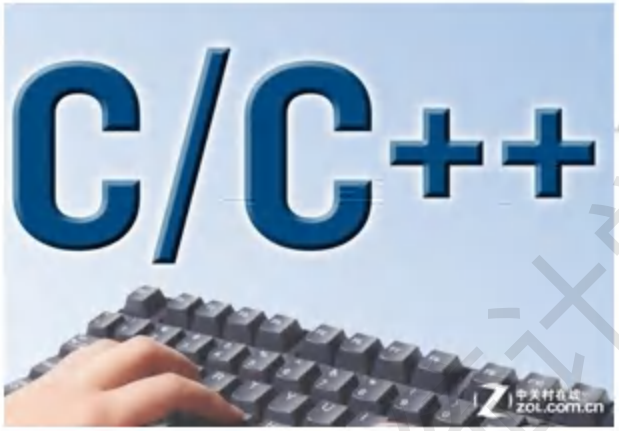
- 林轩田《机器学习基石》《机器学习技法》
  - https://www.csie.ntu.edu.tw/~htlin/mooc/
- 李宏毅《1天搞懂深度学习》
  - http://speech.ee.ntu.edu.tw/~tlkagk/slide/Tutorial_HYLee_Deep.pptx
- 李宏毅《Generative Adversarial Network (GAN)》
  - http://speech.ee.ntu.edu.tw/~tlkagk/slide/Tutorial_HYLee_GAN.pptx

生物统计学2024

生物统计学2024

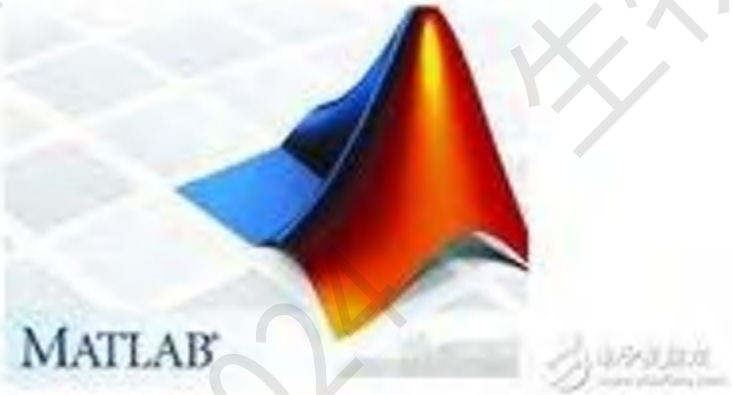
编程语言的选择

生物统计学2024


C/C++
zol.com.cn

Java

python


MATLAB

R
数据科学
网络
科学
学习
大学
信息
研究
会议
计算
方法
语言
问题
可能
预测
中国
公司
时间
选择
变量
系统
关键
影响
应用
学院
data
cos

生物统计学2024

R 与 Python 语言的区别

学习难度大

入门简单

命令式编程

适合处理大量数据

统计功能强大

功能强大

生物统计学2024

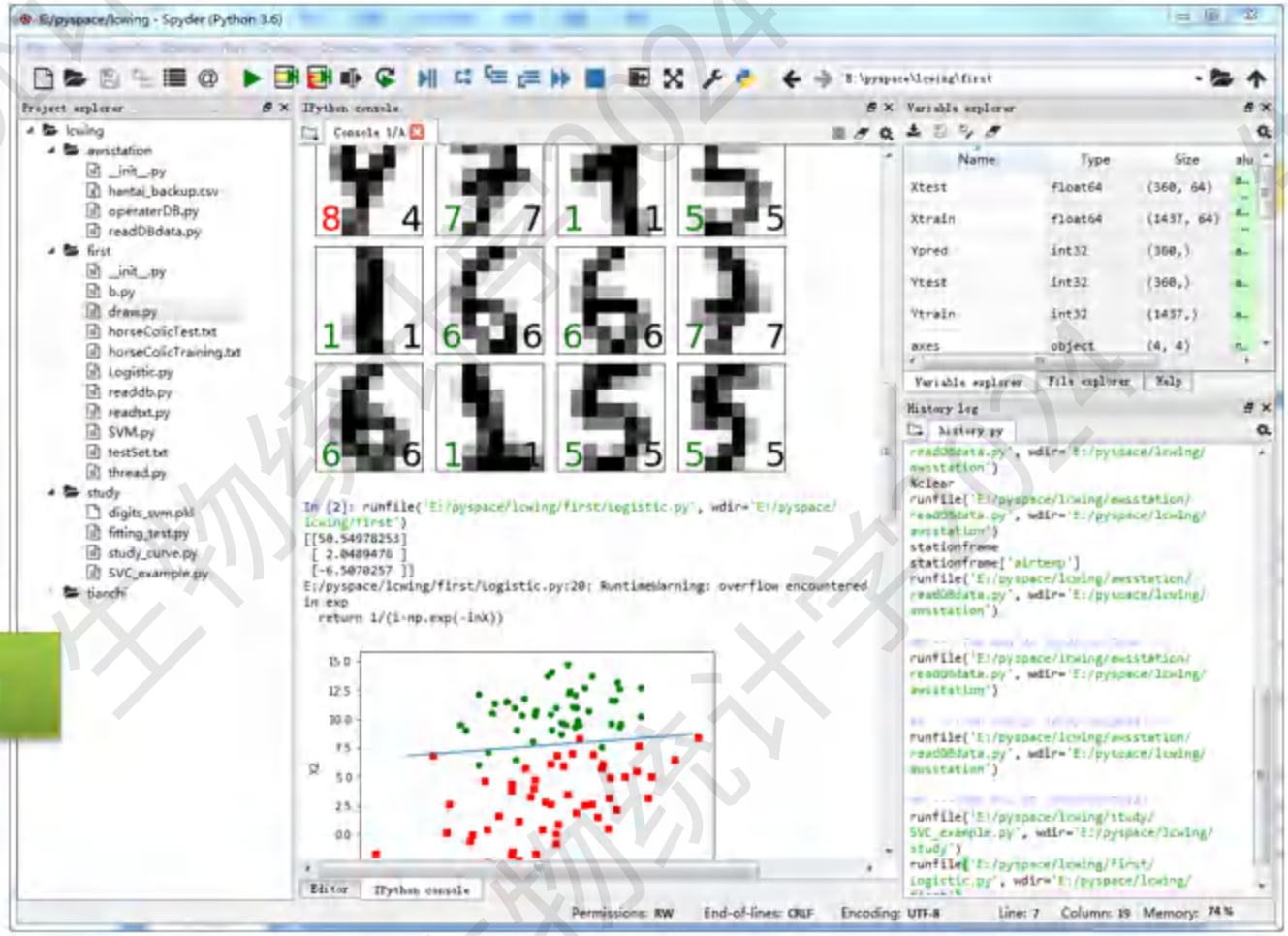
Nov 2018	Nov 2017	Change	Programming Language	Ratings	Change
1	1		Java	16.746%	+3.51%
2	2		C	14.396%	+5.10%
3	3		C++	8.282%	+2.94%
4	4		Python	7.683%	+3.20%
5	7	^	Visual Basic .NET	6.490%	+3.58%
6	5	v	C#	3.952%	+0.94%
7	6	v	JavaScript	2.655%	-0.32%
8	8		PHP	2.376%	+0.48%
9	-	^^	SQL	1.844%	+1.84%
10	14	^^	Go	1.495%	-0.07%
11	19	^^	Objective-C	1.476%	+0.06%
12	20	^^	Swift	1.455%	+0.07%
13	9	vv	Delphi/Object Pascal	1.423%	-0.32%
14	11	v	R	1.407%	-0.20%
15	10	vv	Assembly language	1.108%	-0.61%
16	13	v	Ruby	1.091%	-0.50%
17	12	vv	MATLAB	1.030%	-0.57%
18	15	v	Perl	1.001%	-0.56%
19	18	v	PL/SQL	1.000%	-0.45%
20	17	v	Visual Basic	0.854%	-0.63%

生物统计学2024

Python在线学习推荐----

Anaconda：初学Python、入门机器学习的首选

- NumPy
- Scipy
- Matplotlib
- Pandas
- Scikit-Learn
- TensorFlow



生物统计学2024

References

- Angermueller, C., et al., Deep learning for computational biology. Molecular systems biology, 2016. 12(7): p. 878.
- Jurtz, V.I., et al., An introduction to deep learning on biological sequence data: examples and solutions. Bioinformatics, 2017. 33(22): p. 3685-3690.
- Wang, S., et al., Protein secondary structure prediction using deep convolutional neural fields. Scientific reports, 2016. 6.
- Angermueller, C., et al., DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome biology, 2017. 18(1): p. 67.
- Yuan, L., et al., Applications of Deep Learning in Biological and Medical Data Analysis. PROGRESS IN BIOCHEMISTRY AND BIOPHYSICS, 2016. 43(5): p. 472-483.
- Chai, G., et al., HMMCAS: a web tool for the identification and domain annotations of Cas proteins. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2017.

生物统计学2024

深度学习的问题。。。

?

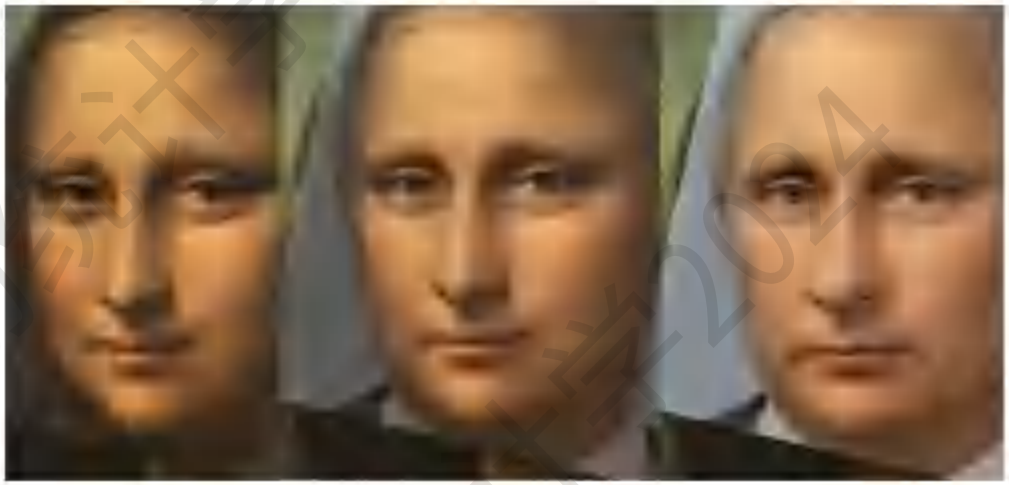
Feature extraction? Or D&C

Case 1

为何对基因的分析结果难以达到共识？“其中一个原因是遗传变异太常见。一个典型的基因包含了数以百计的变异，有些变异可能对健康没有任何影响。有些疾病可能是多个基因变异相互作用的结果，这种作用可能经常被算法所遗漏”，威康信托基金会桑格研究所遗传学家 Matthew Hurles说。 Hurles目前正引领一个关于破解发育障碍的项目，该项目分析了1400个家庭中未确诊重症儿的外显子。他说，“即使当一个单一变异能解释疾病，也还需要进行统计调查、分析数据以及开展临床试验，以提供最终的确诊。”

生物统计学2024

深度学习的问题。。。



Majority vote? Or rely on expert?

Case 1

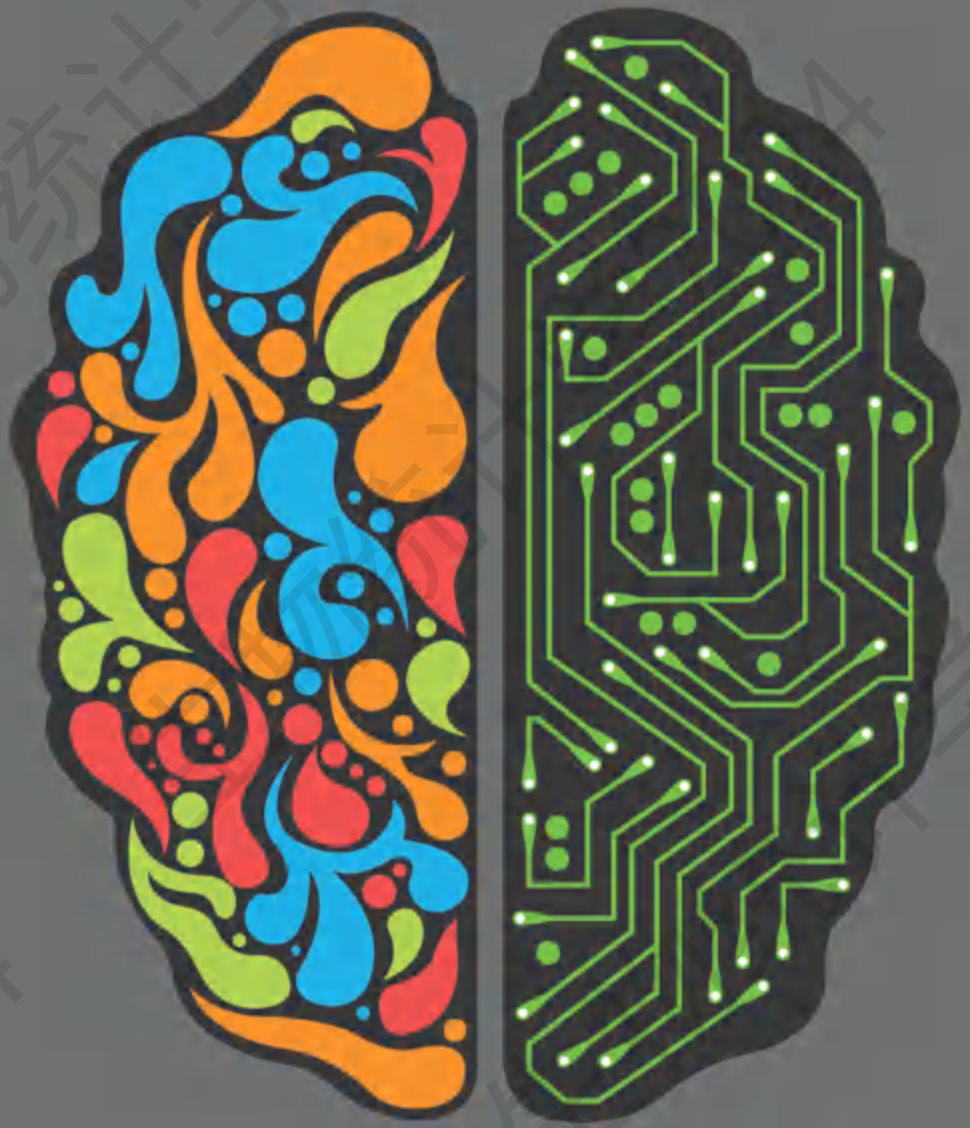
一名40岁女子从小被诊断为重型地中海贫血及血小板过多症，后来并被切除脾臟及长期服用抑制血小板功能的药物，进一步做全外显子基因体定序，才发现真正的致病机转是因细胞内质网相关的基因突变引起「先天性红血球生成异常性贫血」，输血治疗地中海贫血反受其害。

Case 2

2007年，美国最大的分子诊断公司Quest Diagnostics子公司——Athena Diagnostics 给患儿Christian Millare进行了SCN1A基因突变检测。2008年5月1日，两岁的Christian疾病发作并不幸去世。此后死亡患儿母亲Amy Williams 通过咨询专家和查阅发表的文献，再结合Christian的病历，她坚信儿子的生命不应该这样过早的结束。8年后，Williams对Quest 诊断公司和 Athena诊断公司提出

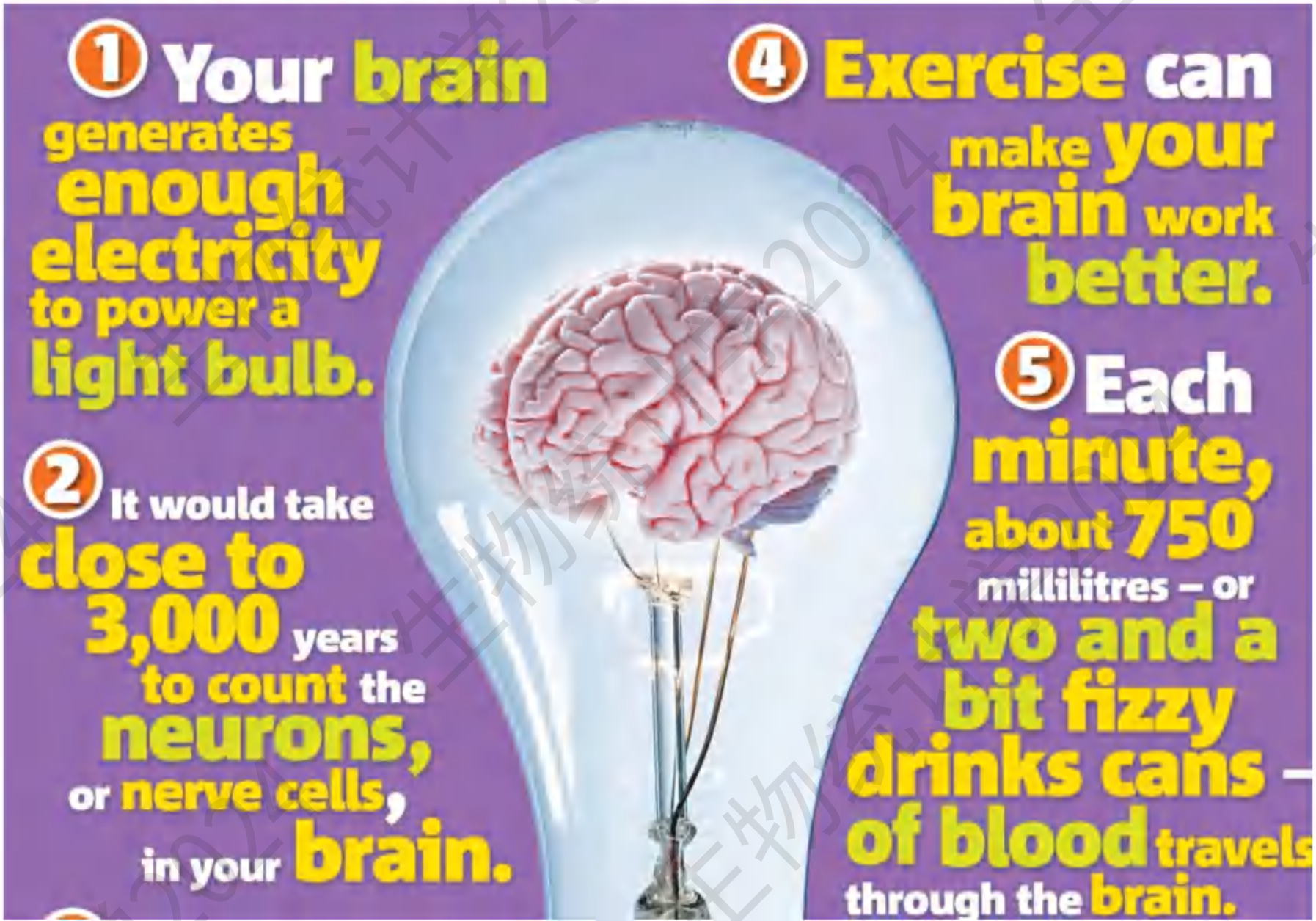
生物统计学2024

深度学习的问题：脑科学。。。



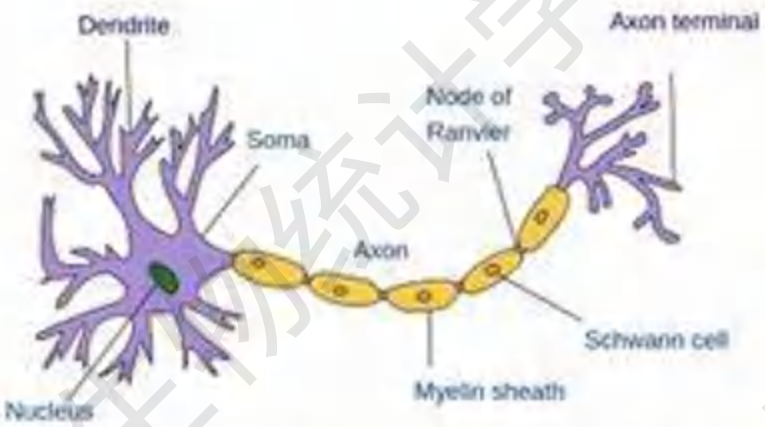
生物统计学2024

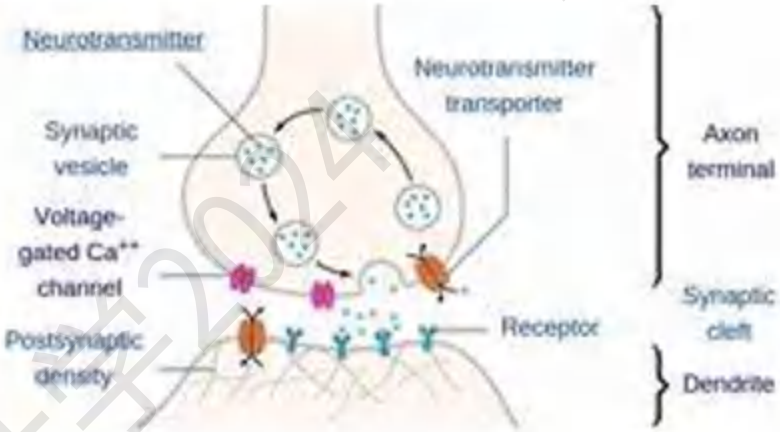
深度学习的问题：脑科学。。。

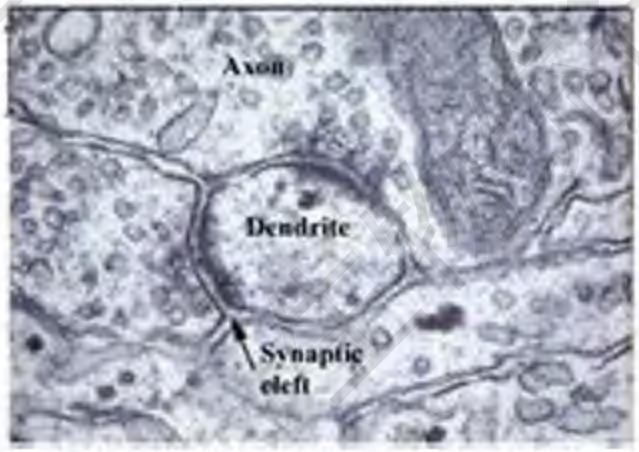


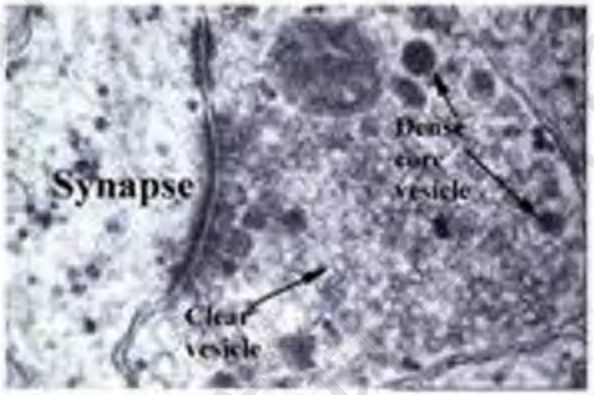
生物统计学2024 生物统计学2024 生物统计学2024

深度学习的问题：脑科学。。。







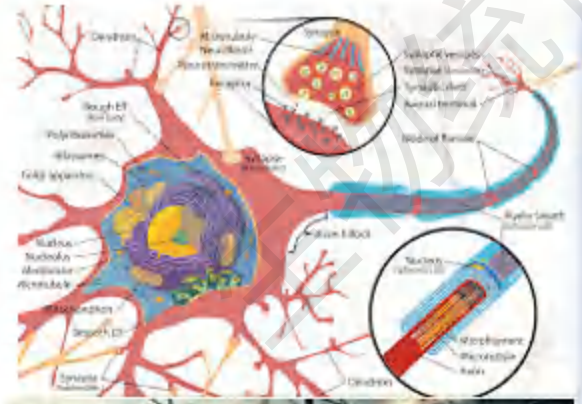


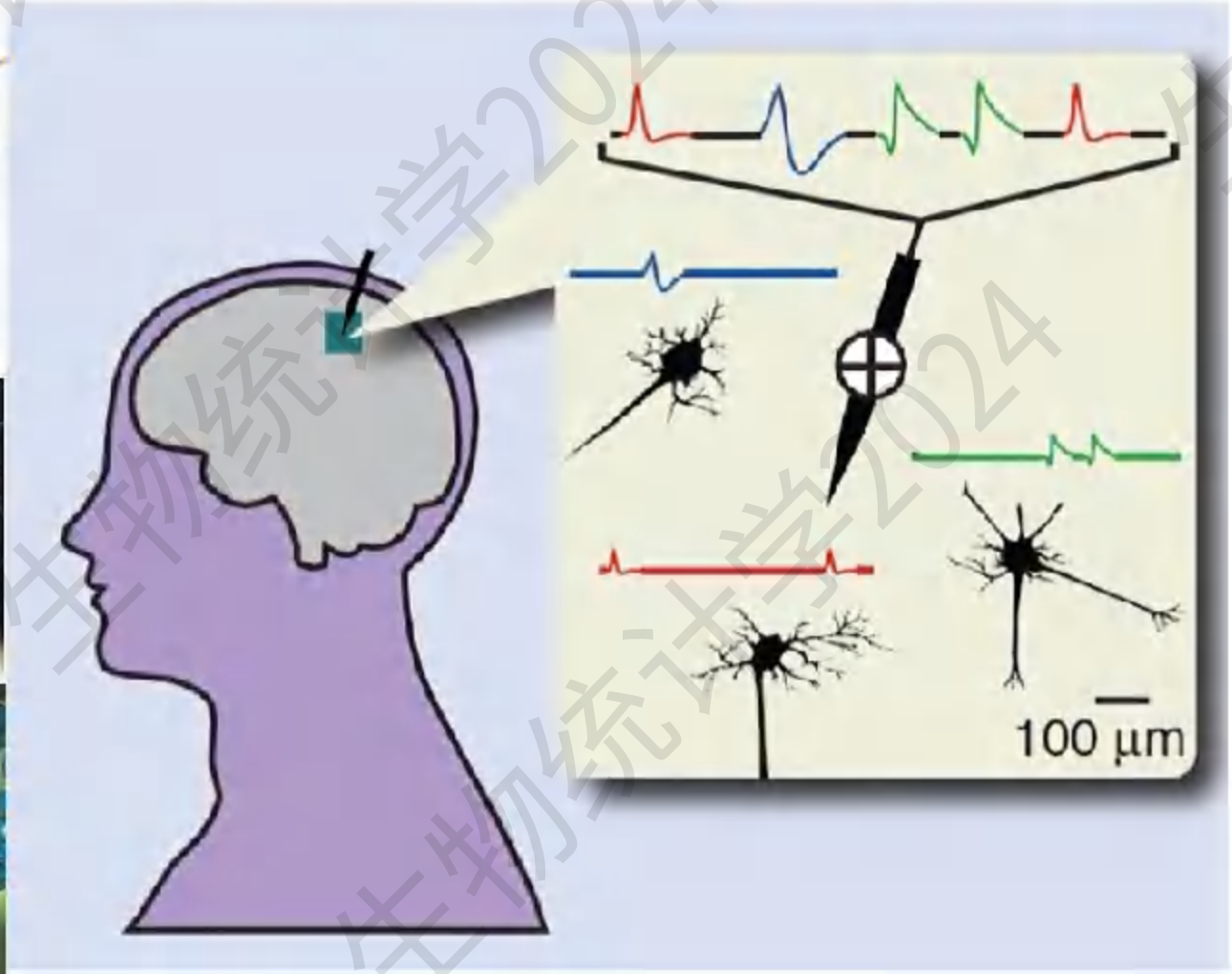
The Brain vs. Deep Learning vs. Singularity

生物统计学2024

深度学习的问题：脑科学。。。

We are actually HMM (or Deep learning) animal...





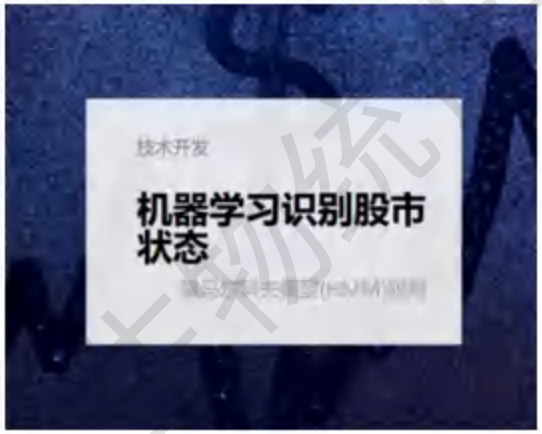
生物统计学2024

深度学习的问题。。。

More problems that you can think of?

生物统计学2024

不建议用简单的深度学习来解决的问题。。。



生物统计学2024

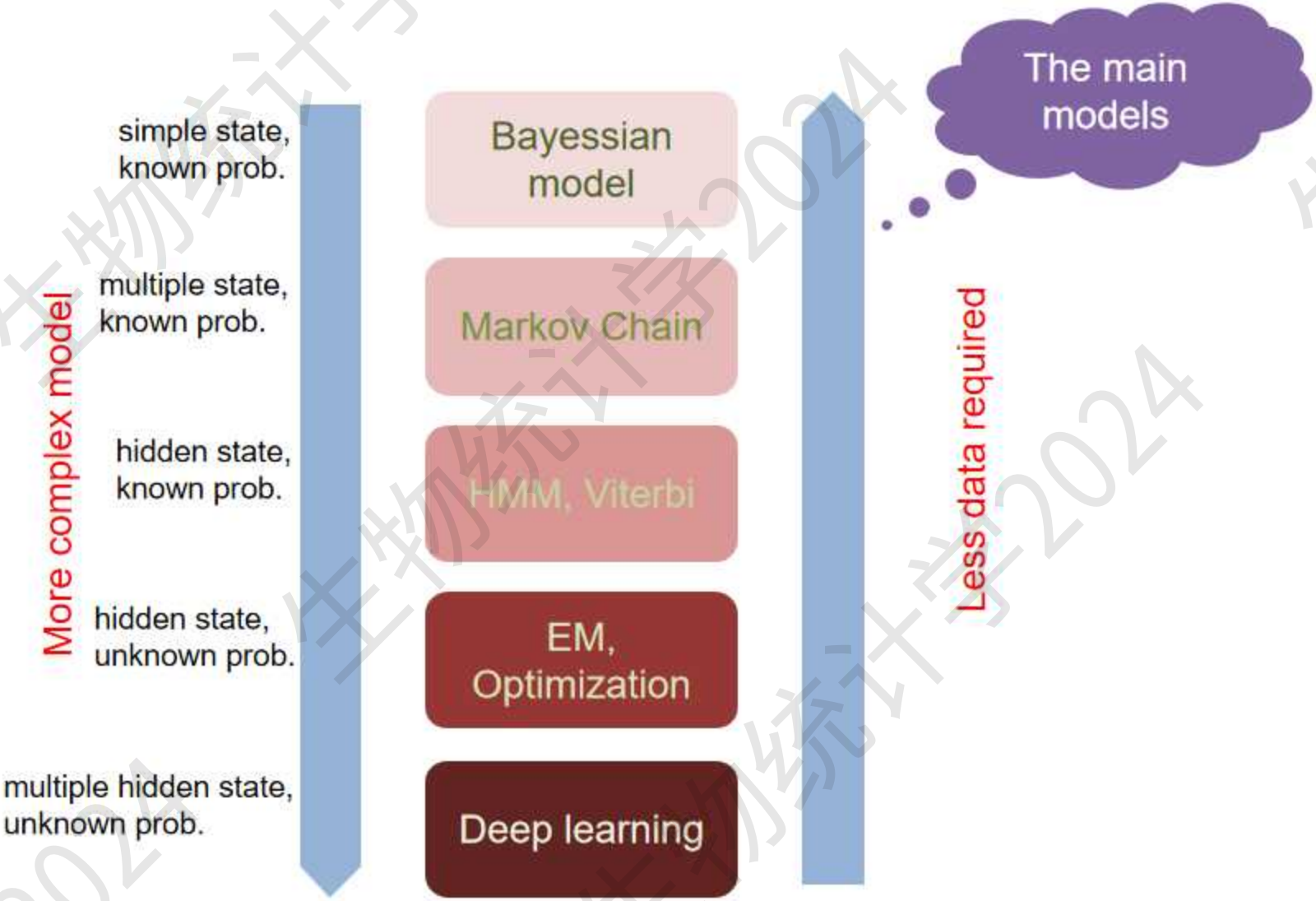
生物统计学2024

Statistical modeling

Distribution, Modeling, Prediction

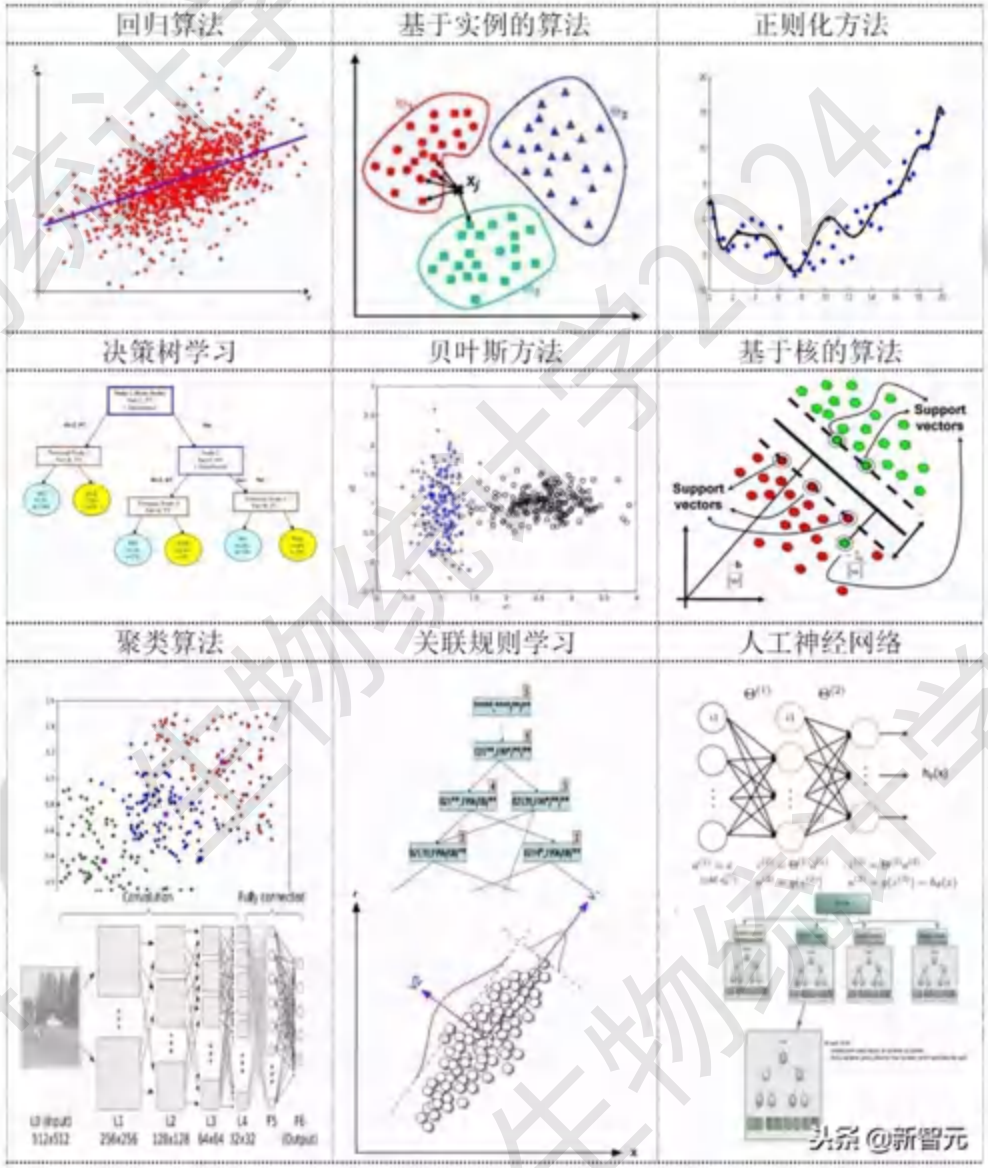
生物统计学2024

Statistical modeling



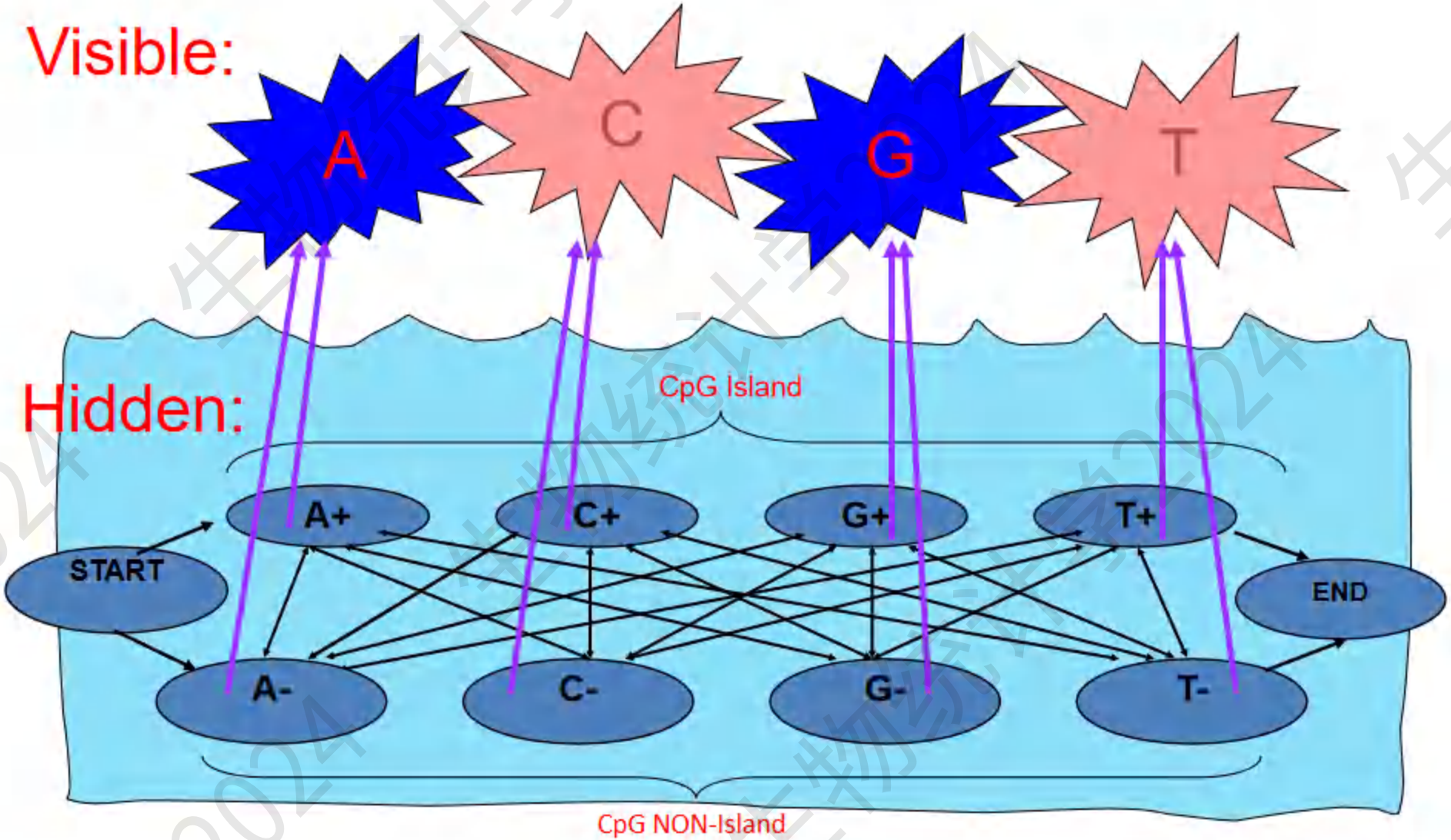
生物统计学2024

Statistical modeling



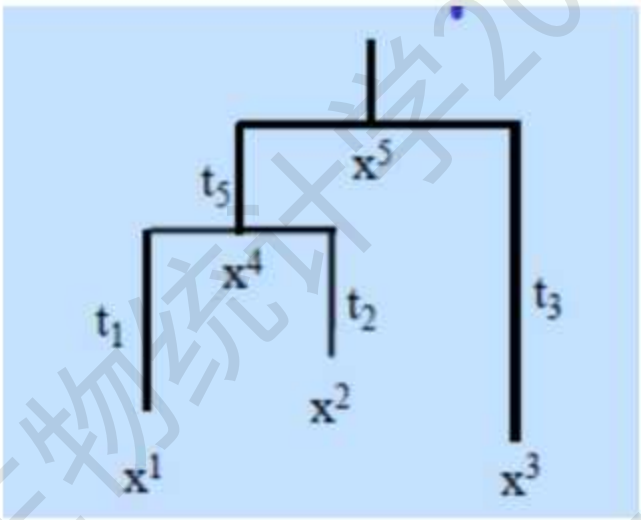
生物统计学2024

HMM算法：基因预测


Visible:
A
C
G
T
Hidden:
CpG Island
A+
C+
G+
T+
START
END
A-
C-
G-
T-
CpG NON-Island

Bayessian算法：进化树

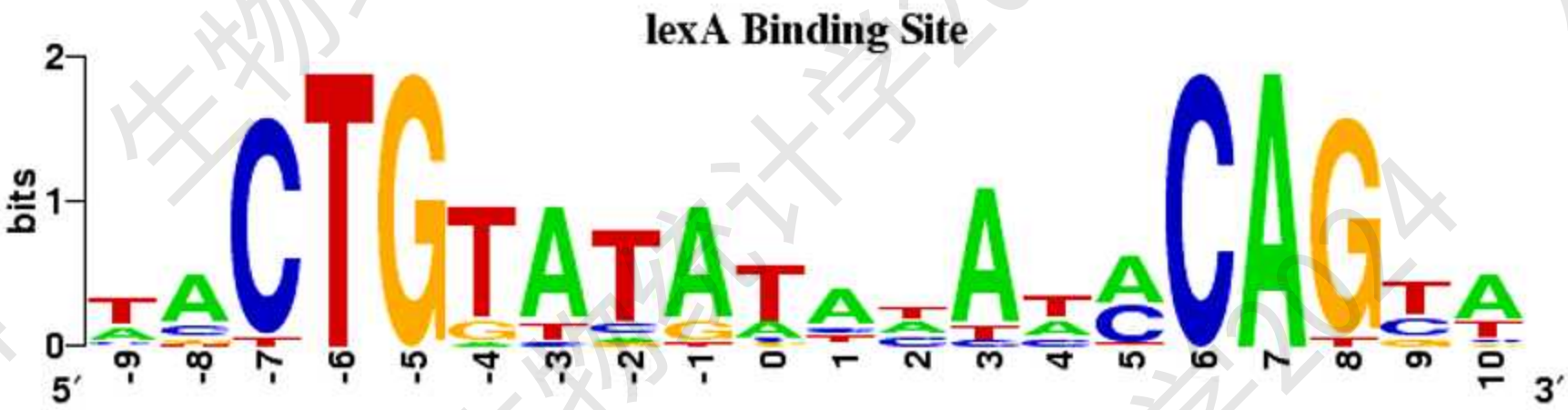
- Given the tree



$$P(x_1, x_2, x_3, x_4, x_5 | T, t^*)$$
$$= P(x_1|x_4, t_1)P(x_2|x_4, t_2)P(x_3|x_5, t_3)P(x_4|x_5, t_5)$$

生物统计学2024

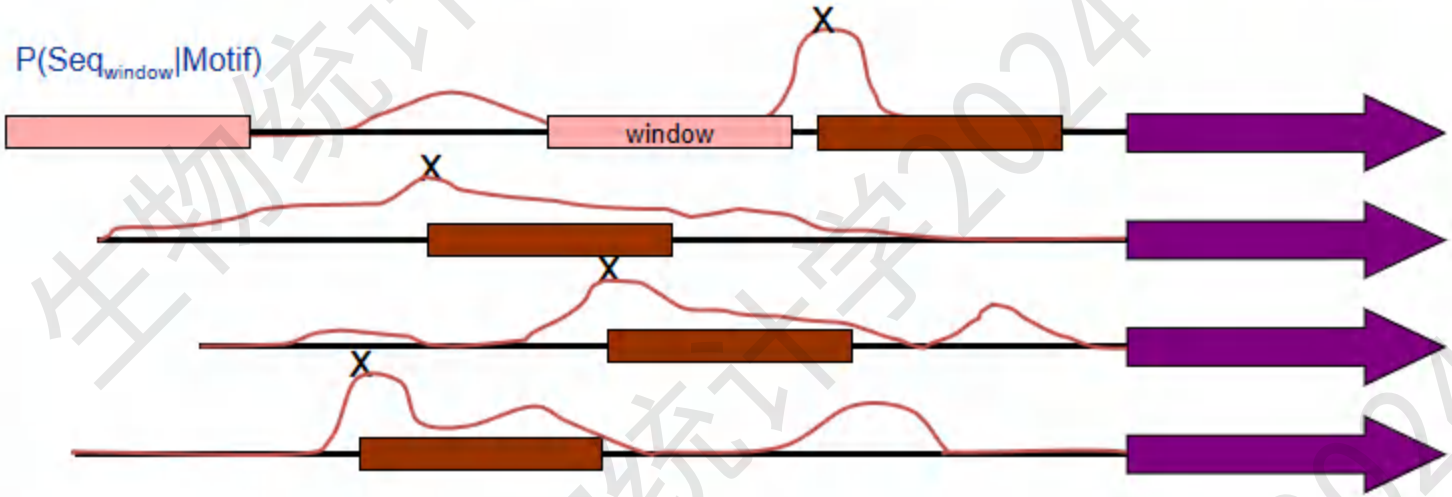
信息熵算法：量化保守区域



生物统计学2024

EM算法：motif识别

Given multiple sequences and motif model but no motif *locations*

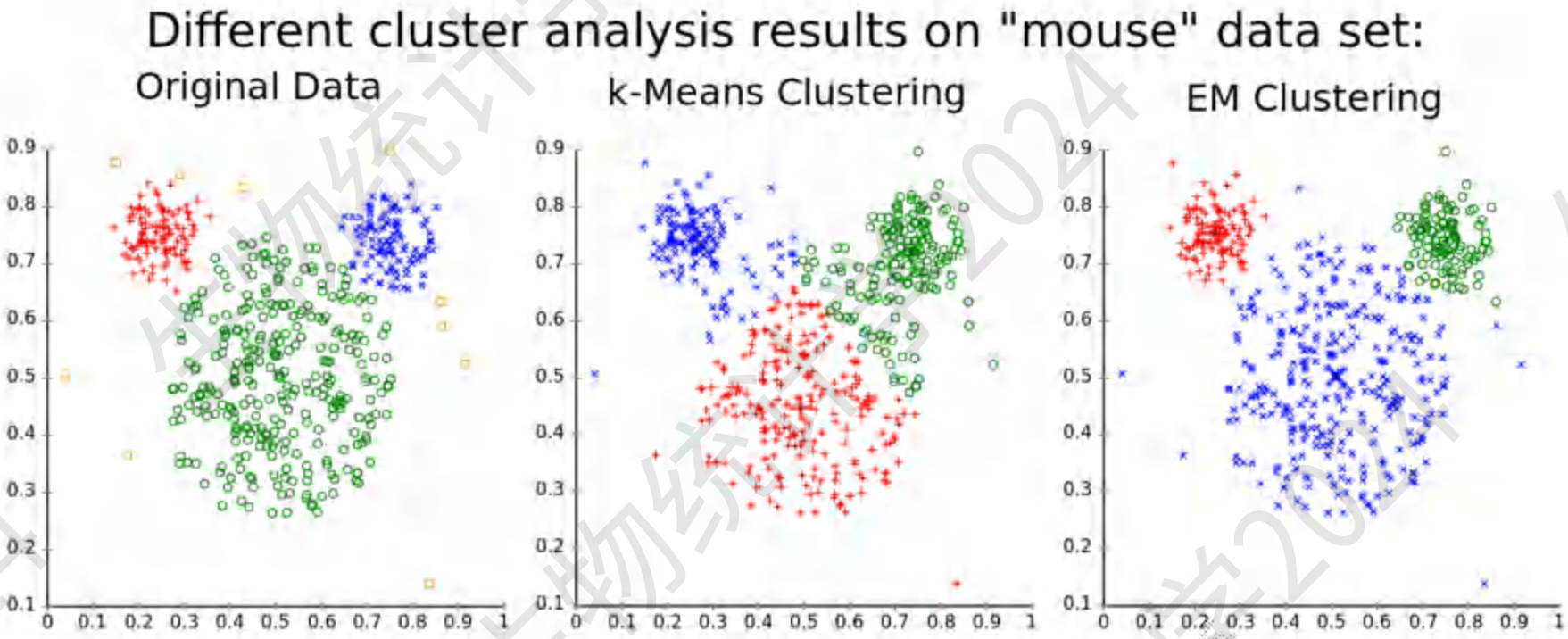


Calculate P(Seq$_{window}$|Motif) for every starting location

Choose best starting location in each sequence

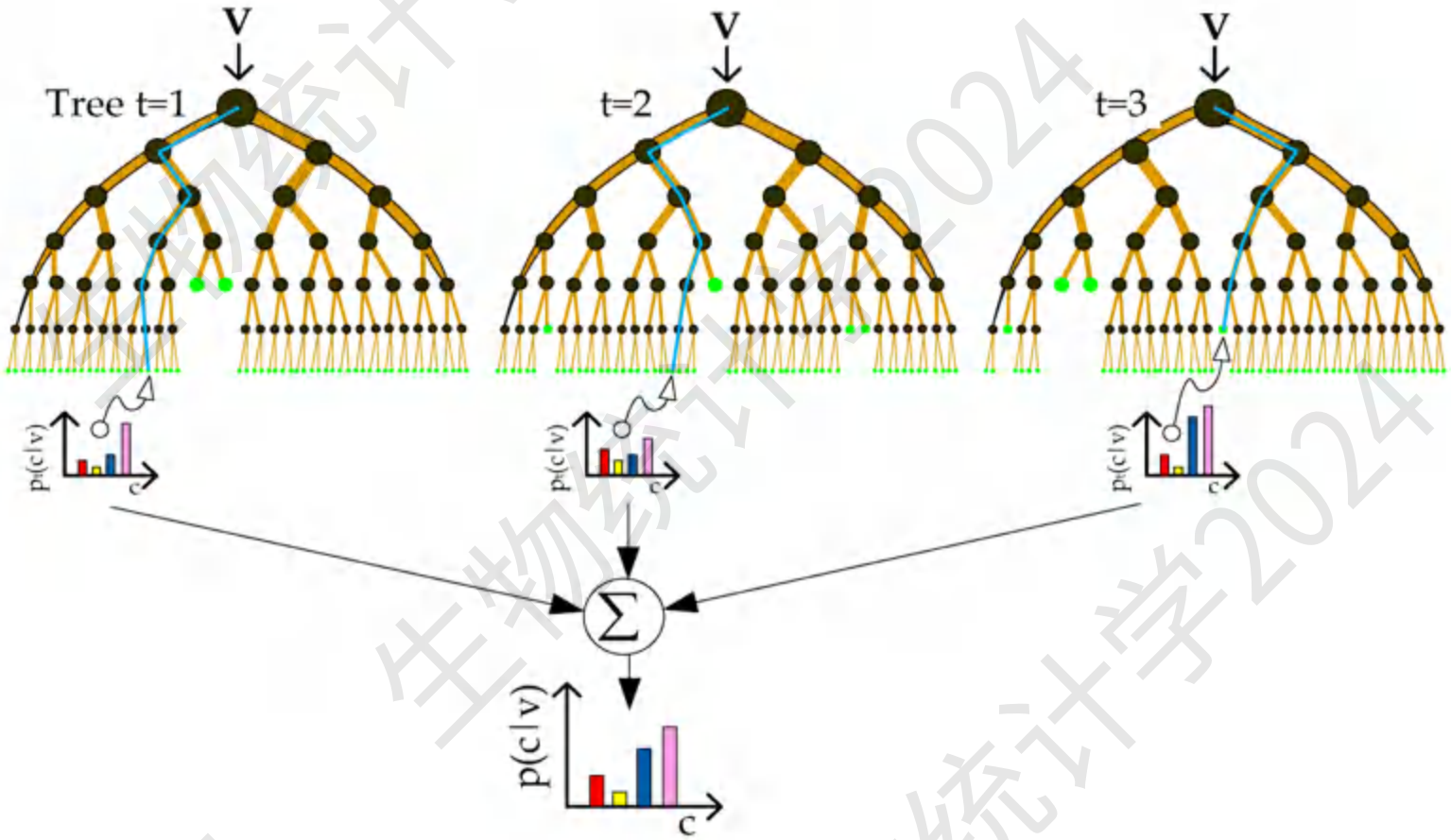
生物统计学2024

EM聚类算法：样本聚类



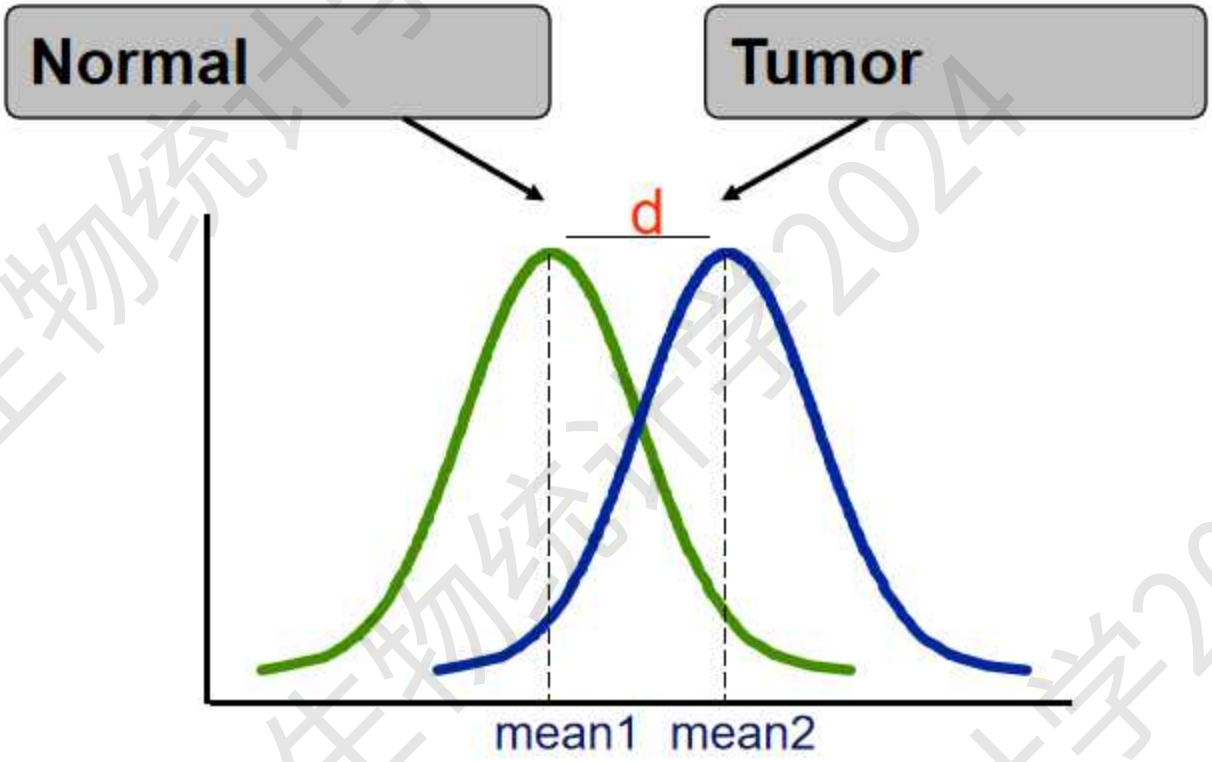
EM聚类与K-Means不同之处：并不计算距离，而是计算概率（并且明显要比K-Means复杂的多），用一个给定的多元高斯概率分布模型来估计出一个数据点属于一个聚类的概率，即将每一个聚类看作是一个高斯模型。

Random Forest算法：分类和特征选择



生物统计学2024

假设检验：分类问题

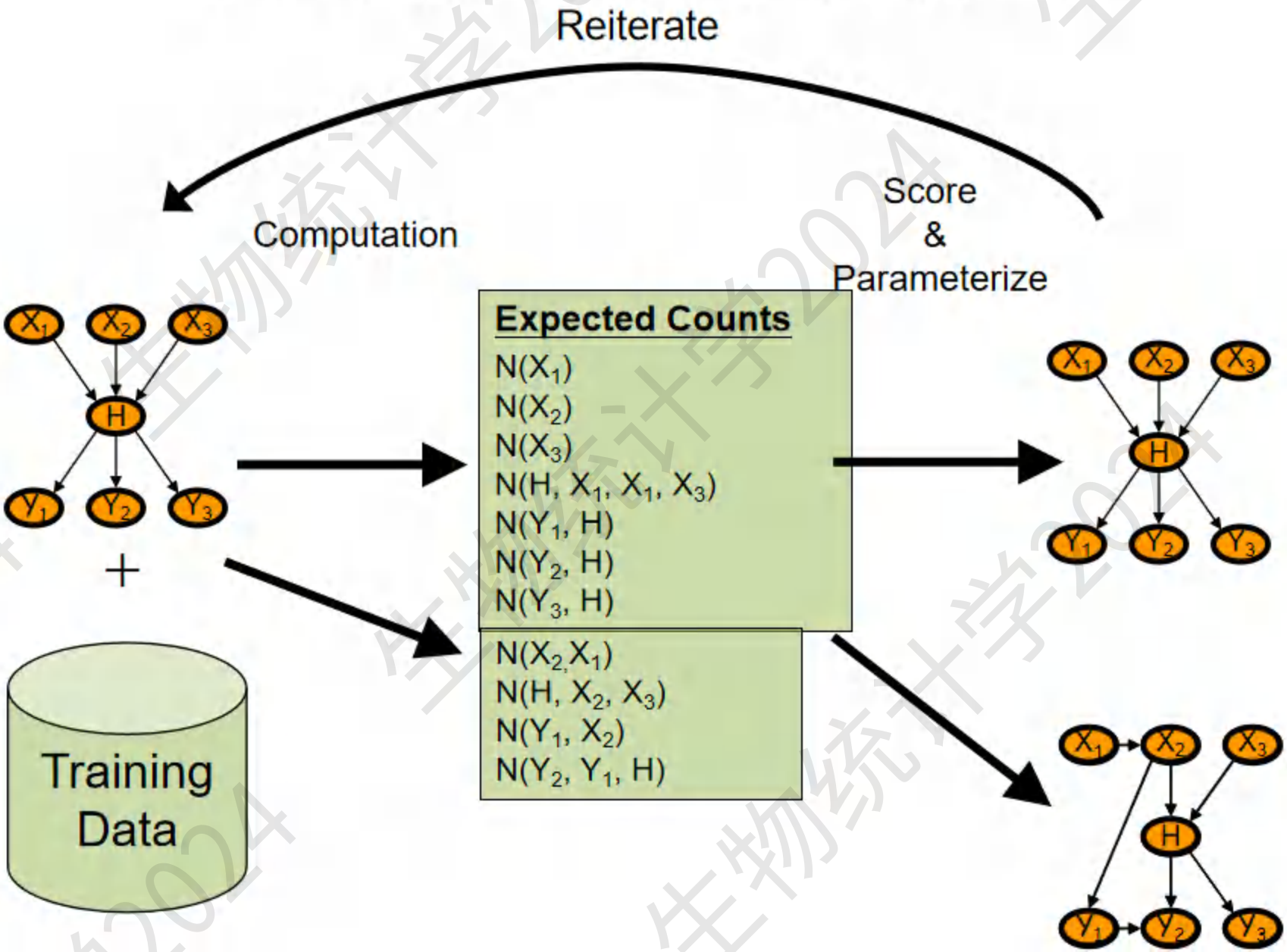


Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternative hypotheses $H_1 : \mu_1 \neq \mu_2$

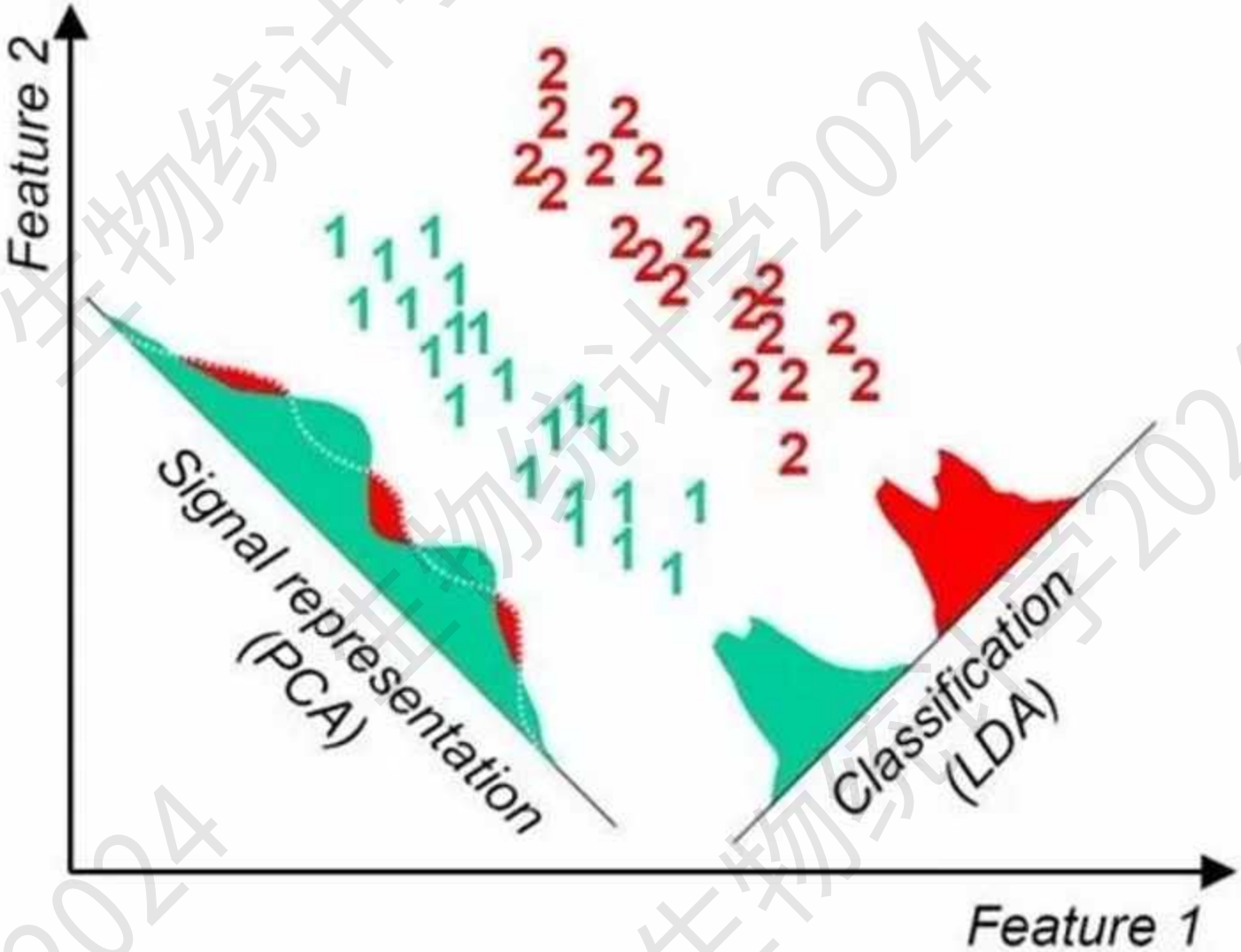
生物统计学2024

贝叶斯网络：因果推断



生物统计学2024

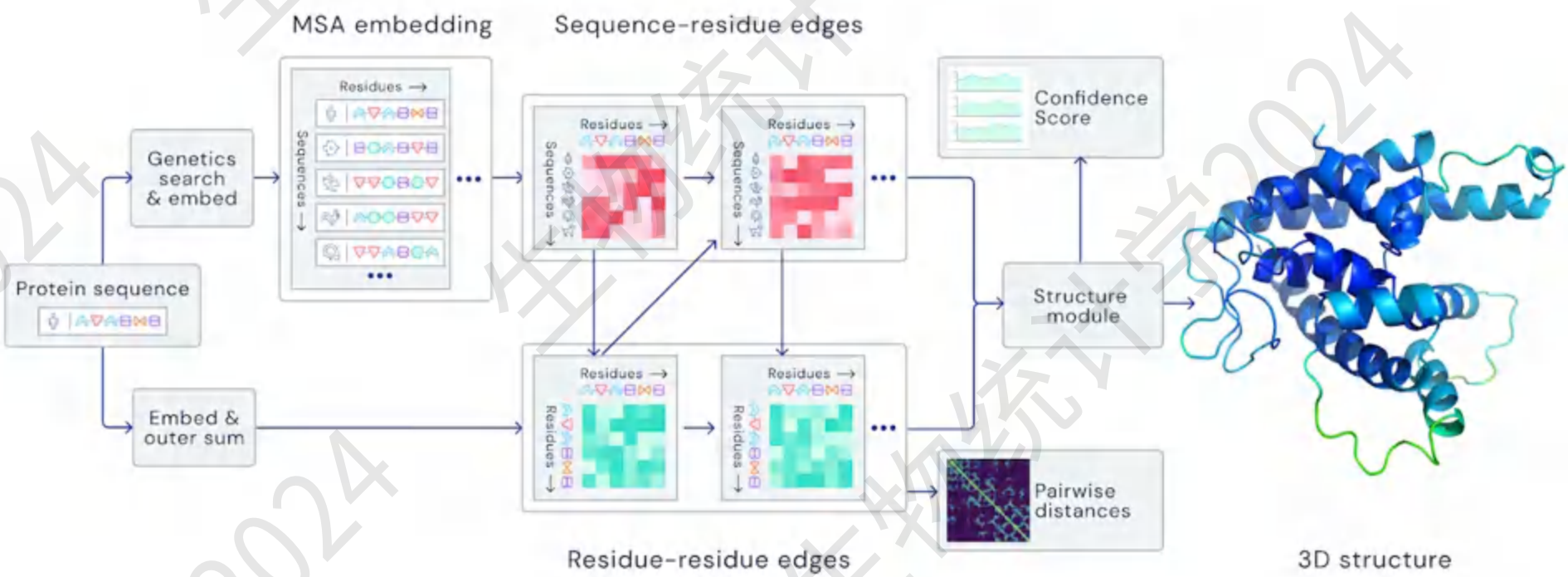
PCA，LDA：降维分析



生物统计学2024

深度学习：广泛应用

AlphaFold2



生物统计学2024

课程安排

- 生物背景和课程简介
- 传统生物统计学及其应用
- 生物统计学和生物大数据挖掘
  - Hidden Markov Model (HMM)及其应用
    - Markov Chain
    - HMM理论
    - HMM和基因识别 (Topic I)
    - HMM和序列比对 (Topic II)
  - 进化树的概率模型 (Topic III)
  - Motif finding中的概率模型 (Topic IV)
    - EM algorithm
    - Markov Chain Monte Carlo (MCMC)
  - 基因表达数据分析 (Topic V)
    - 聚类分析-Mixture model
    - Classification-Lasso Based variable selection
  - 基因网络推断 (Topic VI)
    - Bayesian网络
    - Gaussian Graphical Model
  - 基因网络分析 (Topic VII)
    - Network clustering
    - Network Motif
    - Markov random field (MRF)
  - Dimension reduction及其应用 (Topic VIII)
- 面向生物大数据挖掘的深度学习

研究对象:
生物序列,
进化树,
生物网络,
基因表达
...

方法:
生物计算与生物统计

生物统计学2024

Good luck in biostatistics!

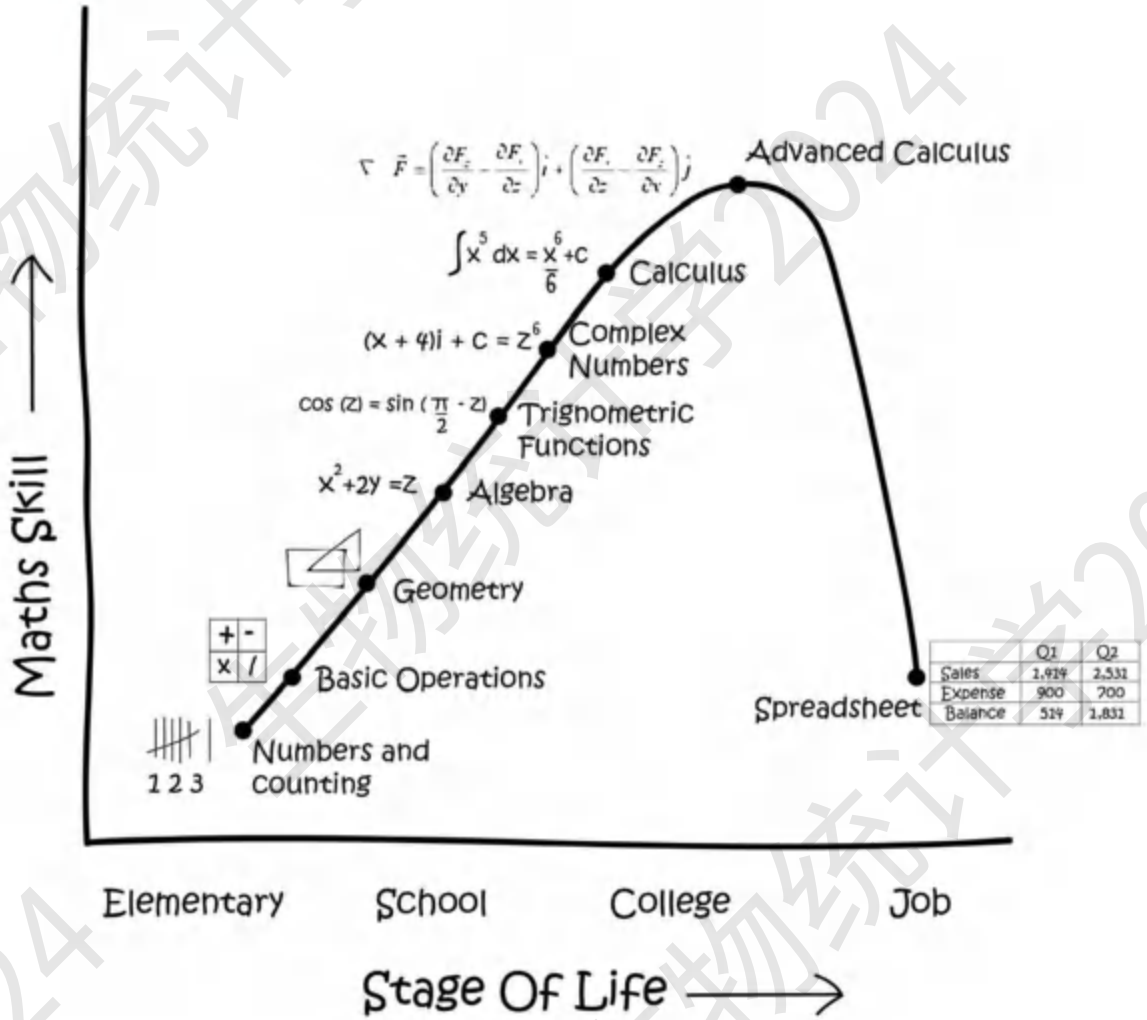
Noah Williams
@Bellmanequation

Found on the roadside: @SpringerStats yellow book, beside empty vodka bottles & cigarette packs

p<0.05

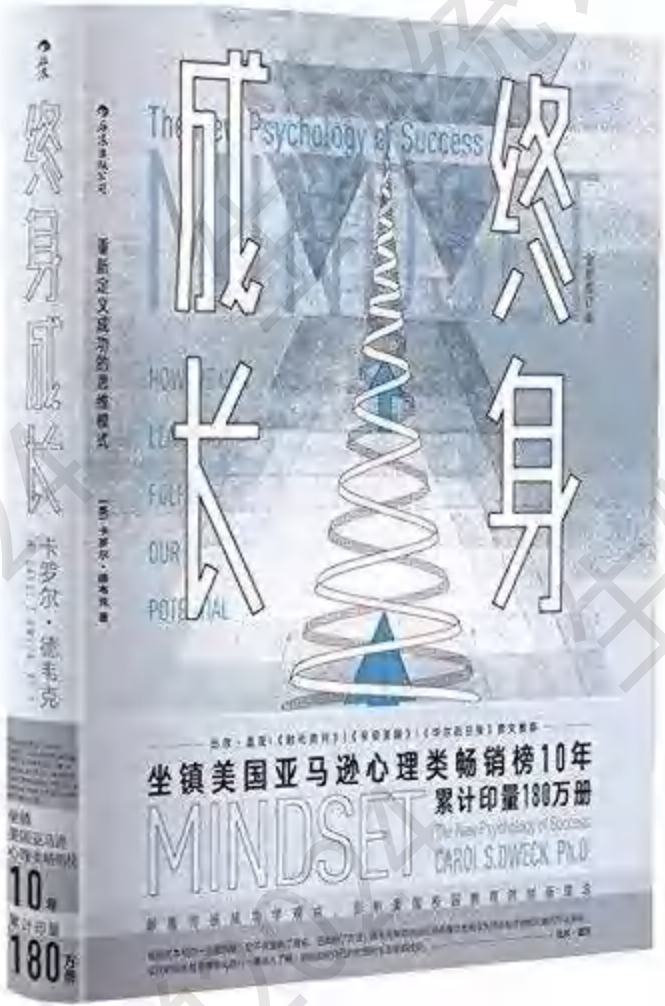
生物统计学2024

Maybe you will not face these many statistical problems...



生物统计学2024

But you have the knowledge



生物统计学2024

Statistical modeling

More to learn:

- Fast Fourier Transformation: https://zhuanlan.zhihu.com/p/19763358
- Causality analysis: https://zhuanlan.zhihu.com/p/33860572
- Ecological modeling (mathematical modeling): https://zhuanlan.zhihu.com/p/112822183
- ...

生物统计学2024

Statistical modeling

More to learn:

- 生物统计与生物信息的区别与联系？（https://www.zhihu.com/question/30284194）
  - 生物学的生物信息学：首先是讲进化，做序列比对，blast，构建进化树；然后，讲基因功能，基因功能富集分析等。侧重的是，将进化中的生物学原理，参数如何设置，软件如何使用，如何将这些软件应用到生物学问题。
  - 计算机系的生物信息学：首先讲的也是进化，但是，讲的是算法设计；然后讲了很多马尔科夫链，HMM模型，序列比对中的blosum62矩阵是怎么来的，如何加快计算效率，如何降低空间存储等等。侧重的是，如何设计合理的算法，给生物学的人使用。
  - 数学系的系统生物学：讲了回归中的penalty function，然后我学了LASSO回归；后来学了SVM中的核函数；还有Gibbs samping，MCMC；还有时间序列中的Granger因果推断等等。
  - 生统对数学要求更高，生信基本不需要。国内生统大多是挂XXX卖XXX

- Yoshua Bengio:
  - **你死我活那是邪教，开放包容才是正道。**

- What do you think?

生物统计学2024

Biostatistics in the market...

[公司文件] 加强与国内大学合作，吸纳全球优秀人才，共同推动中国基础研究
——任总与中国科学技术大学包信和校长座谈的讲话

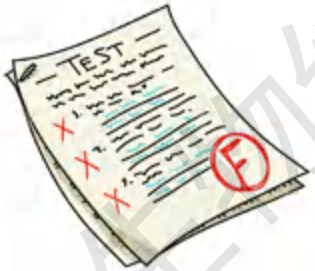
总 裁 办 电 子 邮 件

电邮讲话【2018】128号　　签发人：任正非

加强与国内大学合作，吸纳全球优秀人才，共同推动中国基础研究
——任总与中国科学技术大学包信和校长座谈的讲话
2018年11月19日

在高校学科设置上，我特别支持你们重视**统计学**。计算机科学不仅仅是技术，还应该以**统计学**为基础。大数据需要**统计学**，信息科学需要**统计学**，**生命科学也需要统计学**。国家要搞人工智能，更要重视**统计学**。**统计学**不是一个纯粹的学科，而是每一个学科都要以**统计学**为基础。

生物统计学2024

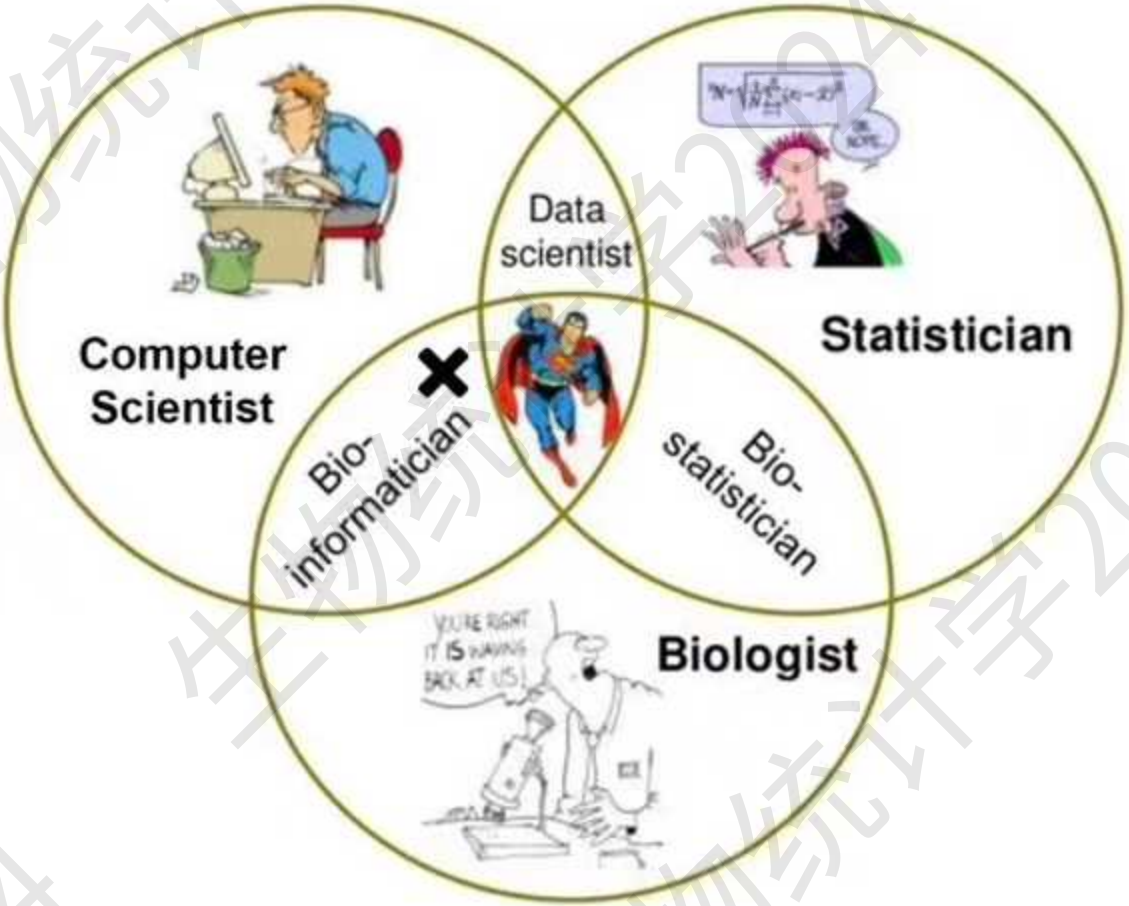


Or

Good luck in final exam!

生物统计学2024

Superman/Wonder woman



Good luck in your life!

生物统计学2024

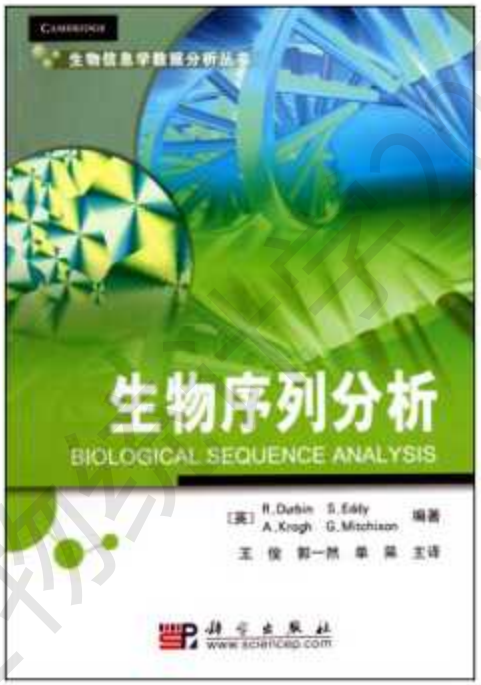
谢谢大家！

生物统计学2024

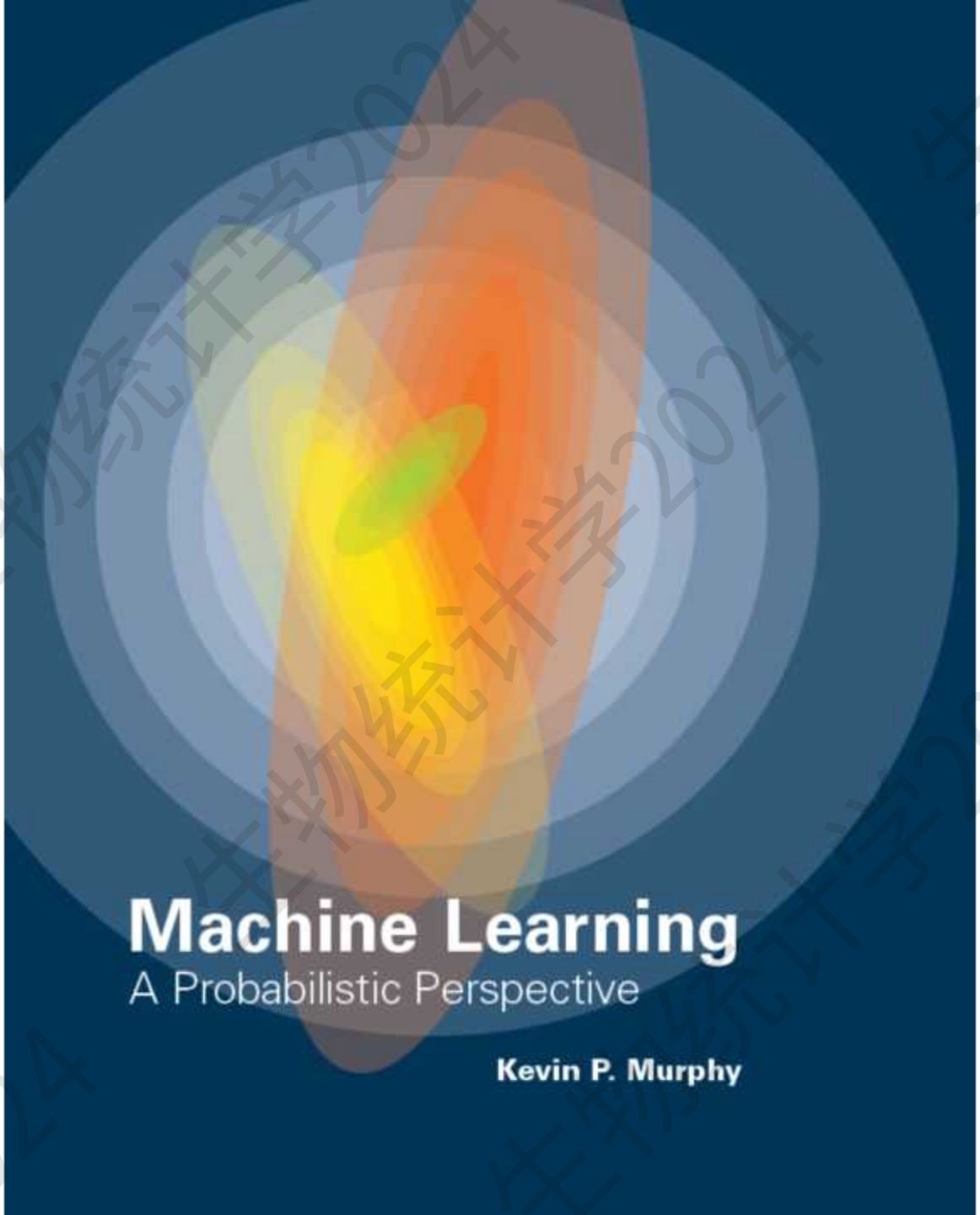
希望大家和我一起，终身学习！

生物统计学2024

References



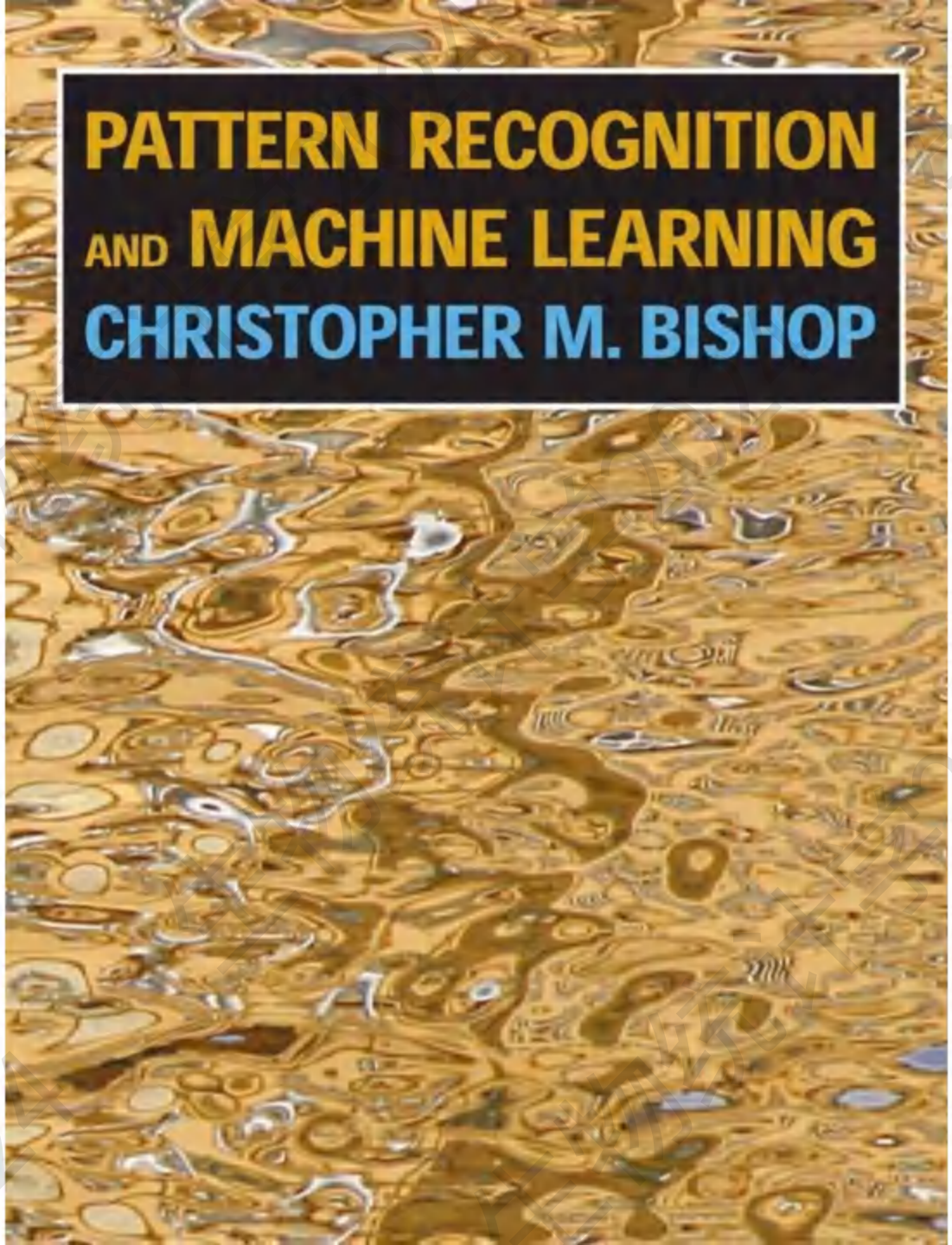
生物统计学2024


Machine Learning
A Probabilistic Perspective
Kevin P. Murphy

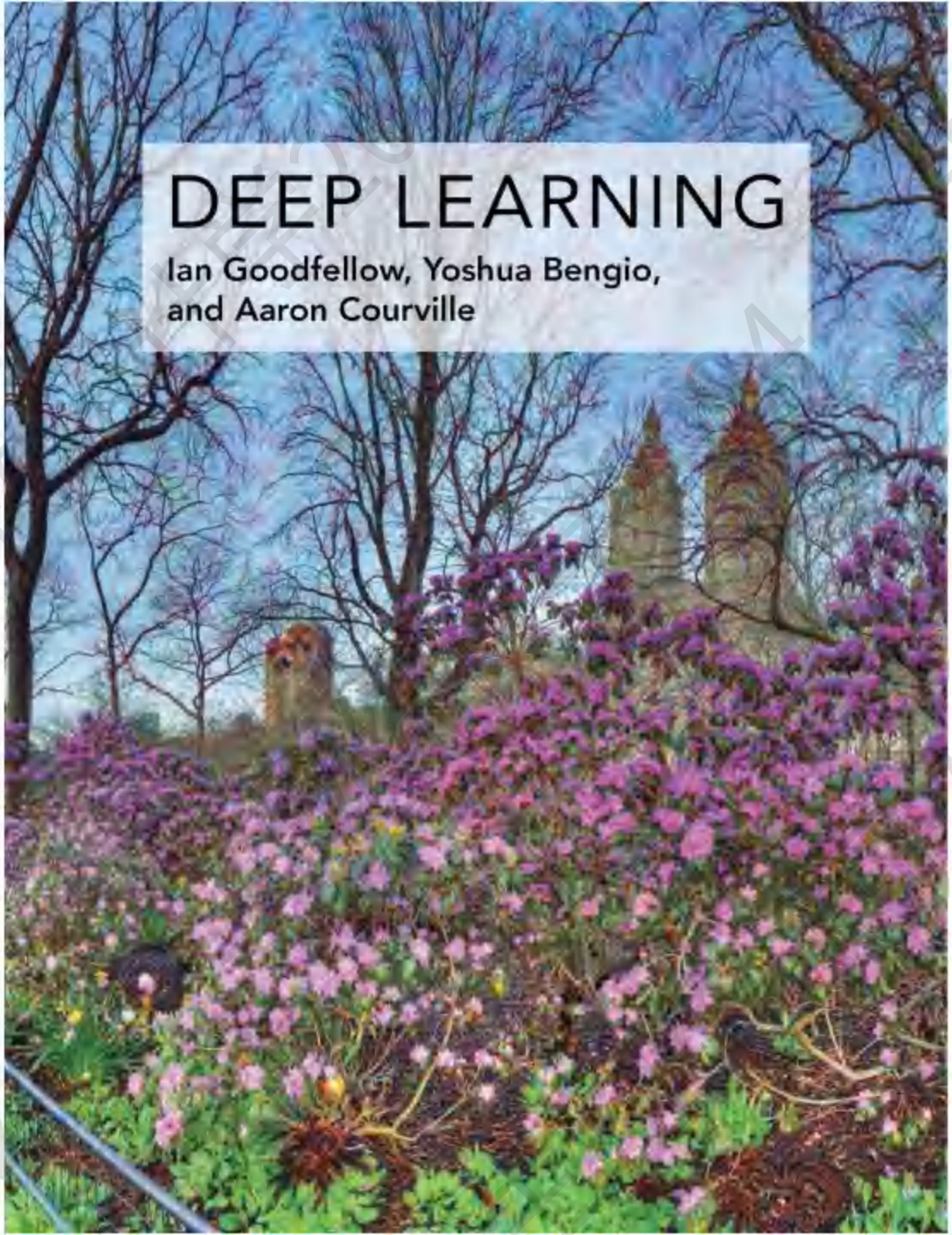
生物统计学2024

PATTERN RECOGNITION AND MACHINE LEARNING

CHRISTOPHER M. BISHOP



生物医学工程2024


DEEP LEARNING
Ian Goodfellow, Yoshua Bengio,
and Aaron Courville

Slides credits

——生物信息学研究方法概述：北京大学生物信息中心

——生物统计学：卜东波@中国科学院计算技术研究所，邓明华@北京大学

——神经网络与深度学习: 邱锡鹏@复旦大学

——Introduction to Computational Biology and Bioinformatics: Xiaole Shirley Liu Lab@Havard University

——Combinatorial Methods in Computation Biology: Ken Sung Lab@NUS

——Deep Learning in the Life Sciences: MIT

——Probabilistic Graphical Models: Eric Xing@CMU

——Numerous other leading researchers and leading labs.....

生物统计学2024

生物统计学2024

补充知识

生物统计学与深度学习：

- Boost方法的适用性
- 从AlphaGo的四世同堂看深度学习的方法进步
- 深度学习工具介绍

生物统计学2024

Boost方法的适用性

nature machine intelligence

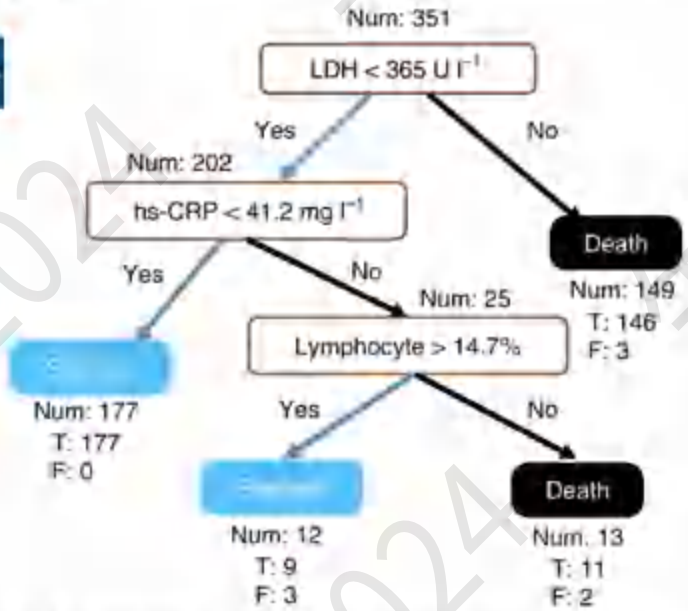
We'd like to understand how you use our websites in order to improve

Article Published: 14 May 2020

An interpretable mortality prediction model for COVID-19 patients

Li Yan, Han-Tao Zhang, Jorge Goncalves, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun, Xiuchuan Tang, Liang Jing, Mingyang Zhang, Xiang Huang, Ying Xiao, Haosen Cao, Yanyan Chen, Tongxin Ren, Fang Wang, Yaru Xiao, Sufang Huang, Xi Tan, Niannian Huang, Bo Jiao, Cheng Cheng, Yong Zhang, Ailin Luo, Laurent Mombaerts, Junyang Jin, Zhiguo Cao, Shusheng Li, Hui Xu & Ye Yuan ~Show fewer authors

Fig. 2: A decision rule using three key features and their thresholds in absolute value.



Num, the number of patients in a class; T, the number of correctly classified; F, the number of misclassified patients.

- 该团队用于建模的样本为2020年1月10日至2月18日间从同济医院的485患者中采集，团队从中寻找可靠且有意义的死亡风险标志物。其中110个样本用于测试模型，其余375例用于建模。在分析所囊括的375例病例中，201例在COVID-19康复后出院，其余174例死亡。
- 作者根据机器学习算法设计了一套数学建模方法，这些算法专门用于识别最能预测患者死亡率的生物标志物。作者在算法中将问题变成一个分类任务，输入数据包括基本信息、症状、血液标本，以及普通、重症、危重症患者的实验室检查结果，如肝功能、肾功能、凝血功能、电解质和炎症因子。
- 该模型将乳酸脱氢酶（LDH）、淋巴细胞、超敏C反应蛋白水平选为识别病危患者的三个最关键的生物标志物。这一结果与当前的医学知识相符，即LDH水平较高本身与不同疾病出现的组织损伤有关，包括肺炎等肺部疾病。大部分患者在住院期间都采集过多次血样，但模型只使用了患者最终标本的数据。尽管如此，模型也适用于所有其他血液标本，对生物标志物的预测潜力也可以估计。

生物统计学2024

Boost方法的适用性

nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

Replication of a mortality prediction model in Dutch patients with COVID-19

Marian J. R. Quanjel, Thijs C. van Holten, Pieternel C. Gunst-van der Vliet, Jette Wielaard, Bekir Karakaya, Maaike Söhne, Hazra S. Moeniralam & Jan C. Grutters

nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

Limited applicability of a COVID-19 specific mortality prediction rule to the intensive care setting

C. Dupuis, E. De Montmollin, M. Neuville, B. Mourvillier, S. Ruckly & J. F. Timsit

nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

External validation demonstrates limited clinical utility of the interpretable mortality prediction model for patients with COVID-19

生物统计学2024

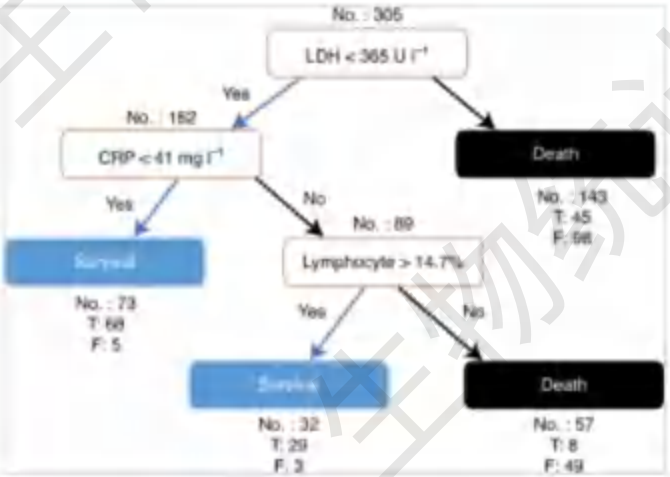
Boost方法的适用性

nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

Replication of a mortality prediction model in Dutch patients with COVID-19

Marian J. R. Quanjel, Thijs C. van Holten, Pieternel C. Gunst-van der Vliet, Jette Wielaard, Bekir Karakaya, Maaike Söhne, Hazra S. Moeniralam & Jan C. Grutters



The decision rule using three key features and their thresholds in absolute value. Num, the number of patients in a class; T, the number of correctly classified; F, the number of misclassified patients.

“在我们的303名患者中，有42名进入重症监护室（ICU）。在这些患者中，36人（86%）根据决策规则会有不利的结果，但实际只有15名（36%）死亡（都满足LDH >365 U l-1).”

生物统计学2024

Boost方法的适用性

nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

Limited applicability of a COVID-19 specific mortality prediction rule to the intensive care setting

C. Dupuis, E. De Montmollin, M. Neuville, B. Mourvillier, S. Ruckly & J. F. Timsit

Table 3 Performance of the decision rule of Yan et al. on the French Outcomerea dataset

From: Limited applicability of a COVID-19 specific mortality prediction rule to the intensive care setting

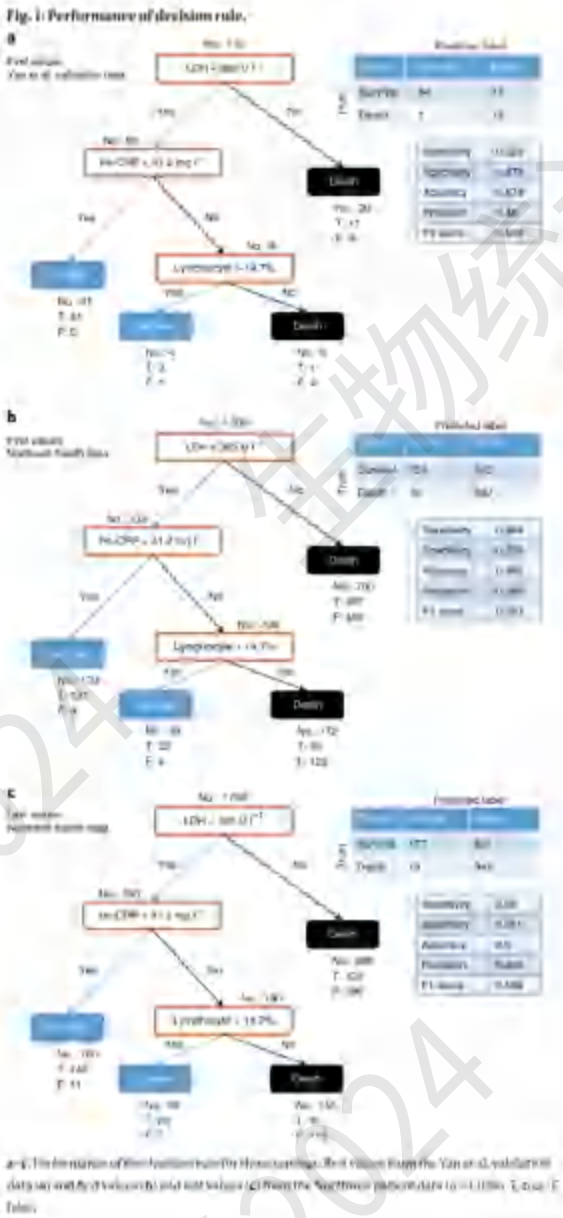
		Precision	Recall	F1 score	Support[a]
Day 7	Survival	1	0.14	0.24	21
	Death	0.86	1	0.71	155
	Accuracy			0.16	176
Day 14	Survival	0.87	0.14	0.24	21
	Death	0.91	0.92	0.85	155
	Accuracy			0.80	176
Day 28	Survival	0.63	0.16	0.27	21
	Death	0.37	0.93	0.53	155
	Accuracy			0.43	176

[a]Predicted number of patients.

“这些结果可以用我们队列的真实特异性来解释。确实，只有约5%的COVID-19患者因急性低氧血症性呼吸衰竭（AHRF）被送入ICU 。因此，我们的ICU人群不包括①绝大多数LDH和hs-CRP血清水平很低且淋巴细胞计数很高的丘疹症状患者（这些患者预后良好），以及②一些病情最严重的患者hs-CRP和LDH血清水平高且淋巴细胞计数低的患者，由于治疗受限而未入ICU（这些患者的预后最差）。由此，Yan等人的预测规则在我们的队列中应用不准确也并不奇怪。”

生物统计学2024

Boost方法的适用性



nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

External validation demonstrates limited clinical utility of the interpretable mortality prediction model for patients with COVID-19

Matthew Barish, Siavash Bolourani, Lawrence F. Lau, Sareen Shah & Theodoros P. Zanos

"为了测试死亡率预测模型的临床可移植性，我们使用Northwell Health电子健康记录数据库在外部对其进行了验证。在总共13,106名患者中，最终纳入了1,038名患者进行模型验证。

我们最初使用所有三个实验室值均可用的第一个时间点测试了模型性能（图.1b）。模拟该模型在此初始分类点的操作，死亡的精度为0.40（F1得分为0.56），总精度为0.48。

据报道，随着实验室值越来越接近患者的结果，该模型的准确性会提高。如前所述，取决于事先知道结果日期的临床模型值得怀疑。不过，我们使用数据集中的最终（死亡或出院前）实验室值从外部验证了模型，死亡精度仍然很低，为0.41，总体模型精度为0.50（图.1c）。"

Boost方法的适用性

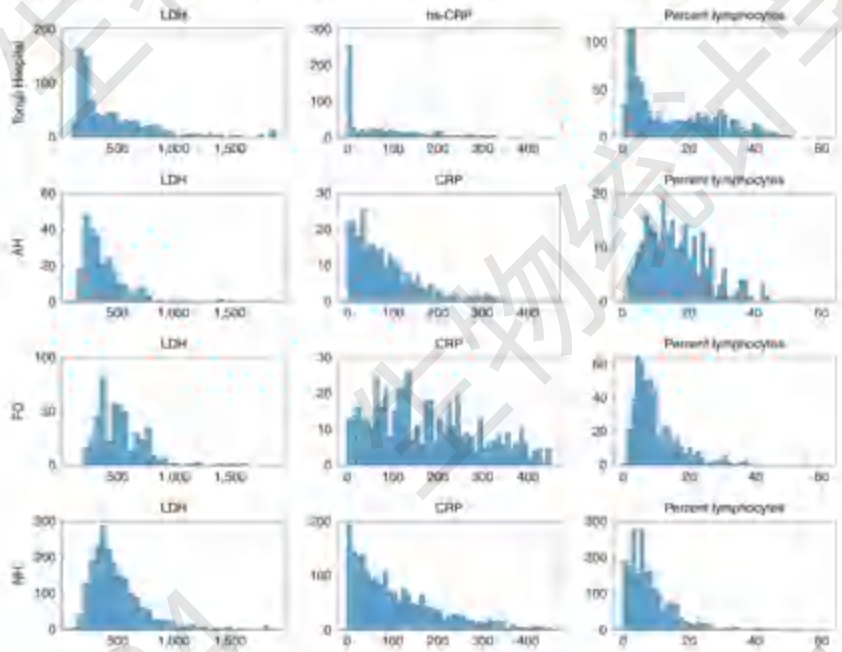
nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

Li Yan et al. reply

Jorge Goncalves, Li Yan, Hai-Tao Zhang, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun, Xiuchuan Tang, Zhiguo Cao, Shusheng Li, Hui Xu, Cheng Cheng, Junyang Jin & Ye Yuan

Fig. 1: Distribution of all three biomarkers for the different datasets.



From left to right: distributions of lactate dehydrogenase (LDH/LD), high-sensitivity C reactive protein (hs-CRP) or C reactive protein (CRP) and percent lymphocytes in blood samples from the different datasets: Tongji Hospital (last 10 days), AH (last 10 days), FO (all data) and NH (all data).

"第一：如下图显示，同济医院和其他医院的所有三种生物标志物的数据分布在统计上是不同的。另外，来自同济医院的数据中的三个生物标志物（训练和外部测试数据相结合）在存活和死亡之间有着清晰的区分。但是，AH（St Antonius医院）、FO（法国Outcomerea）和NH（Northwell Health）的数据集却不是这种情况，他们患者的出院方案与同济医院的出院方案相比较也有明显不同。"

Boost方法的适用性

nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

Li Yan et al. reply

Jorge Goncalves, Li Yan, Hai-Tao Zhang, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun, Xiuchuan Tang, Zhiguo Cao, Shusheng Li, Hui Xu, Cheng Cheng, Junyang Jin & Ye Yuan

“第二：可能与医院中使用的不同实验室规程有关。

第三：LDH表达似乎有亚洲人和白种人之间的实质性的遗传异质性。

第四：不同的医院治疗方法或患者的基线特征会影响预后。

第五：自2020年4月起，重症监护室和非重症监护室的死亡率每周下降2%至5%。这可能会使我们的实验数据与AH、FO、NH数据集的数据之间产生差异。

第六：据报道，SARS-CoV-2病毒至少有两个谱系。患者可能因这些生物标记物的不同表达而感染了不同的菌株。而中国、欧洲和美国似乎被划分为不同的类别。

第七：FO数据集的患者选择未遵循我们使用的完整的患者选择过程，因此不能用作该模型的无偏验证。从本质上讲，本来可以通过我们的模型正确分类的患者被删除，仅留下了一些难以分类的中级患者，从而降低了我们模型的整体准确性。”

生物统计学2024

Boost方法的适用性

nature > nature machine intelligence > matters arising > article

Matters Arising | Published: 12 November 2020

Li Yan et al. reply

Jorge Goncalves, Li Yan, Hai-Tao Zhang, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun, Xiuchuan Tang, Zhiguo Cao, Shusheng Li, Hui Xu, Cheng Cheng, Junyang Jin & Ye Yuan

“我们利用了来自荷兰的St Antonius医院（AH）、法国Outcomerea（FO）和美国Northwell Health（NH）的三个数据集测试了该模型。关键信息如下：

我们使用AH的数据对模型进行了重新训练，整体交叉验证的准确性从53%提高到了83%。

FO患者的选择遵循的是与我们论文所述的截然不同的标准，因此它不能用作本文中所提出的该模型的无偏验证。

我们关于NH对模型进行了重新训练，最后一次血液样本的准确性从50%提高到了78%，而第一次血液样本的检测准确性从48%增加到了72%（入院）。

武汉金银滩医院和深圳市第三人民医院的新数据已验证了该模型（无重新训练），准确性分别为94%和90%。

在某些情况下，不需要进行重新训练，而在另一些情况下，重新训练可以大大提高模型的性能。”

生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步

nature

Article | Published: 23 December 2020

Mastering Atari, Go, chess and shogi by planning with a learned model

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap & David Silver

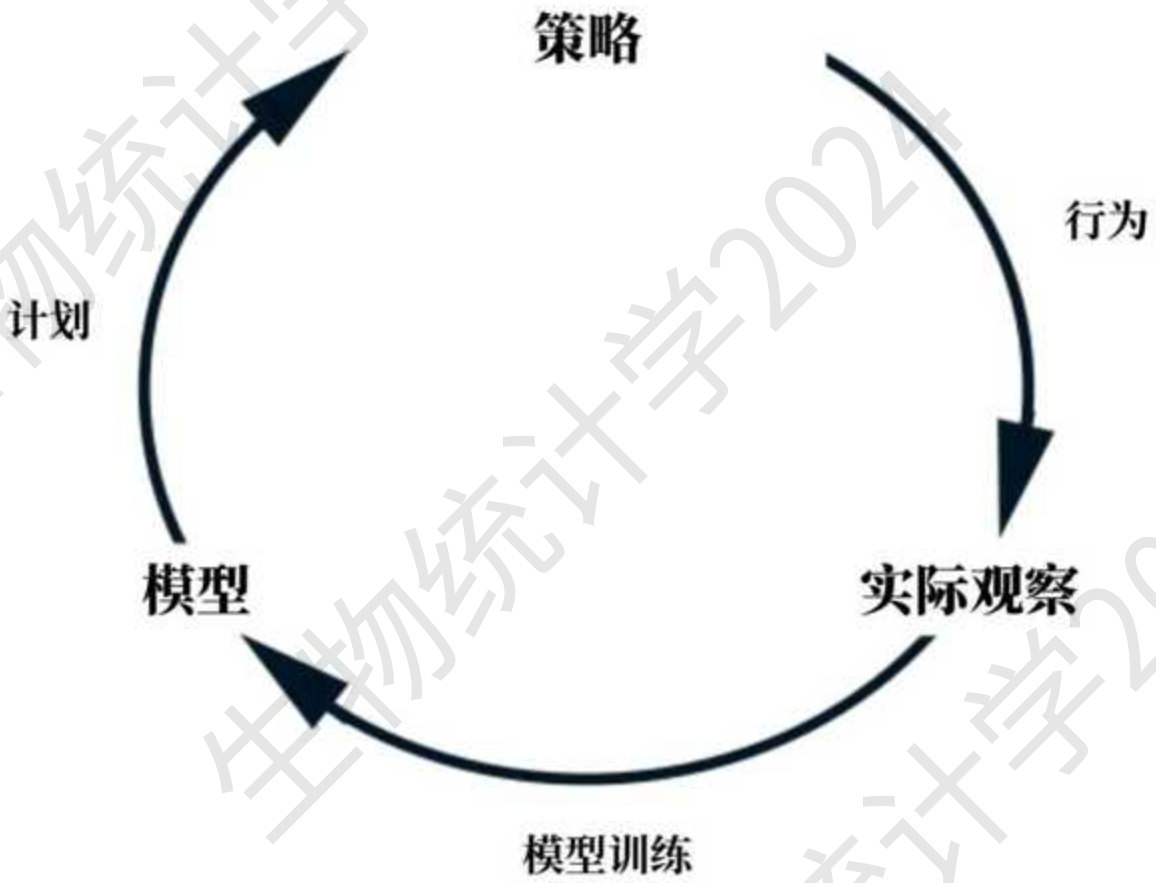
Nature **588**, 604–609(2020) | Cite this article

716 Altmetric | Metrics

生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步

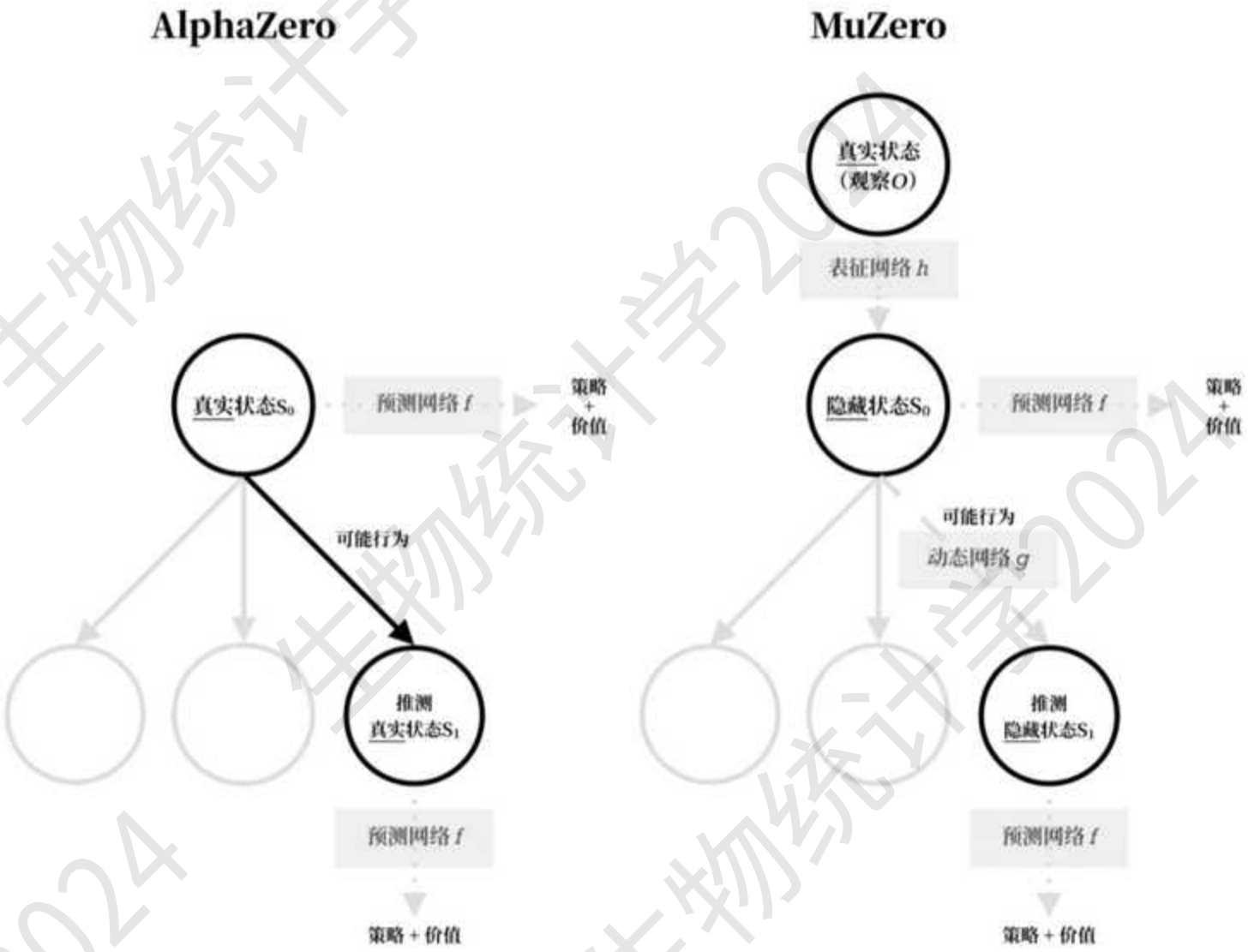
从AlphaGo的四世同堂看深度学习的方法进步



针对三个最重要的要素进行建模：即价值（The value）、策略（The policy）和奖励（The reward）

生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步



生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步

AlphaZero只需1个神经网络：

状态 s → 预测网络 f → 策略 p + 价值 v

MuZero缺乏规则信息，需要3个神经网络：

观察 o → 表征网络 h → 隐藏状态 s

状态 s → 预测网络 f → 策略 p + 价值 v

隐藏状态 s + 可能行为 a → 动态网络 g → 推测奖励 r + 推测隐藏状态 s

生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步

基于蒙特卡洛树状搜索和 MuZero 神经网络进行规划

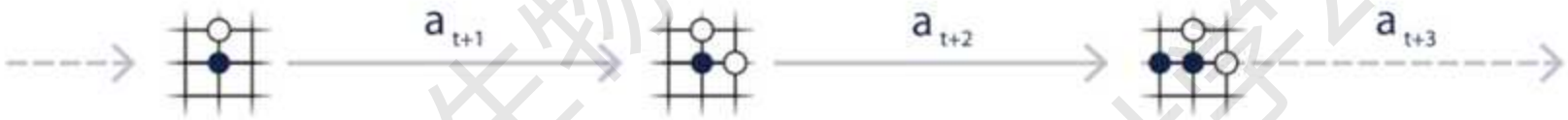
生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步

MuZero 通过模拟下棋走向训练神经网络

生物统计学2024

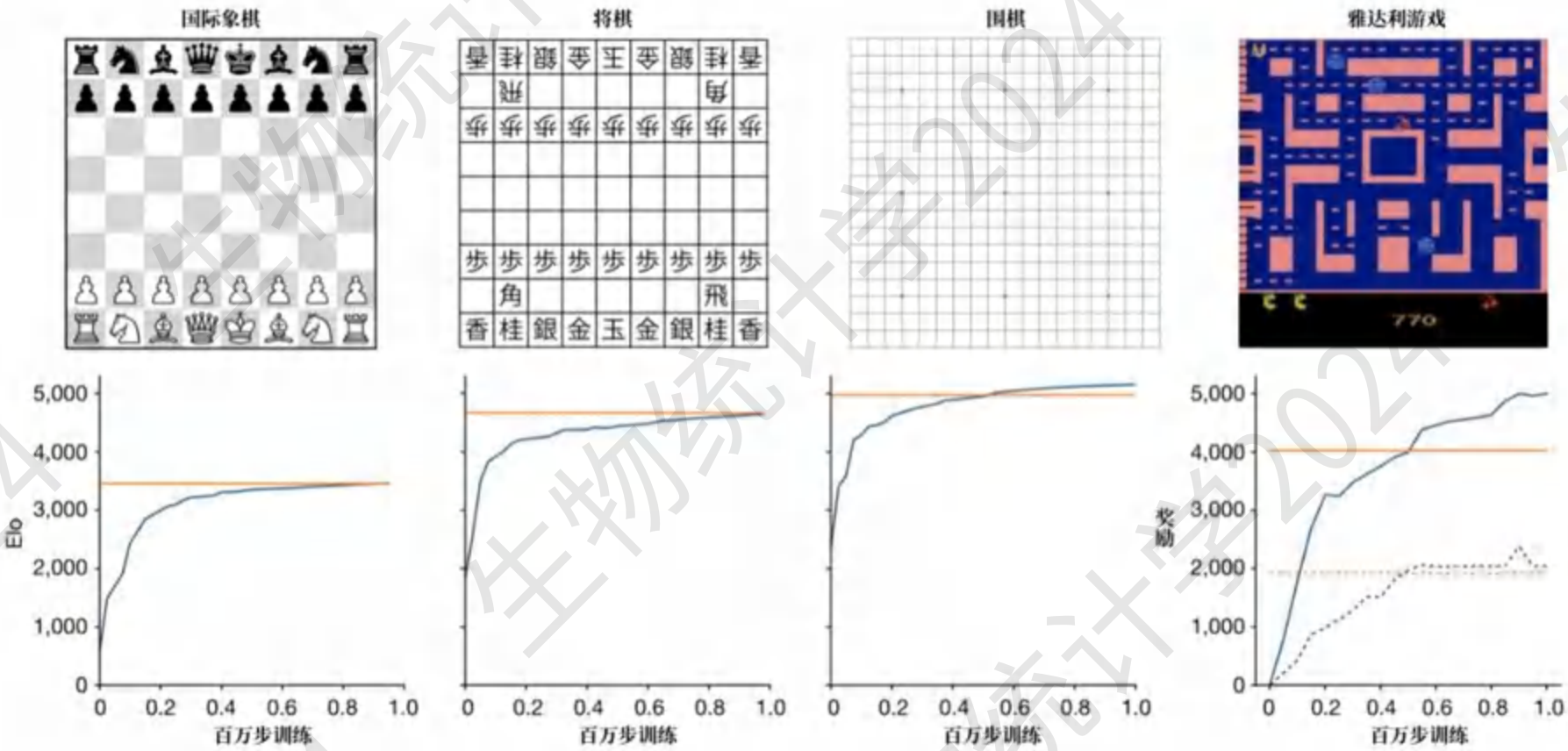
从AlphaGo的四世同堂看深度学习的方法进步



预测走法带来奖励

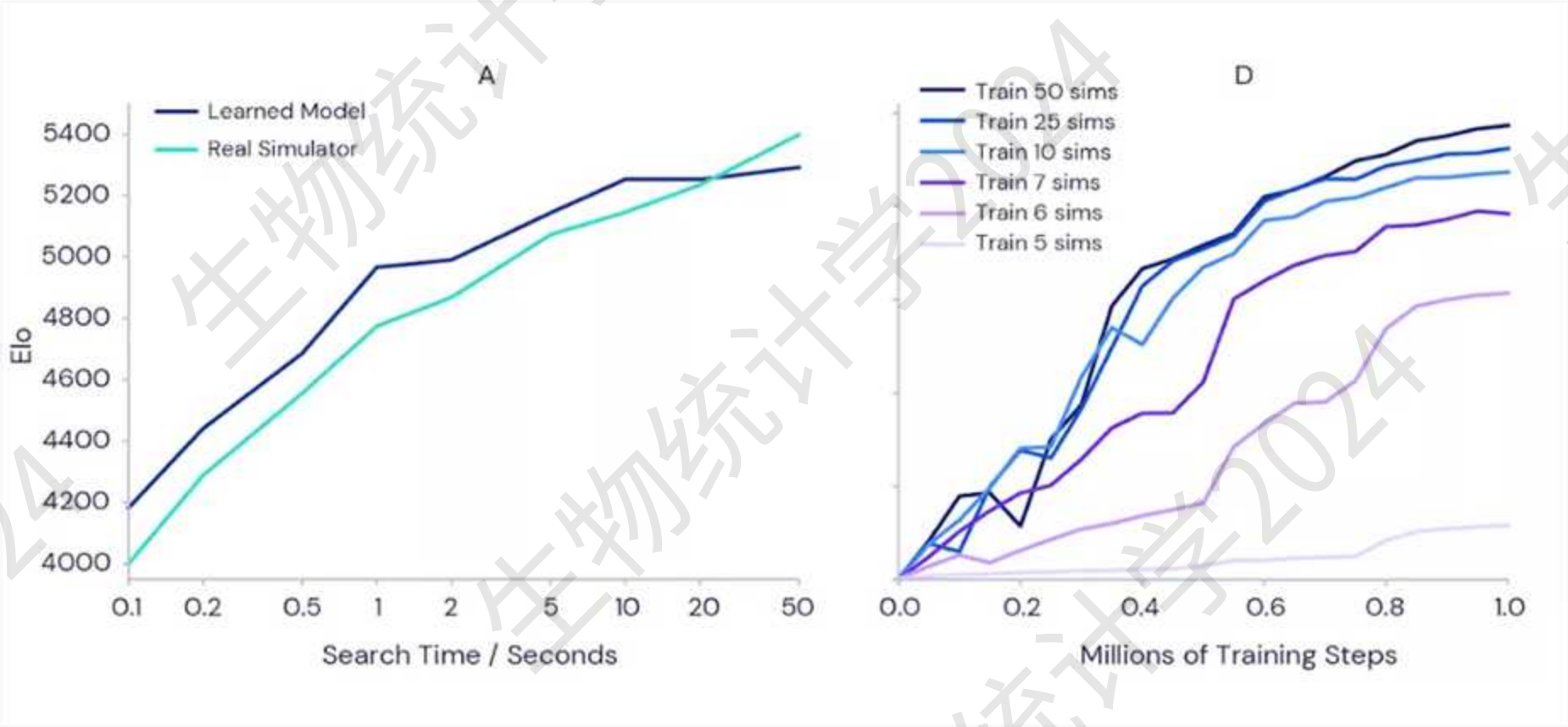
生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步



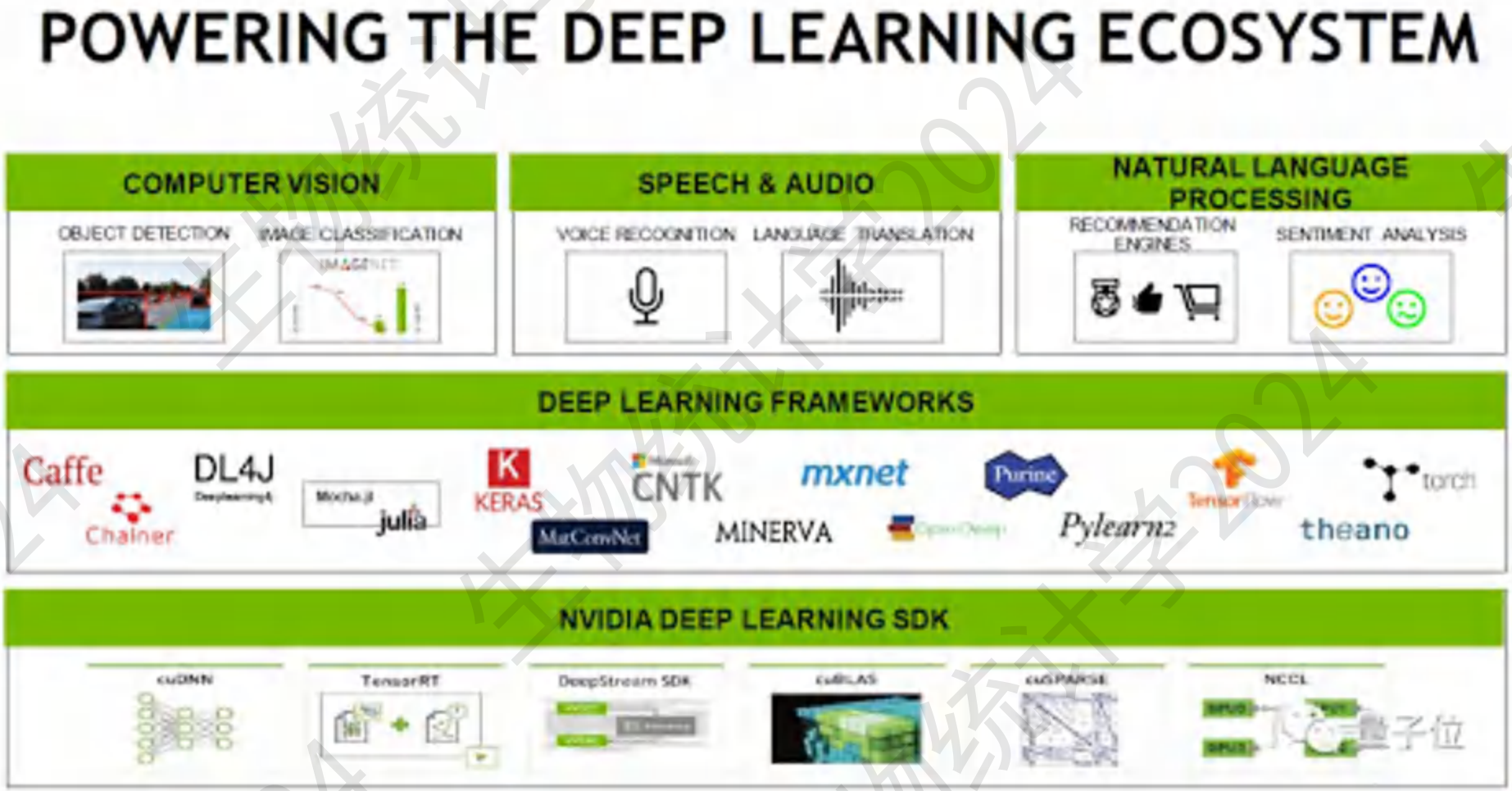
生物统计学2024

从AlphaGo的四世同堂看深度学习的方法进步



生物统计学2024

深度学习工具介绍



生物统计学2024

生物统计学2024

生物统计学2024

生物统计学2024