模式识别与统计学习

赵海涛

haitaozhao@ecust.edu.cn

大纲

- 什么是模式识别与统计学习
- 模式识别与统计学习的应用
- 研究什么问题
- 如何进行

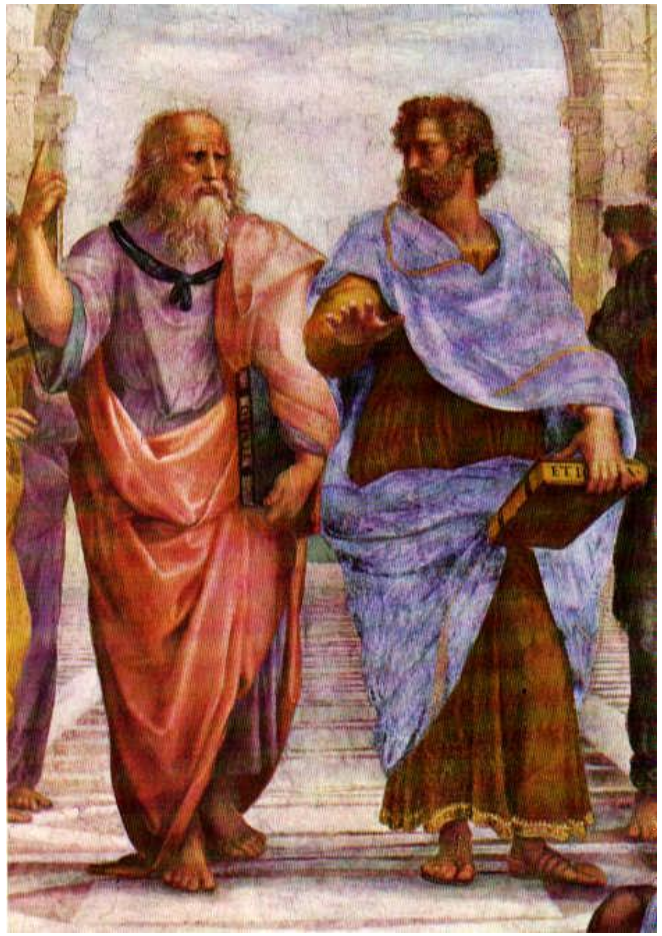
模式识别

"The problem of searching for patterns in data is a fundamental one and has a long and successful history."

模式识别

Western: Plato, Aristotle

They distinguished between an "*essential property*" from an "accidental property". Pattern recognition can be cast as the problem of finding such essential properties of a category.



模式识别

Laozi (6th Century BC)

道可道非常道 名可名非常名
无名天地之始 有名万物之母

模式识别

维基百科：

模式识别的定义：模式识别是通过使用计算机算法自动发现数据中的规律性，并利用这些规律性采取行动，例如将数据分为不同的类别。

A modern definition of pattern recognition is:

The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories.

Bishop, Christopher M. (2006). Pattern Recognition and Machine Learning. Springer.

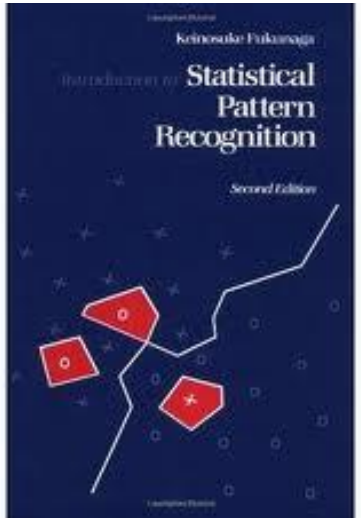
机器学习

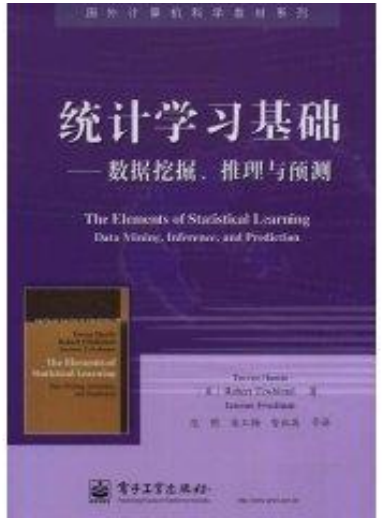
维基百科：

机器学习是研究通过经验自动改进的计算机算法。它被视为**人工智能**的一部分。机器学习算法基于样本数据建立一个模型，称为“训练数据”，进行预测或决策。

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions.

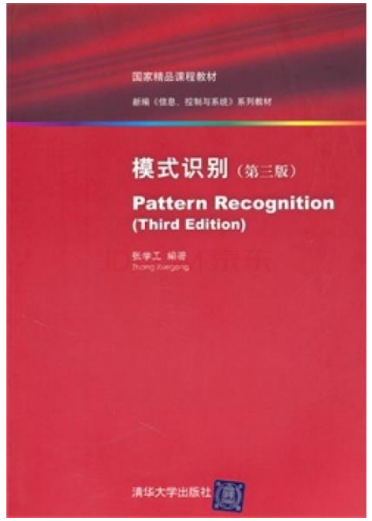
参考书





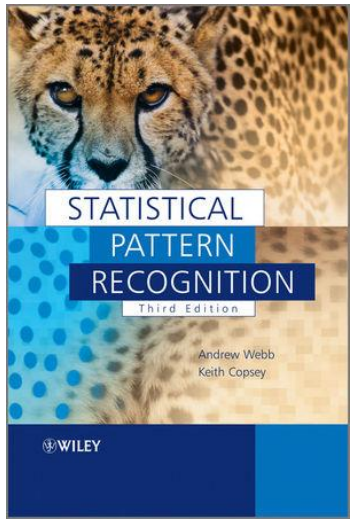
- 模式分类 (原书第二版). Duda等. 机械工业出版社. 2005.
- Introduction to Statistical Pattern Recognition (2nd Edition). Fukunaga. Academic Press. 1990.
- 统计学习基础–数据挖掘、推理与预测. Hastie等. 电子工业出版社. 2004.

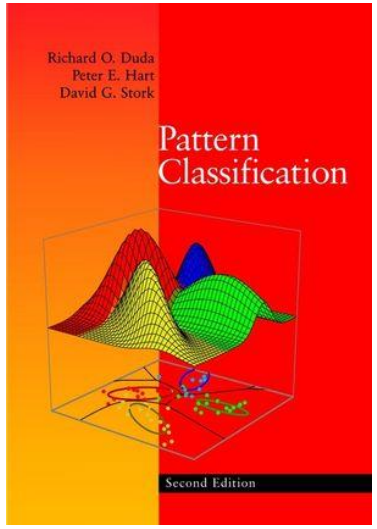
参考书

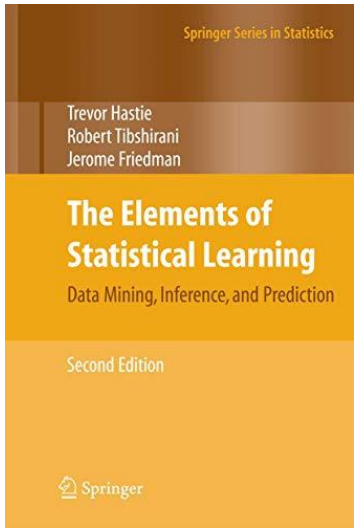


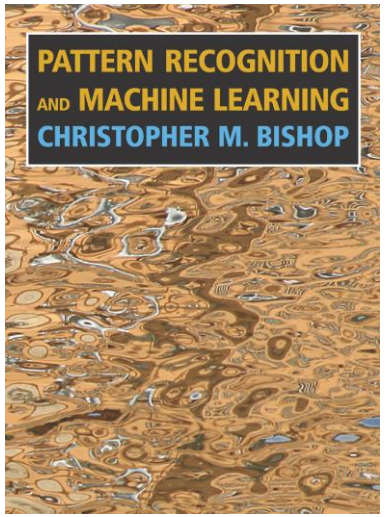
- 模式识别 (第二版). 边肇祺. 清华大学出版社. 2000.
- 模式识别 (第四版). Theodoridis. 电子工业出版社. 2010.
- 模式识别 (第三版). 张学工. 清华大学出版社. 2010.

Reference



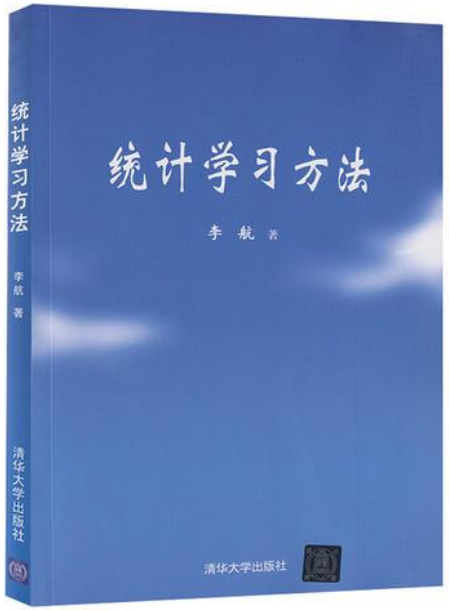






- Andrew R. Webb, Keith D. Copsey. Statistical Pattern Recognition, 3rd edition, Wiley, 2011.
- Richard O. Duda, Peter E. Hart, David G. Stock: Pattern Classification, 2nd edition, John Wiley & Sons, New York, 2001.
- Trevor Hastie, Robert Tobshiani, Jerome Friedman: The Elements of Statistical Learning - Data Mining, Inference, and Prediction, 2nd edition, Springer, New York, 2009.
- Christopher M. Bishop: Pattern Recognition and Machine Learning, Springer, New York, 2006.

教材



- 统计学习方法 李航. 清华大学出版社.

教材

模式识别与统计学习
统计学习方法概述
模式识别与统计学习介绍
统计学习三要素、监督与无监督
模型评估、正则化、泛化能力
分类问题、标注问题、回归问题
线性感知机
感知机模型介绍
感知机学习算法
k近邻法
K非参数估计
K近邻法模型介绍
构造kd树与搜索kd树
贝叶斯方法
条件概率公式与全概率公式
贝叶斯决策与朴素贝叶斯
最大似然估计，最大后验概率估计
贝叶斯估计
决策树
决策树模型 ID3
CART模型
随机森林
Logistic回归
Logistic回归模型
Softmax 模型
模型学习的最优化算法
支持向量机
线性可分支持向量机算法
线性不可分支持向量机算法
非线性支持向量机与核函数
AdaBoost算法
AdaBoost算法
前向分布算法与AdaBoost
提升树与GBDT
EM算法
EM算法
EM算法的收敛性
聚类算法
相似度与距离
层次聚类
K均值聚类
奇异值分解算法
奇异值分解定理
奇异值分解的性质
奇异值分解的应用
主成分分析算法
主成分分析的基本思想
主成分分析的推导
主成分分析的扩展
潜在语义分析算法
潜在语义分析简介
基于SVD分解的潜语义分析
概率潜在语义分析
概率潜在语义分析简介
概率潜在语义分析推导
马尔可夫链与蒙特卡洛方法
有限马尔可夫链
蒙特卡洛方法、Gibbs抽样
PageRank算法
PageRank算法简介
PageRank算法的推导与编程

软件

- Python

Spyder (Python 3.7)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\haitao

C:\Users\haitao\pr_class2020\softmax_iris_batch.py

softmax_iris_batch.py | least_squares_iris.py | perceptron_iris.py

```python
# -*- coding: utf-8 -*-
"""
Created on Tue Oct 20 14:01:05 2020

@author: haitao
"""

from sklearn import datasets
from sklearn.model_selection import train_test_split
import numpy as np
import matplotlib.pyplot as plt

def mysoft(x):
    y = np.exp(x)
    y = y/np.sum(y,axis=0)
    return y
def myloss(p,y):
    loss = - sum(y*np.log(p))
    return loss


# 读入 iris数据
iris_df = datasets.load_iris()
x = iris_df.data
x = x/np.max(x)
y = iris_df.target




# 划分训练集和测试集
train_X,test_X, train_y, test_y = train_test_split(x,
                                                   y,
                                                   test_size = 0.2,
                                                   random_state = 0)

# encode y into onehot format
nb_classes = 3
targets = train_y.reshape(-1)
y_onehot= np.eye(nb_classes)[targets]

# 初始化w
```

Source Console Object

Usage

Here you can get help of any object by pressing **Ctrl+I** in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in *Preferences > Help*.

New to Spyder? Read our tutorial

Variable explorer Help Plots Files

Console 1/A

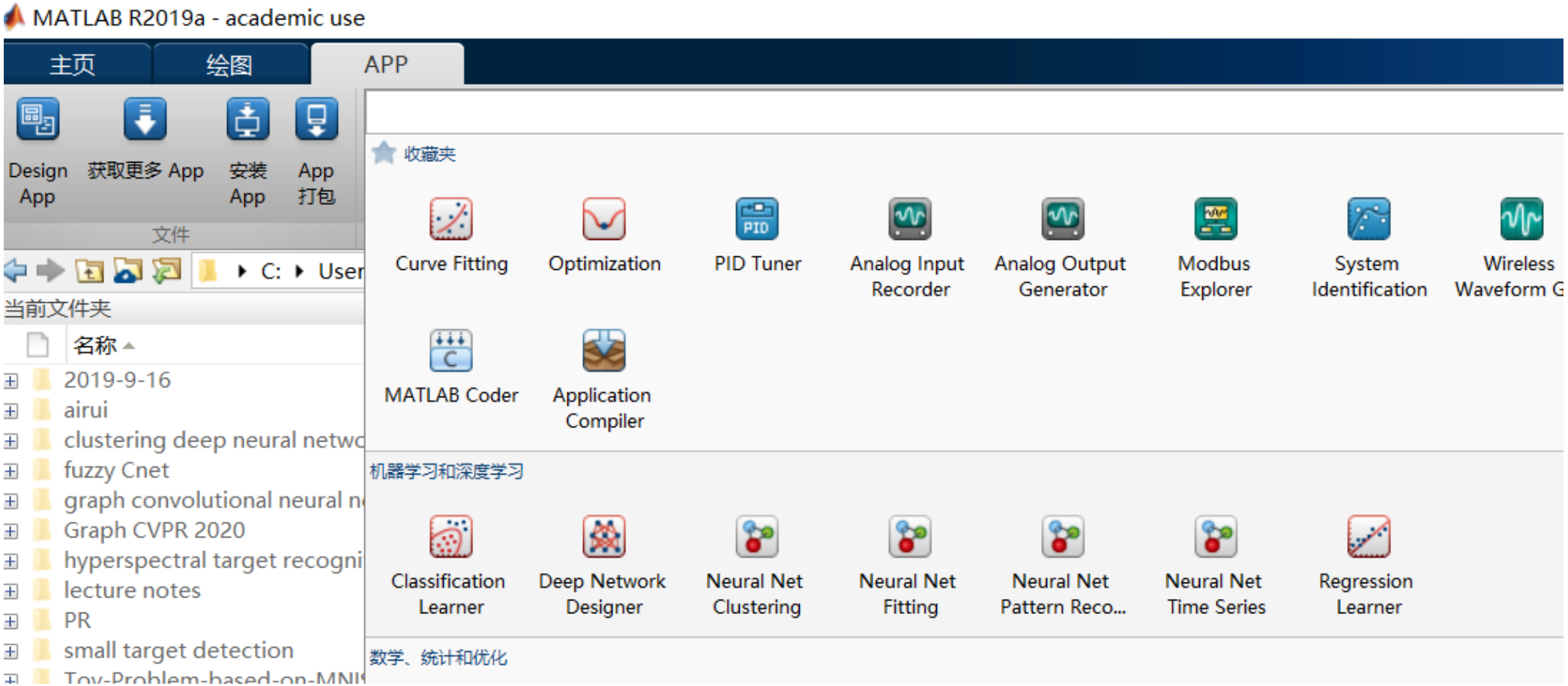
```
Python 3.7.3 (default, Mar 27 2019, 17:13:21) [MSC v.1915 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 7.4.0 -- An enhanced Interactive Python.

In [1]:
```

IPython console History

LSP Python: ready | conda: base (Python 3.7.3) | Line 60, Col 1 | UTF-8 | CRLF | RW | Mem 50%

软件

- Matlab
  - can be downloaded from our university



GitHub

github.com/haitaozhao/Python-Examples-for-Pattern-Recognition

haitaozhao / Python-Examples-for-Pattern-Recognition Public

Unpin | Unwatch 4 | Fork 9 | Star 16

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

master | 1 branch | 0 tags | Go to file | Add file | Code

haitaozhao revise dir — efc8286 on 23 Sep 2019 — 46 commits

-1 Toy problem - Predicting the c...	upload codes for car insurance prediction problem	4 years ago
-2 Toy problem - Predcting Cat o...	upload codes of Cat and Dog Problem	3 years ago
-3 Toy problem - Predcting Ham ...	upload new code (ham and spam)	3 years ago
0 Toy problem - Predicting the S...	upload matlab code	3 years ago
1 Least squares - noniterative sol...	revise dir	2 years ago
2 Least squares - with gradient d...	revise dir	2 years ago
3 Fully connected neural network...	upload fcn	3 years ago
4 Example of PCA for Dimension ...	upload example of PCA	3 years ago
5 Logistic Regression on iris data...	Update readme.md	3 years ago

About

Python codes for my course (Pattern Recognition)

Readme
16 stars
4 watching
9 forks

Releases

No releases published
Create a new release

Packages

No packages published

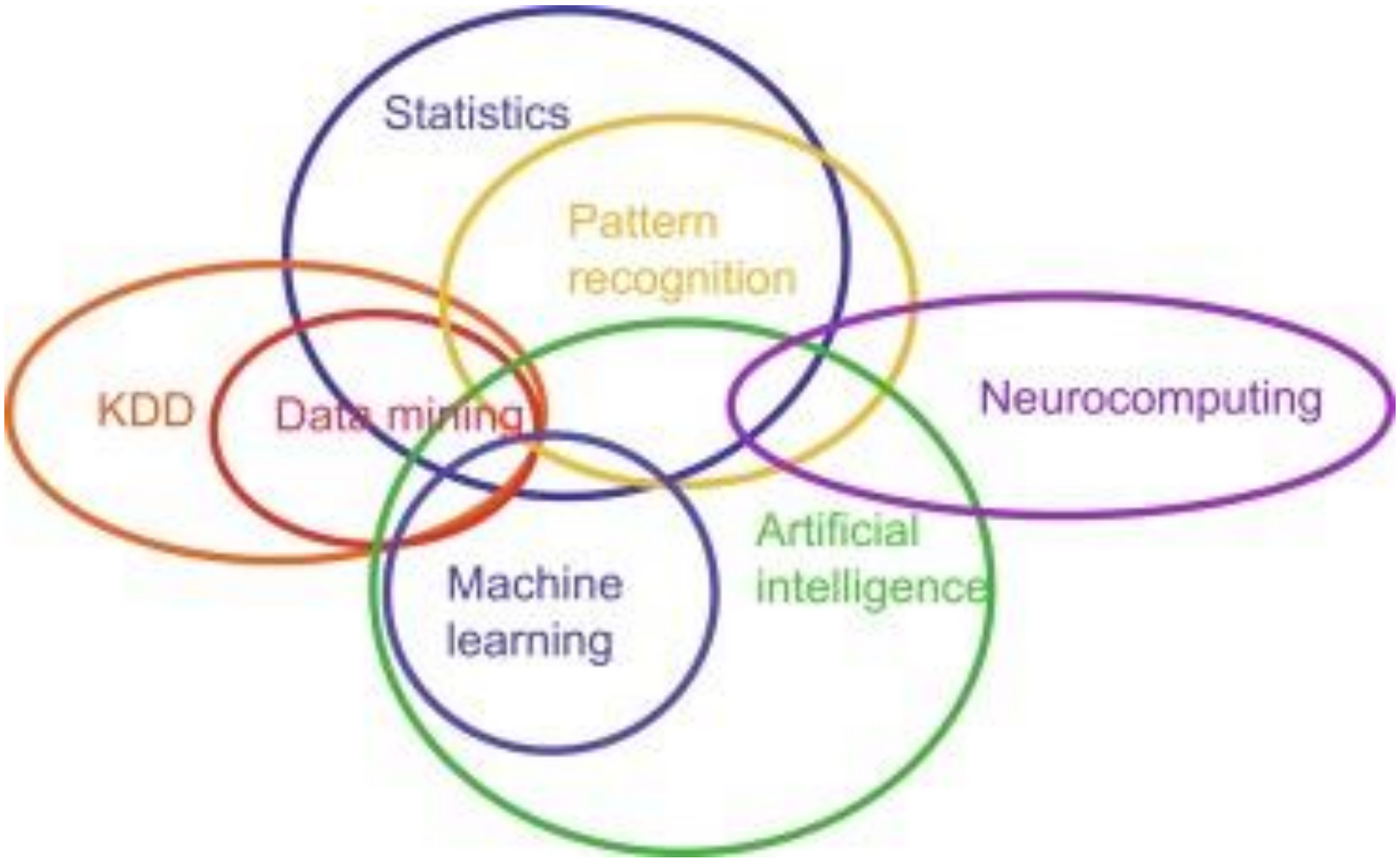
模式识别与机器学习、数据挖掘、人工智能的关系

人工智能

维基百科：

Intelligence has been defined in many ways: the capacity for logic, understanding, self-awareness, learning, emotional knowledge, reasoning, planning, creativity, critical thinking, and problem-solving. More generally, it can be described as the ability to perceive or infer information, and to retain it as knowledge to be applied towards adaptive behaviors within an environment or context.

Intelligence in computers or other machines is called artificial intelligence.


Statistics
Pattern recognition
KDD
Data mining
Neurocomputing
Artificial intelligence
Machine learning

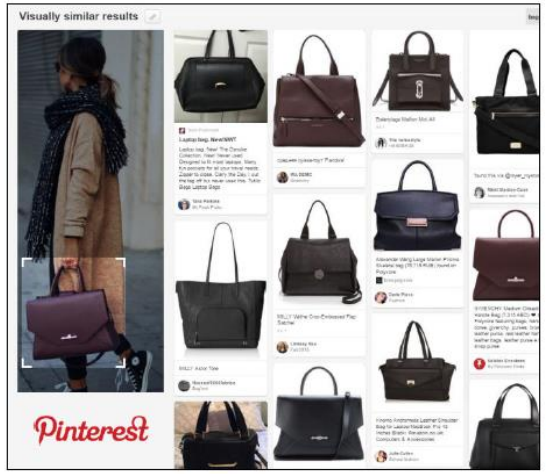
大纲

- 什么是模式识别与统计学习
- 模式识别与统计学习的应用
- 研究什么问题
- 如何进行

模式识别的应用

Healthcare
Automobile
Finance
Surveillance
Social Media
Entertainment
Education
Space Exploration
Gaming
Robotics
Agriculture
E-commerce

模式识别的应用



Increasing public safety with smart video surveillance at airports & malls

Providing intelligent services in hotels, banks and stores

Separating weeds as it harvests, reduces chemical usage by 90%

模式识别的应用

FUELING ALL INDUSTRIES

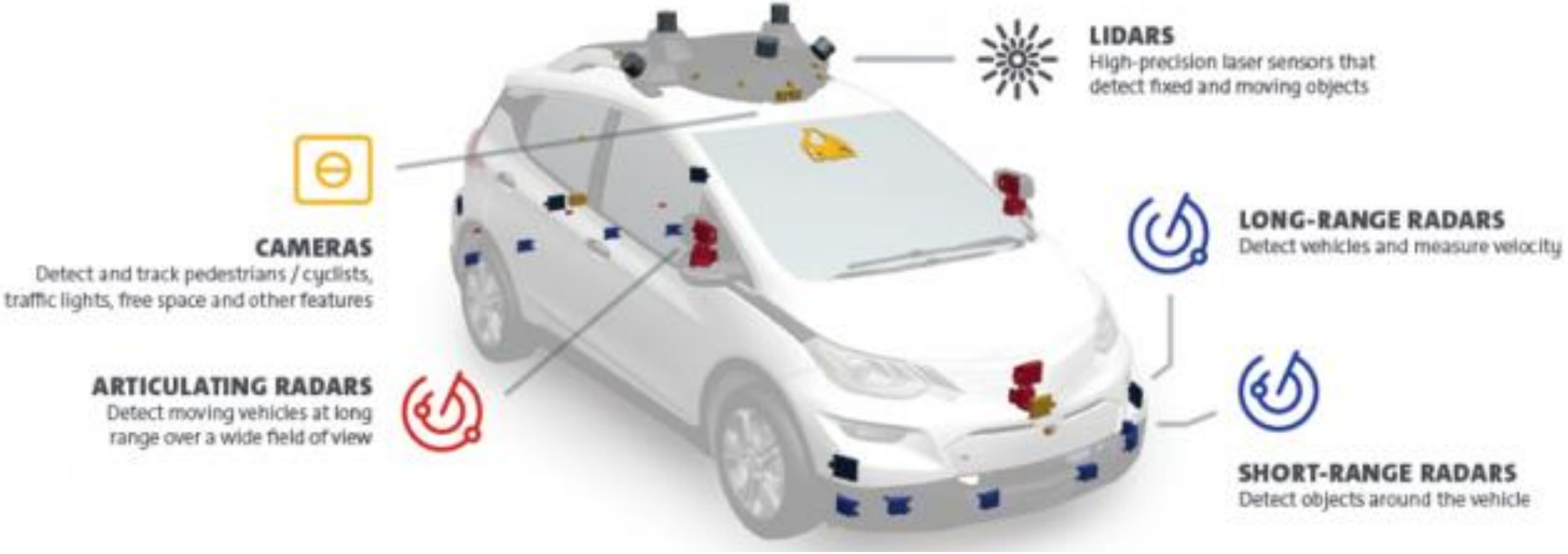
Increasing public safety with smart video surveillance at airports & malls

Providing intelligent services in hotels, banks and stores

Separating weeds as it harvests, reduces chemical usage by 90%

模式识别的应用

模式识别的应用


LIDARS
High-precision laser sensors that detect fixed and moving objects
CAMERAS
Detect and track pedestrians / cyclists, traffic lights, free space and other features
LONG-RANGE RADARS
Detect vehicles and measure velocity
ARTICULATING RADARS
Detect moving vehicles at long range over a wide field of view
SHORT-RANGE RADARS
Detect objects around the vehicle

厉害了人工智能

中国人工智能产业规模爆发式增长，发展潜力大

艾媒咨询 iiMedia Research

数据来源：工信部，艾媒数据中心（data.iimedia.cn）

艾媒报告中心：report.iimedia.cn ©2020 iiMedia Research Inc

厉害了人工智能

行业在快速发展之际，行业人才缺口数量也是很惊人的，相关数据显示2020年国内人工智能人才缺口达到了500多万，整个行业人才缺口很大，有的企业都是“抢人才”。我国人工智能缺口大，人工智能人才供不应求，所以现在不少高校都在积极培养相关行业人才。

行人才稀缺的情况下从业者的薪资也很喜人，根据《2021年互联网产业求职指南》中显示，2021年人工智能人才平均工资19010元/月，远高于互联网产业产品技术核心岗位，这样的薪资待遇你们心动吗?

这些年中不少人都关注到人工智能行业的发展，不少高校都开通了相关专业，整个行业的技术人才数量飙升，不少学生以及家长都在考虑这一行业。可以说人工智能行业已经进入到了正轨，2018年35所高校获教育部批准首批开设人工智能本科专业，2019、2020年新增人工智能专业的高校分别有180所、130所，社会对于这一行业的关注度也提高了许多。

中共中央政治局在2018年10月31日下午就人工智能发展现状和趋势举行第九次集体学习。

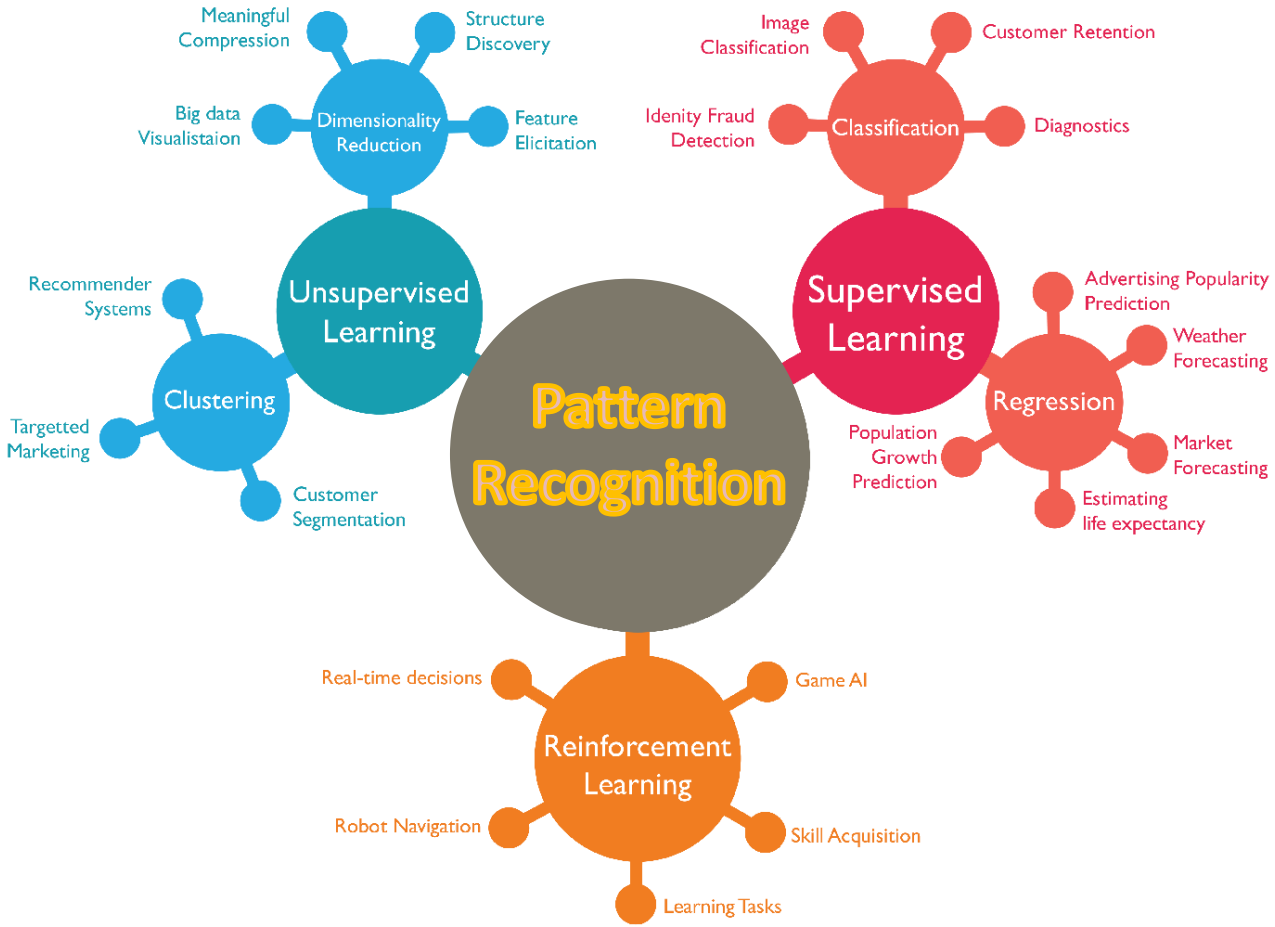
中共中央总书记习近平在主持学习时强调，人工智能是**赢得全球科技竞争主动权的重要战略抓手。**

大纲

- 什么是模式识别与统计学习
- 模式识别与统计学习的应用
- **研究什么问题**
- 如何进行

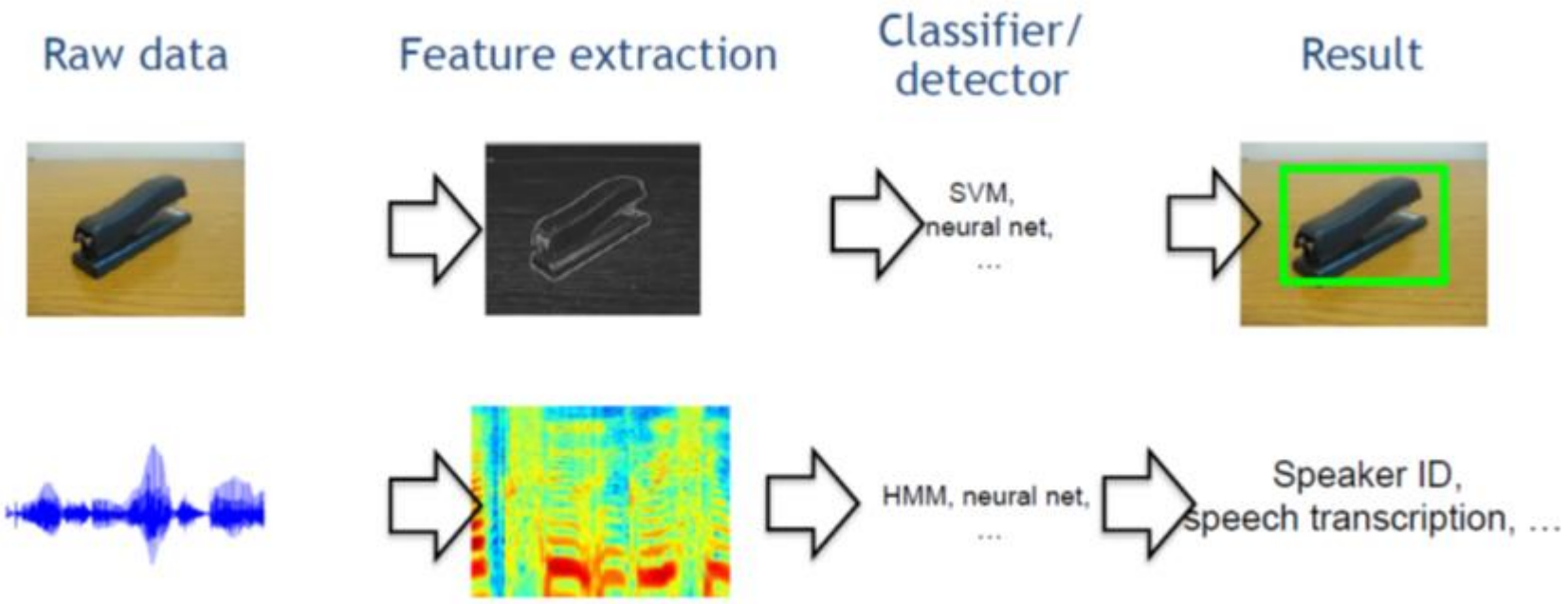
模式识别与统计学习方法的分类

- 有监督学习
- 无监督学习
- 半监督学习
- 强化学习



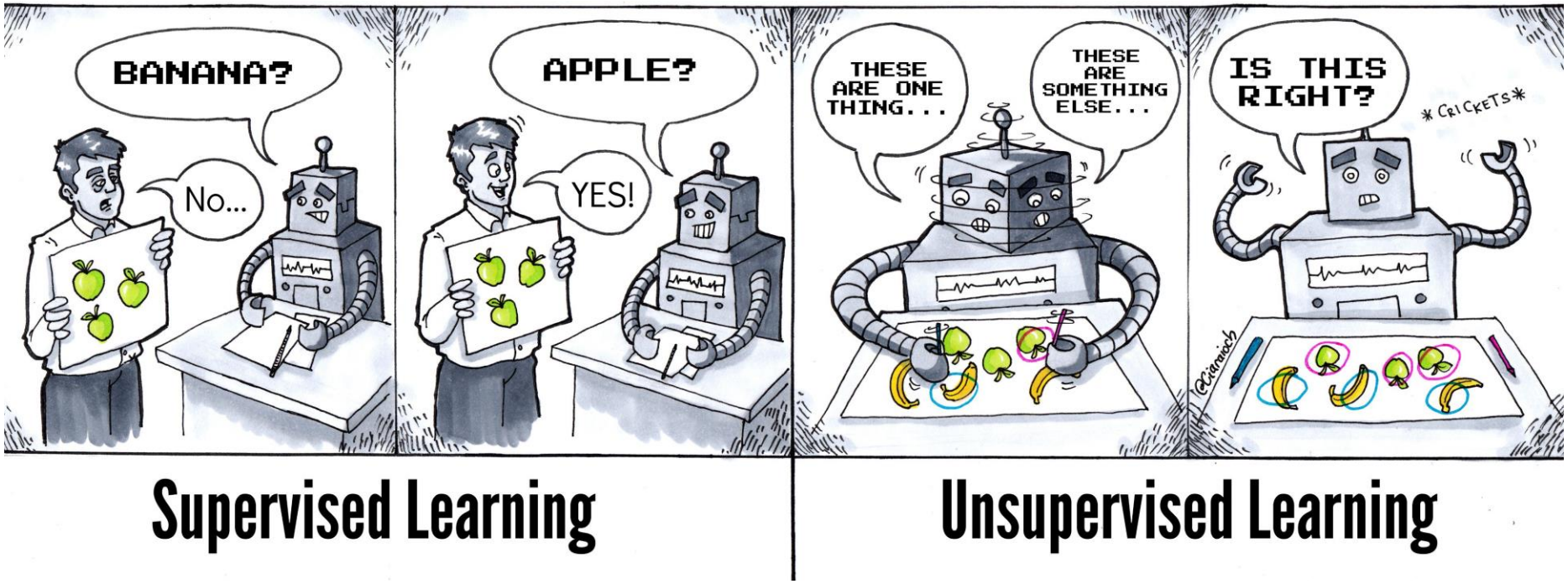
- 有监督学习

"**Supervised learning** is the task of learning a function that maps an input to an output based on example input-output pairs."

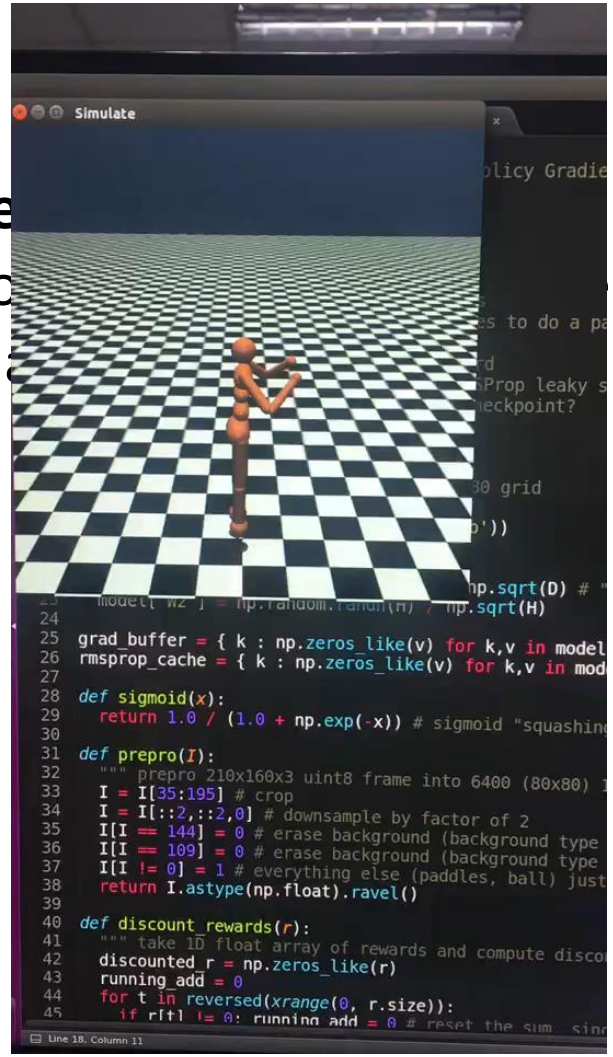


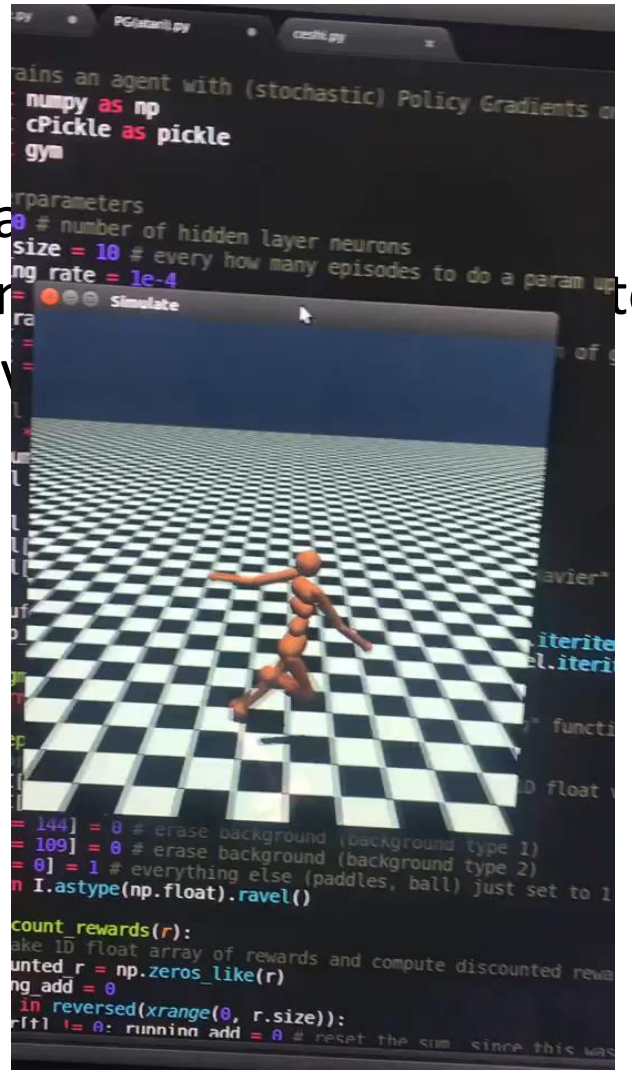
- 无监督学习

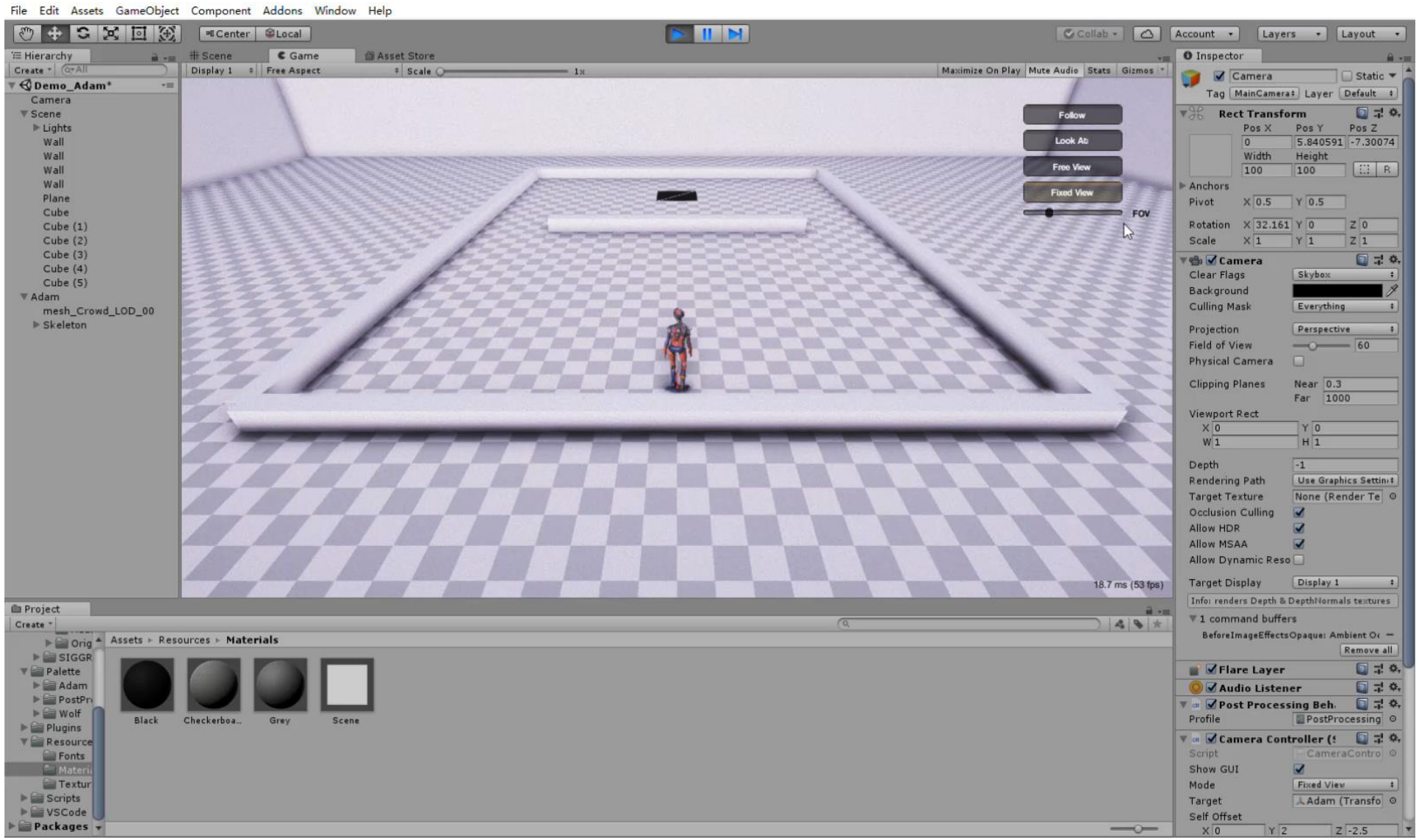
“**Unsupervised machine learning** is the task of inferring a function that describes the structure of "unlabeled" data (i.e. data that has not been classified or categorized).”



- **Re** **RL**) is an a
ho *actions* in to
ma cumulativ






File Edit Assets GameObject Component Addons Window Help
Center
Local
Collab
Account
Layers
Layout
Hierarchy
Create
Demo_Adam*
Camera
Scene
Lights
Wall
Wall
Wall
Wall
Plane
Cube
Cube (1)
Cube (2)
Cube (3)
Cube (4)
Cube (5)
Adam
mesh_Crowd_LOD_00
Skeleton
Scene
Game
Asset Store
Display 1
Free Aspect
Scale
1x
Maximize On Play
Mute Audio
Stats
Gizmos
Follow
Look At
Free View
Fixed View
FOV
18.7 ms (53 fps)
Project
Create
Assets ▸ Resources ▸ Materials
Black
Checkerboa...
Grey
Scene
Orig
SIGGR
Palette
Adam
PostPr
Wolf
Plugins
Resource
Fonts
Materi
Textur
Scripts
VSCode
Packages
Inspector
Camera
Static
Tag MainCamera
Layer Default
Rect Transform
Pos X 0
Pos Y 5.840591
Pos Z -7.30074
Width 100
Height 100
Anchors
Pivot X 0.5 Y 0.5
Rotation X 32.161 Y 0 Z 0
Scale X 1 Y 1 Z 1
Camera
Clear Flags Skybox
Background
Culling Mask Everything
Projection Perspective
Field of View 60
Physical Camera
Clipping Planes Near 0.3 Far 1000
Viewport Rect
X 0 Y 0
W 1 H 1
Depth -1
Rendering Path Use Graphics Settin
Target Texture None (Render Te
Occlusion Culling
Allow HDR
Allow MSAA
Allow Dynamic Reso
Target Display Display 1
Info: renders Depth & DepthNormals textures
1 command buffers
BeforeImageEffectsOpaque: Ambient Oc
Remove all
Flare Layer
Audio Listener
Post Processing Beh.
Profile PostProcessing
Camera Controller (S
Script CameraContro
Show GUI
Mode Fixed View
Target Adam (Transfo
Self Offset
X 0 Y 2 Z -2.5

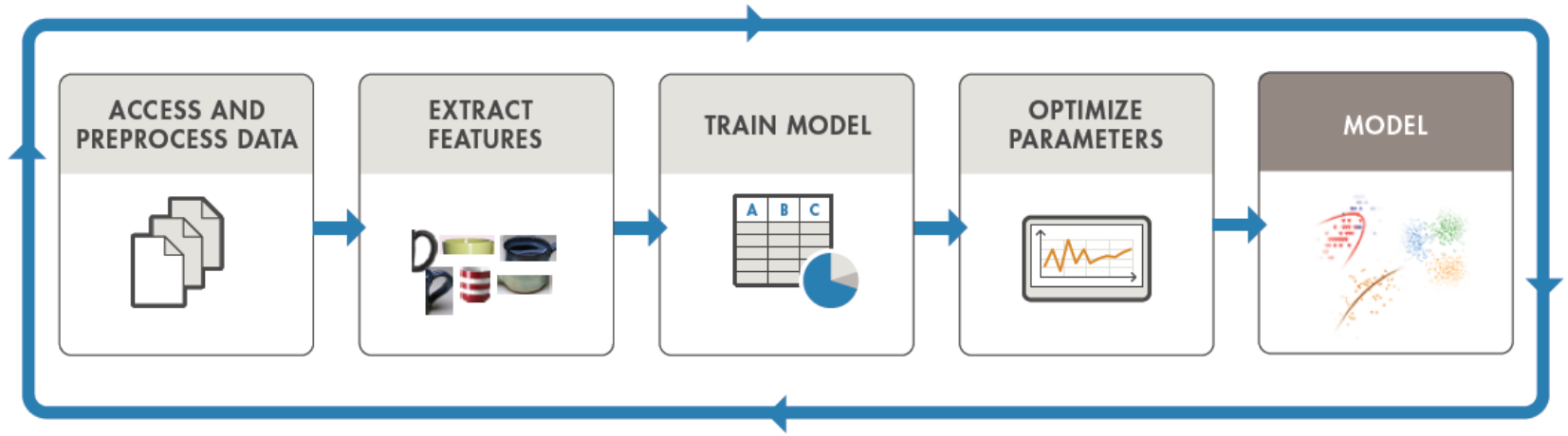
- 强化学习

GOOGLE DEEPMIND ALPHAGO CHALLENGE

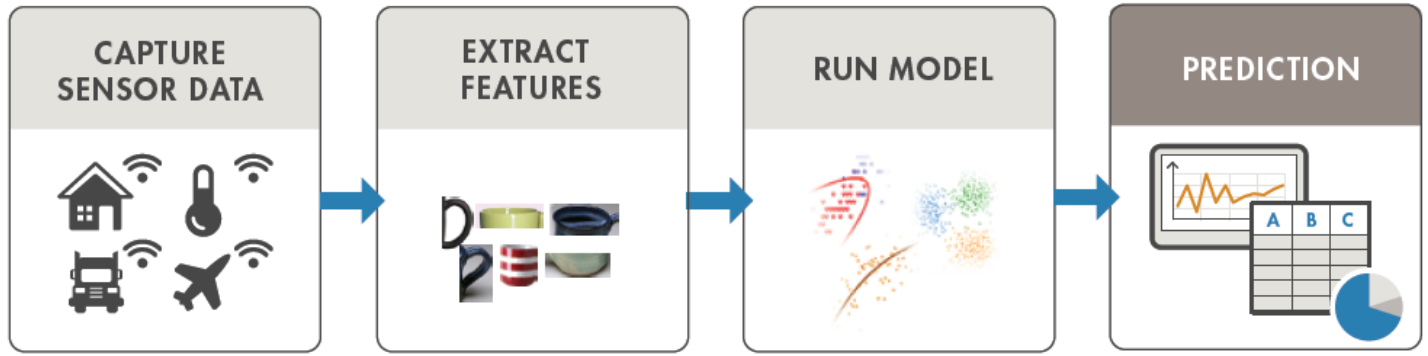
大纲

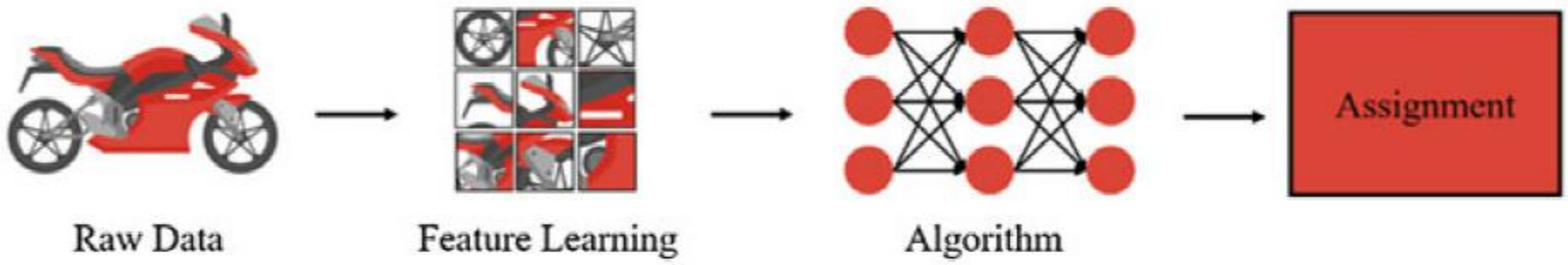
- 什么是模式识别与统计学习
- 模式识别与统计学习的应用
- 研究什么问题
- 如何进行

训练：迭代进行，直到取得令人满意的性能。

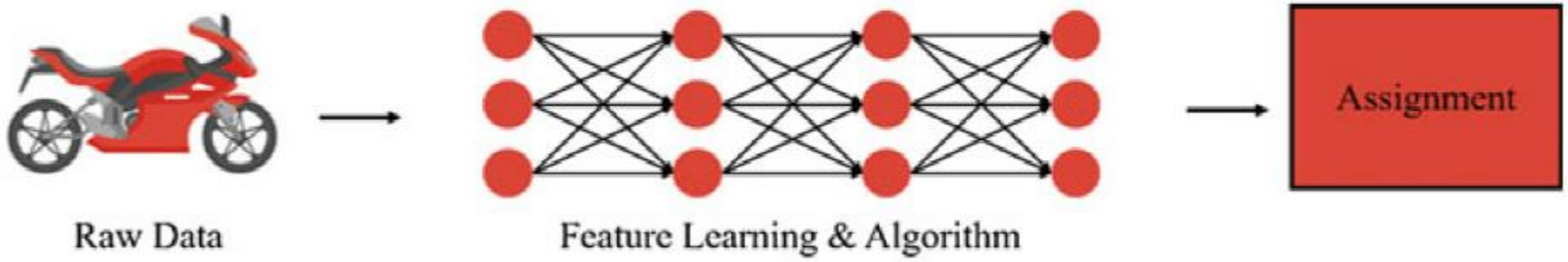


预测：将训练好的模型集成到应用程序中。




Raw Data
Feature Learning
Algorithm
Assignment

(a)


Raw Data
Feature Learning & Algorithm
Assignment

(b)

- 访问和预处理数据
- 提取特征
- 训练模型
- 优化模型
- 应用模型

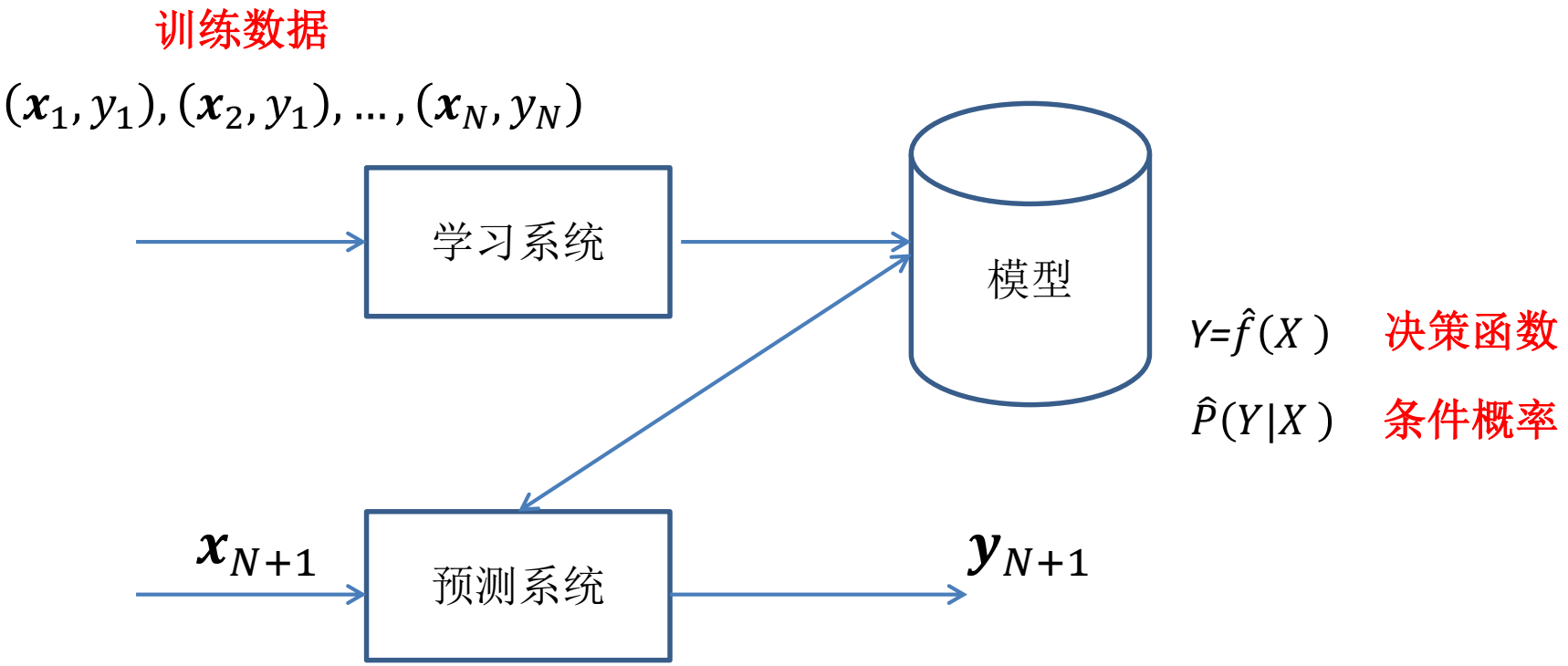
统计学习三要素

- 模型
- 准则
- 优化

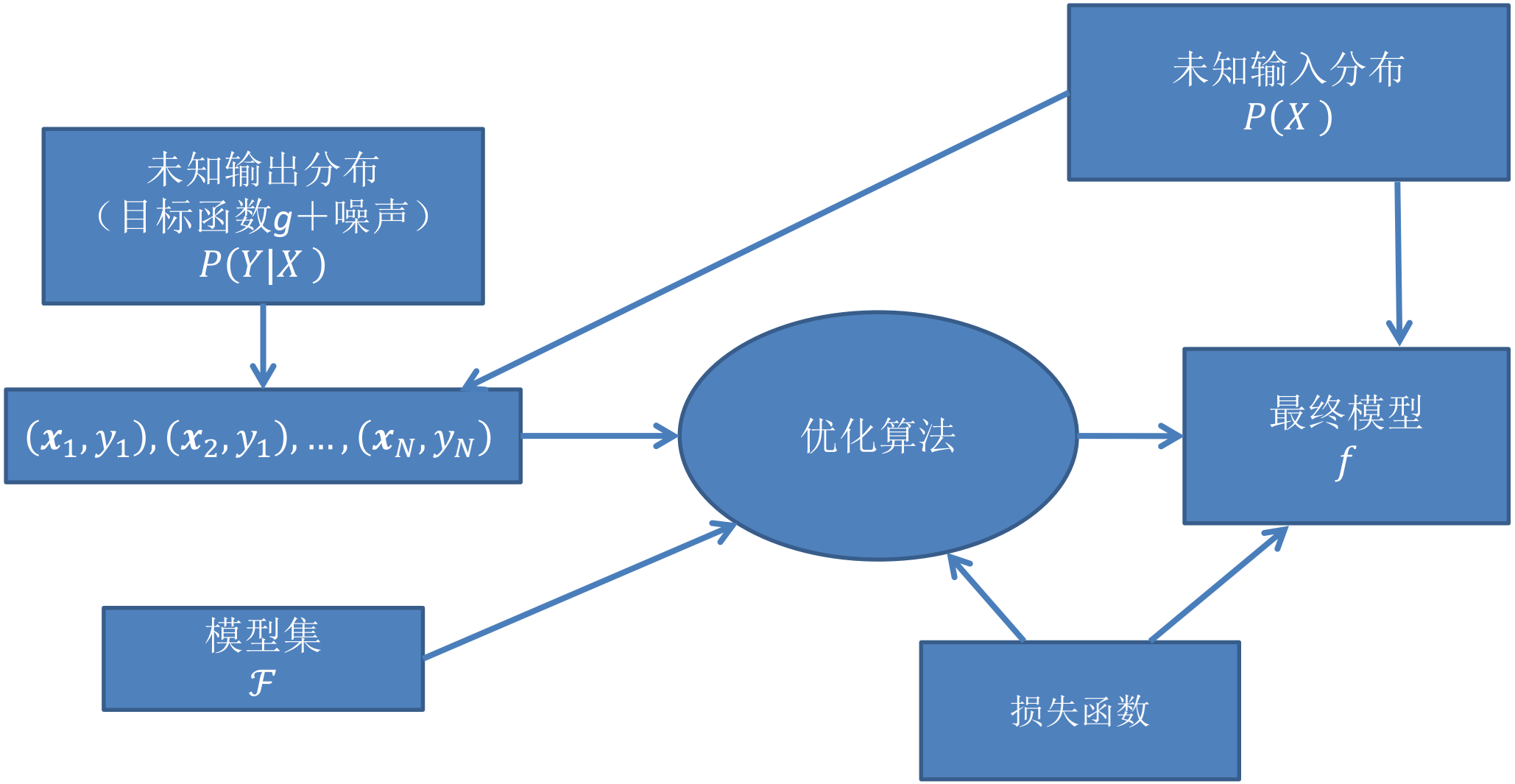
```
# Use the nn package to define our model and loss function.
model = torch.nn.Sequential(
    torch.nn.Linear(D_in, H),
    torch.nn.ReLU(),
    torch.nn.Linear(H, D_out),
)
loss_fn = torch.nn.MSELoss(reduction='sum')

# Use the optim package to define an Optimizer that will update the weights of
# the model for us. Here we will use Adam; the optim package contains many other
# optimization algorithms. The first argument to the Adam constructor tells the
# optimizer which Tensors it should update.
learning_rate = 1e-4
optimizer = torch.optim.Adam(model.parameters(), lr=learning_rate)
```

```
    print(t, loss.item())

    # Before the backward pass, use the optimizer object to zero all of the
    # gradients for the variables it will update (which are the learnable
    # weights of the model). This is because by default, gradients are
    # accumulated in buffers( i.e, not overwritten) whenever .backward()
    # is called. Checkout docs of torch.autograd.backward for more details.
    optimizer.zero_grad()

    # Backward pass: compute gradient of the loss with respect to model
    # parameters
    loss.backward()

    # Calling the step function on an Optimizer makes an update to its
    # parameters
    optimizer.step()
```

以监督学习为例


训练数据
$(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_1), \ldots, (\boldsymbol{x}_N, y_N)$
学习系统
模型
$Y=\hat{f}(X)$ 决策函数
$\hat{P}(Y|X)$ 条件概率
$\boldsymbol{x}_{N+1}$
预测系统
$\boldsymbol{y}_{N+1}$

以监督学习为例


未知输入分布
$P(X)$
未知输出分布
（目标函数g＋噪声）
$P(Y|X)$
$(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_1), \ldots, (\boldsymbol{x}_N, y_N)$
优化算法
最终模型
f
模型集
$\mathcal{F}$
损失函数

模型：

决策函数的集合： $\mathcal{F} = \{f \mid Y = f(X)\}$

参数空间 $\mathcal{F} = \{f \mid Y = f_\theta(X), \theta \in \boldsymbol{R}^n\}$

条件概率的集合： $\mathcal{F} = \{P \mid P(Y|X)\}$

参数空间 $\mathcal{F} = \{P \mid P_\theta(Y|X), \theta \in \boldsymbol{R}^n\}$

准则：

损失函数：一次预测的好坏
风险函数：平均意义下模型预测的好坏

0-1损失函数 0-1 loss function

$$L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$$

平方损失函数 quadratic loss function

$$L(Y, f(X)) = (Y - f(X))^2$$

绝对损失函数 absolute loss function

$$L(Y, f(X)) = |Y - f(X)|$$

对数损失函数 logarithmic loss function
（或对数似然损失函数 loglikelihood loss function）

$$L(Y, P(Y|X)) = -logP(Y|X)$$

损失函数的期望

$$R_{\exp}(f) = E_P[L(Y, f(X))] = \int_{\mathcal{X}\times y} L(y, f(\boldsymbol{x}))P(\boldsymbol{x}, y)\mathrm{d}\boldsymbol{x}\mathrm{d}y$$

经验风险 empirical risk

$$R_{emp}(f) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(\boldsymbol{x}_i))$$

算法：

- 解析解法，矩阵分解等
- 梯度下降法
- 凸优化方法

模型评估与模型选择

训练误差，训练数据集的平均损失

$$R_{emp}(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}(\boldsymbol{x}_i))$$

测试误差，测试数据集的平均损失

$$e_{test} = \frac{1}{N'}\sum_{i=1}^{N'} L(y_i, \hat{f}(\boldsymbol{x}_i))$$

损失函数是0-1 损失时

$$e_{test} = \frac{1}{N'}\sum_{i=1}^{N'} I(y_i \neq \hat{f}(\boldsymbol{x}_i))$$

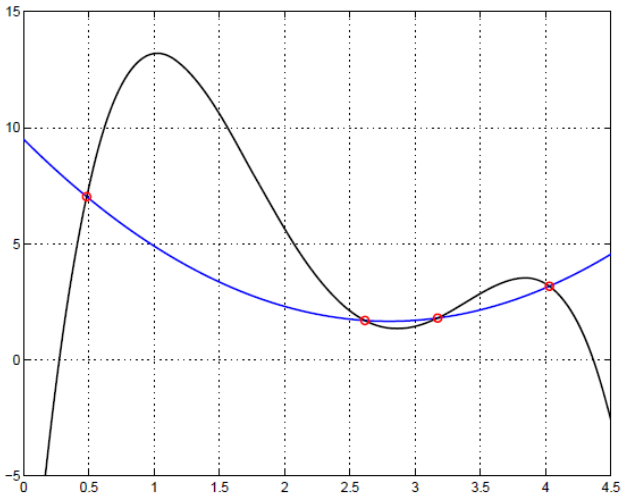
测试数据集的准确率

$$r_{test} = \frac{1}{N'}\sum_{i=1}^{N'} I(y_i = \hat{f}(\boldsymbol{x}_i))$$

泛化能力 generalization ability

假设空间有无限个函数？

经验风险都为0？



经验风险最小化函数： $f_N = \arg\min_{f \in \mathcal{F}} \hat{R}(f)$

泛化能力： $R(f_N) = E[L(Y, f_N(X))]$

定理：泛化误差上界，二分类问题，

当假设空间是有限个函数的结合 $\mathcal{F} = \{f_1, f_2, \dots, f_d\}$，

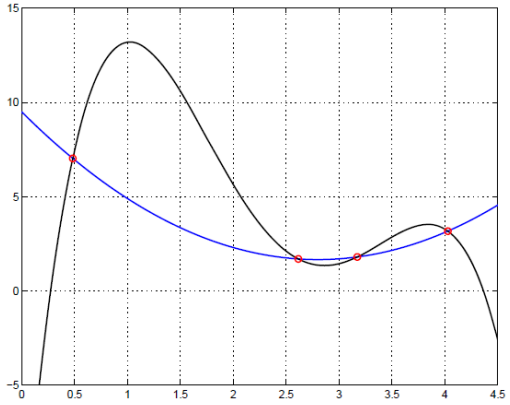
对任意一个函数f， 至少以概率$1-\delta$，以下不等式成立：

$$R(f) \leqslant \hat{R}(f) + \varepsilon(d, N, \delta) \qquad \varepsilon(d, N, \delta) = \sqrt{\frac{1}{2N}\left(\log d + \log\frac{1}{\delta}\right)}$$

**泛化能力 generalization ability

假设空间有无限个函数？

经验风险都为0？



经验风险最小化函数： $f = \arg\min_{f \in \mathcal{F}} \hat{R}(f)$

泛化能力： $R(f) = E[L(Y, f(X))]$

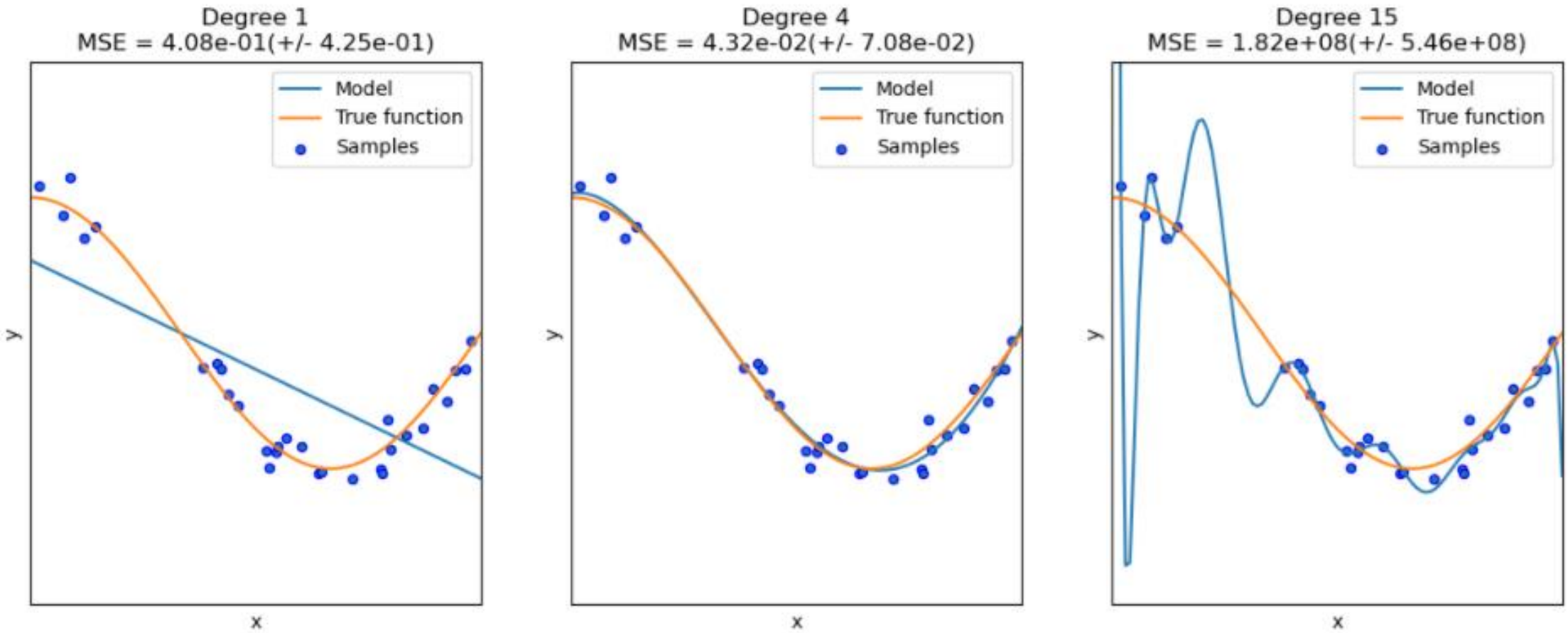
定理：泛化误差上界，二分类问题，

对于给定的$\delta > 0$，假设空间$\mathcal{F} = \{f(\boldsymbol{x}, \boldsymbol{w})\}$的**VC维**为$d^{VC}$

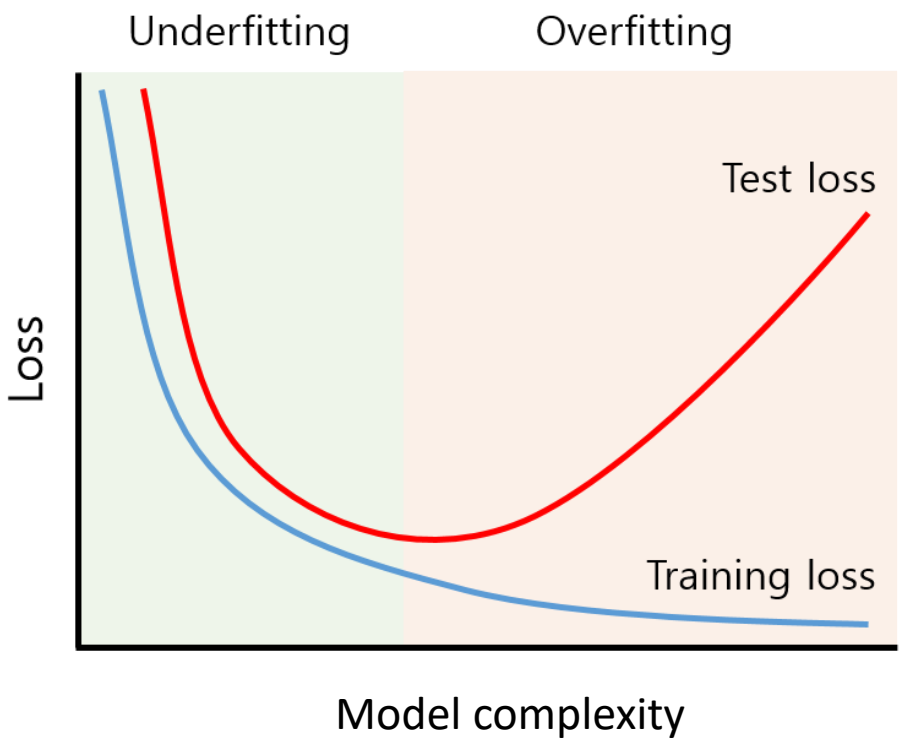
对任意一个函数f， 至少以概率$1 - \delta$，以下不等式成立：

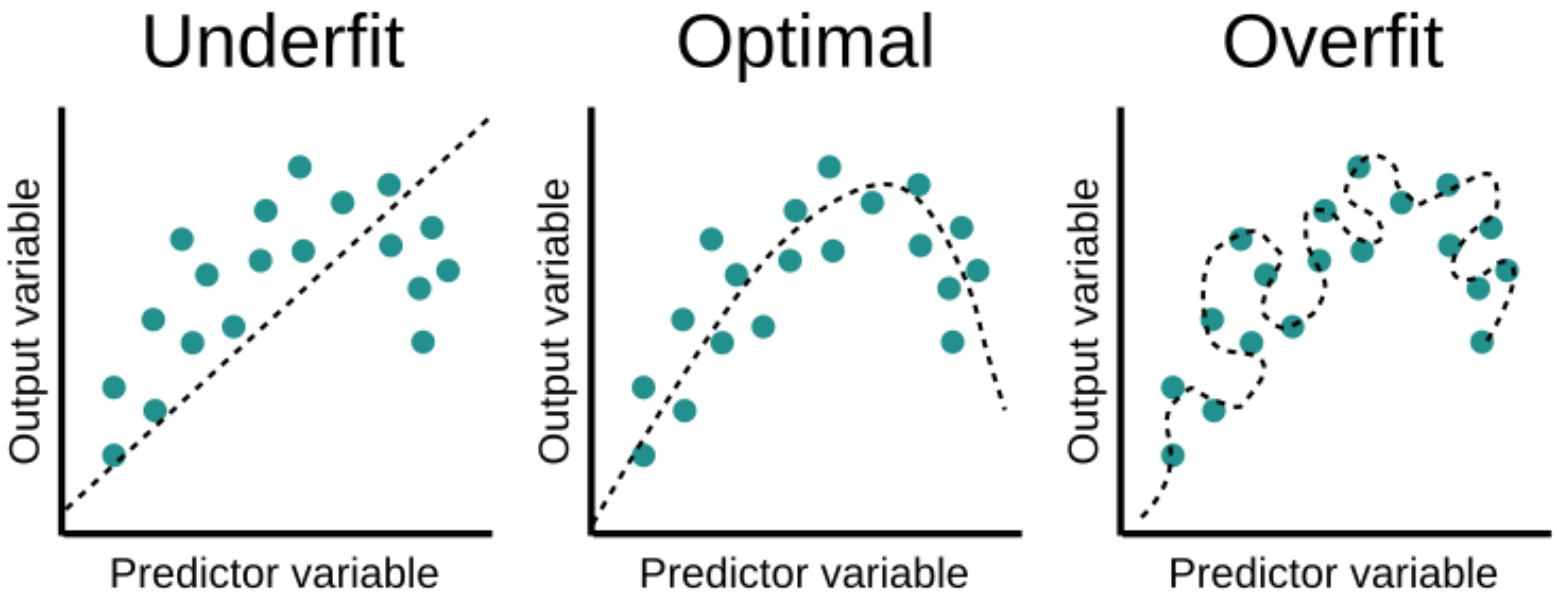
$$R(f) \leqslant \hat{R}(f) + \varepsilon(d^{VC}, N, \delta) \qquad \varepsilon(d^{VC}, N, \delta) = \sqrt{\frac{1}{8N} \log \frac{4(2N)^{d^{VC}} + 4}{\delta}}$$

模型评估与模型选择

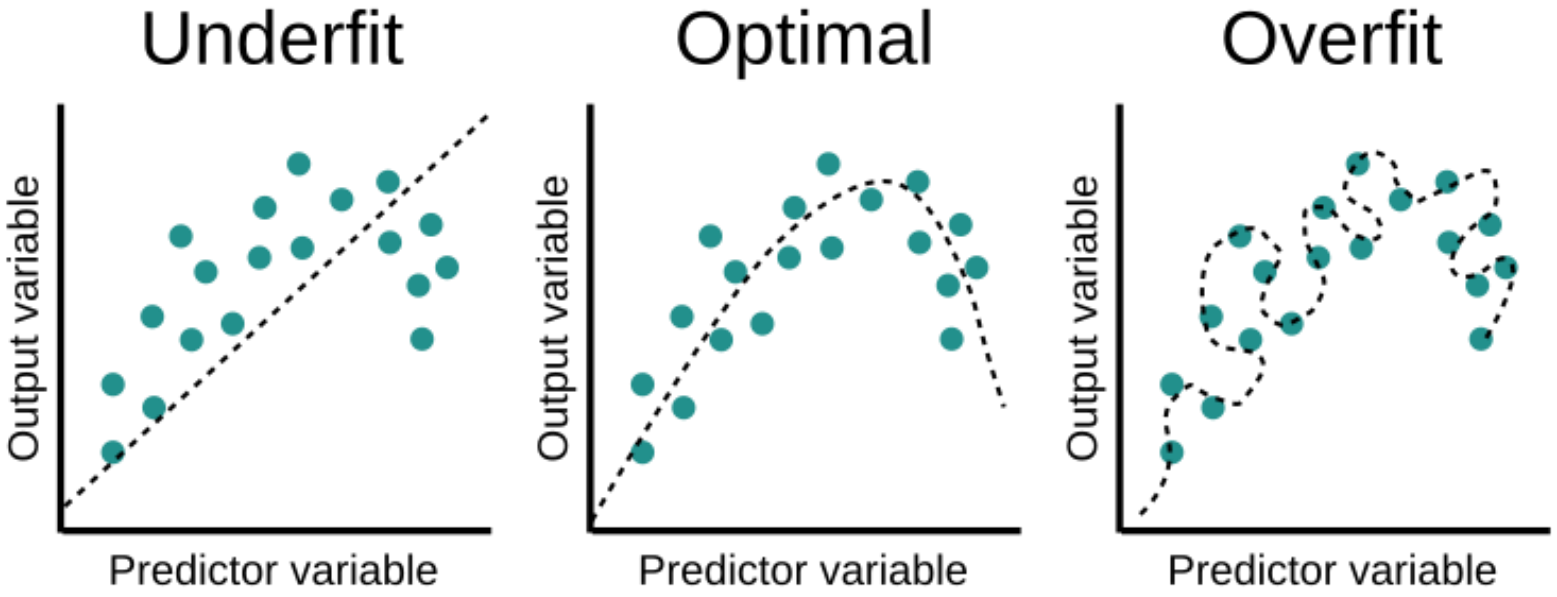

Degree 1
MSE = 4.08e-01(+/- 4.25e-01)
Model
True function
Samples
y
x
Degree 4
MSE = 4.32e-02(+/- 7.08e-02)
Model
True function
Samples
y
x
Degree 15
MSE = 1.82e+08(+/- 5.46e+08)
Model
True function
Samples
y
x

模型评估与模型选择


Underfitting
Overfitting
Test loss
Training loss
Loss
Model complexity


Underfit
Optimal
Overfit
Output variable
Predictor variable
Output variable
Predictor variable
Output variable
Predictor variable

模型评估与模型选择



In statistics, **overfitting** is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably".[1] An **overfitted model** is a statistical model that contains more parameters than can be justified by the data.[2] The essence of overfitting is to have unknowingly extracted some of the residual variation (i.e., the noise) as if that variation represented underlying model structure.[3]:45

Underfitting occurs when a statistical model cannot adequately capture the underlying structure of the data. An **under-fitted model** is a model where some parameters or terms that would appear in a correctly specified model are missing.[2] Under-fitting would occur, for example, when fitting a linear model to non-linear data. Such a model will tend to have poor predictive performance.

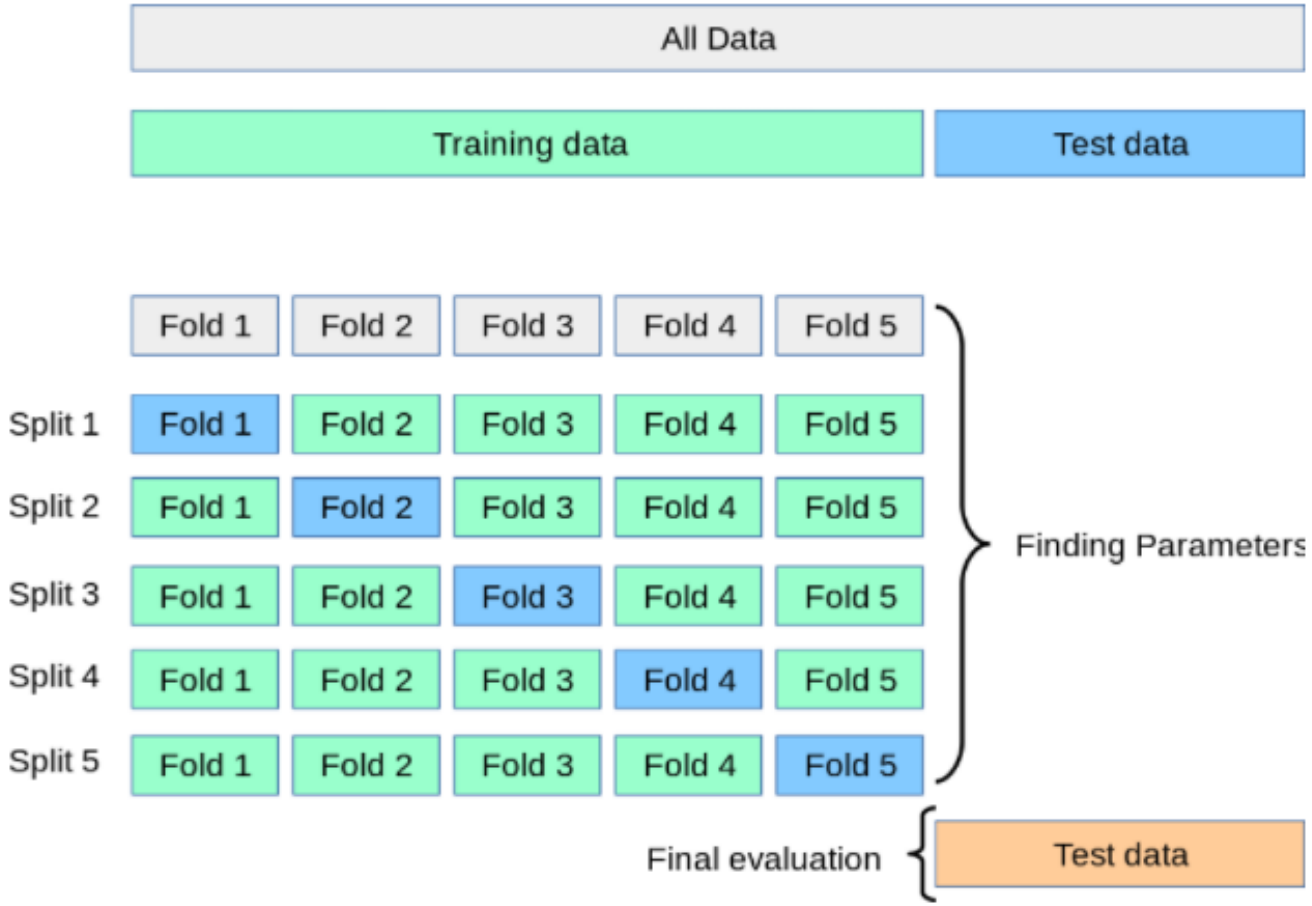
如何发现Overfitting

- **交叉验证**，有时亦称**循环估计**， 是一种统计学上将数据样本切割成较小子集的实用方法。于是可以先在一个子集上做分析，而其它子集则用来做后续对此分析的确认及验证。一开始的子集被称为**训练集**。而其它的子集则被称为**验证集**或**测试集**。交叉验证的目的，是用未用来给模型作训练的新数据，测试模型的性能，以便减少诸如过拟合和选择偏差等问题，并给出模型如何在一个独立的数据集上通用化（即，一个未知的数据集，如实际问题中的数据）

  - **Holdout 验证**
  - ***k*折交叉验证**
  - **留一法交叉验证**

5折交叉验证

训练集 training set：　　用于训练模型
验证集 validation set：　　用于模型选择
~~测试集 test set：　　用于最终对学习方法的评估~~



如何处理Overfitting

- 优化算法早停止
- 获取更多的训练数据
- 数据增强
- 特征选择
- 正则化
- 模型集成

二分类评价指标

混淆矩阵

	P	N
T	TP	FP
F	FN	TN

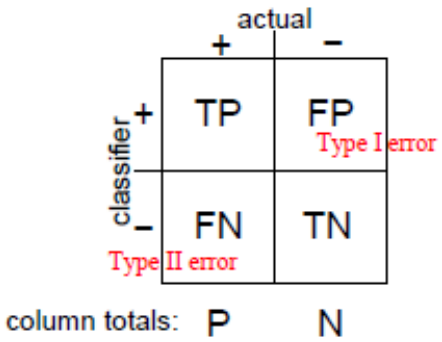
二分类评价指标

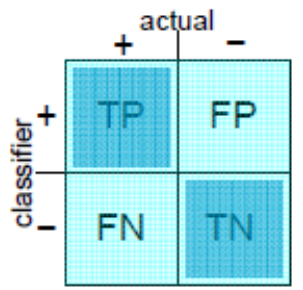
TP true positive
FN false negative
FP false positive
TN true negative

- 精确率 $P = \dfrac{TP}{TP + FP}$
- 召回率 $R = \dfrac{TP}{TP + FN}$
- F_1值 $\dfrac{2}{F_1} = \dfrac{1}{P} + \dfrac{1}{R}$

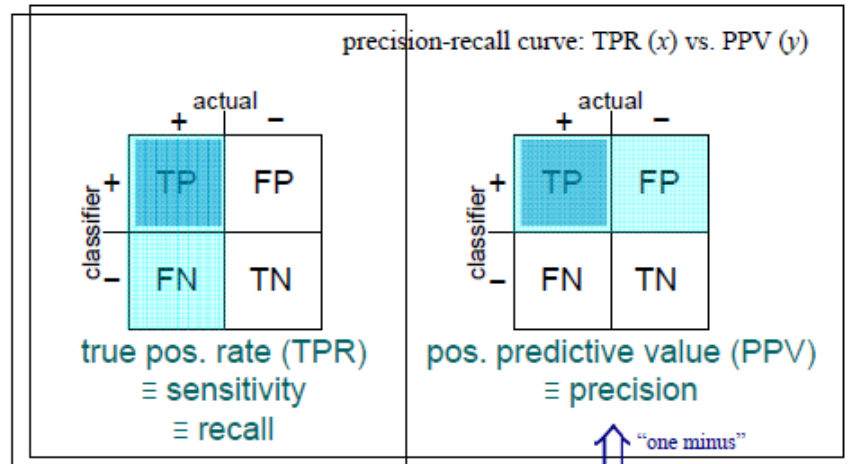
$$F_1 = \frac{2TP}{2TP + FP + FN}$$

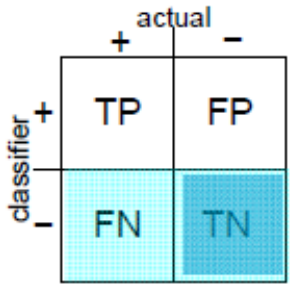
Confusion matrix:



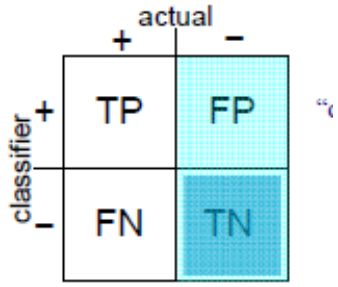


accuracy (ACC)





neg. predictive value (NPV)



specificity (SPC)

one minus"

Map points from ROC to Precision-Recall or vice-versa: (TPR same values in both)

$$\mathrm{PPV} = \frac{\mathrm{P\ TPR}}{\mathrm{P\ TPR} + \mathrm{N\ FPR}} \qquad \textit{(ROC to P-R)}$$

$$\mathrm{FPR} = \frac{\mathrm{P}\,(1 - \mathrm{PPV})\ \mathrm{TPR}}{\mathrm{N\ PPV}} \qquad \textit{(P-R to ROC)}$$

Confusion matrix

No.	1	2	3	4	5	6	7	8	9	10	11	12
Ground Truth	P	P	F	F	F	P	P	F	F	F	F	F
Tempreture	40°C	39°C	38.7°C	38.6°C	38.3°C	38.1°C	37.8°C	37.6°C	37.4°C	37.2°C	37°C	36.6°C

Threshhold parameter T= 36

	P	N
+	4	8
-	0	0

TPR = 1

FPR = 1

T= 38

	P	N
+	3	3
-	1	5

TPR = 0.75

FPR = 0.375

T= 38.5

	P	N
+	2	2
-	2	6

TPR = 0.5

FPR = 0.25

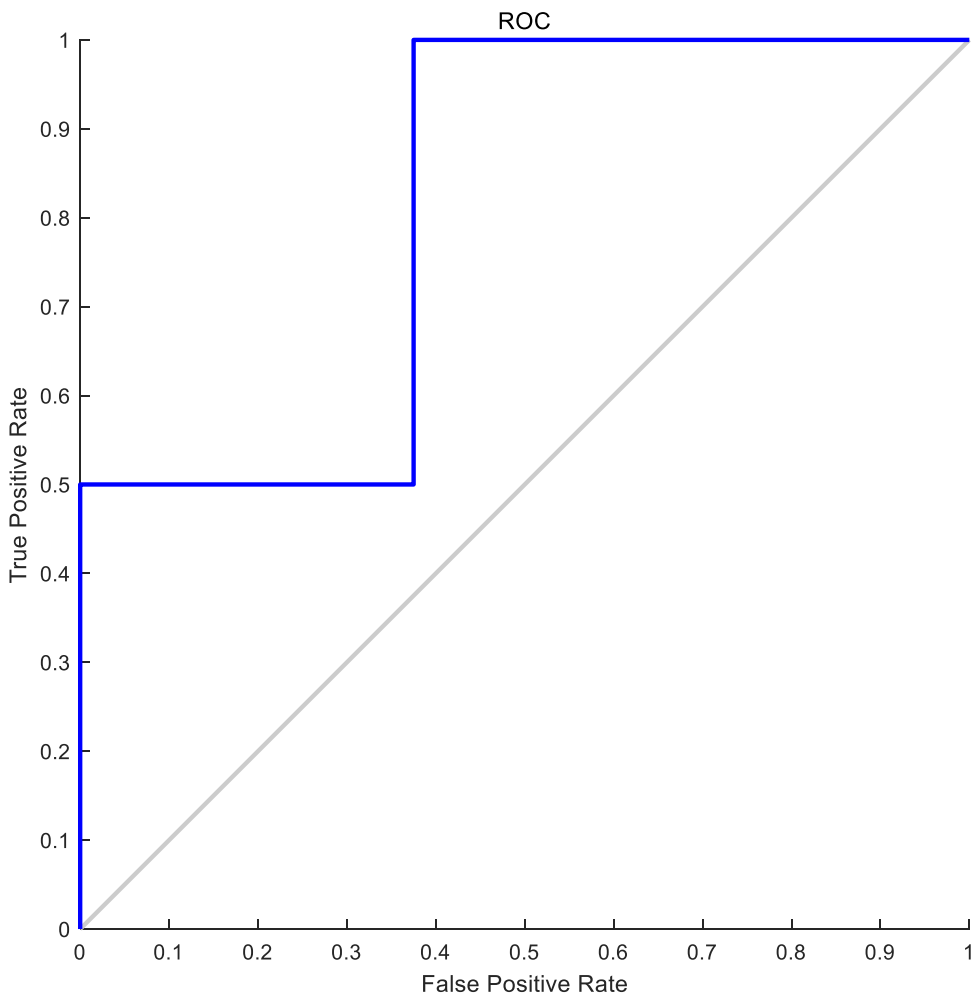
T= 38.8

	P	N
+	2	0
-	2	8

TPR = 0.5

FPR = 0

Receiver Operating Characteristic (ROC) Curve



1. Compute the confusion matrix and plot the ROC curve

The data points and the corresponding labels:

```
label = [1 1 0 0 0 1 1  0 0 0 0 0];
BTmp = [40 39 38.7 38.6 38.3 38.1 37.8 37.6 37.4 37.2 37 36.6];
```

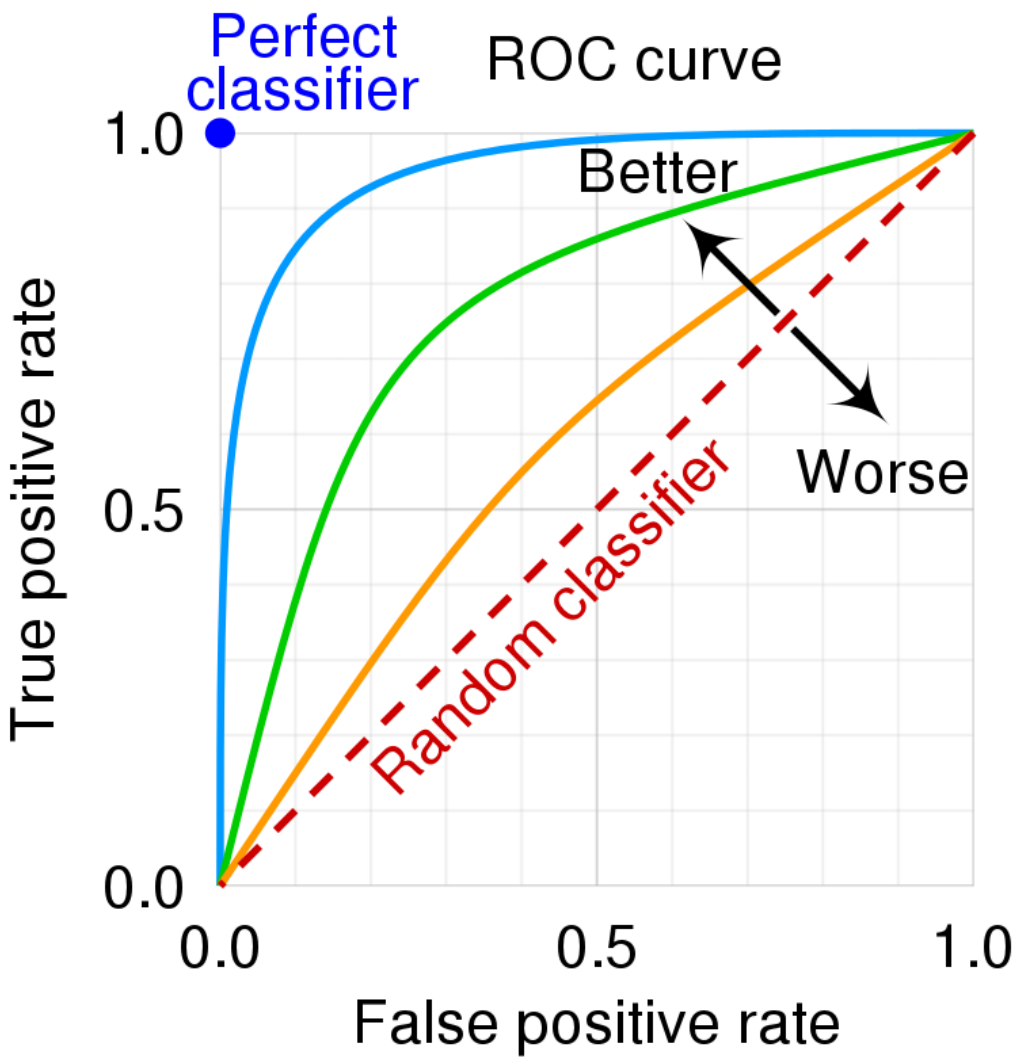
Compute the confusion matrix:

```
% Th is the threshold value
Th = 38;
% prediction
pred = double(BTmp>38);
% compute the confusionchart
cm = confusionchart(label,pred);
```

Plot the ROC curve:

```
% normalize BTmp in [0,1]
BTmp_norm = (BTmp-36.6)/(40-36.6);
% Plot the ROC curve
plotroc(label,BTmp_norm);
```

Receiver Operating Characteristic (ROC) Curve



*多分类评价指标

混淆矩阵

		Predicted		
		Airplane	Boat	Car
Actual	Airplane	2	1	0
	Boat	0	1	0
	Car	1	2	3

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)
Airplane	2	1	1
Boat	1	3	0
Car	3	0	3

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

*多分类评价指标

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

Macro Average: The macro-averaged F1 score (or macro F1 score) is computed by taking the arithmetic mean (aka **unweighted** mean) of all the per-class F1 scores.

Label	Per-Class F1 Score	Macro-Averaged F1 Score
Airplane	0.67	$\frac{0.67 + 0.40 + 0.67}{3}$ **= 0.58**
Boat	0.40	
Car	0.67	

*多分类评价指标

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

Macro Average: The macro-averaged F1 score (or macro F1 score) is computed by taking the arithmetic mean (aka **unweighted** mean) of all the per-class F1 scores.

Label	Per-Class F1 Score	Macro-Averaged F1 Score
Airplane	0.67	$\frac{0.67 + 0.40 + 0.67}{3}$ **= 0.58**
Boat	0.40	
Car	0.67	

*多分类评价指标

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

Weighted Average: The **weighted-averaged** F1 score is calculated by taking the mean of all per-class F1 scores **while considering each class's support**.

Label	Per-Class F1 Score	Support	Support Proportion	Weighted Average F1 Score
Airplane	0.67	3	0.3	$(0.67 * 0.3) + (0.40 * 0.1) + (0.67 * 0.6)$ **= 0.64**
Boat	0.40	1	0.1	
Car	0.67	6	0.6	
Total	-	10	1.0	

*多分类评价指标

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

Micro Average: Micro averaging computes a **global average** F1 score by counting the **sums** of the True Positives (**TP**), False Negatives (**FN**), and False Positives (**FP**).

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Micro-Averaged F1 Score
Airplane	2	1	1	$\frac{TP}{TP+\frac{1}{2}(FP+FN)} = \frac{6}{6+\frac{1}{2}(4+4)}$ = **0.60**
Boat	1	3	0	
Car	3	0	3	
TOTAL	**6**	**4**	**4**	

一个例子：mAP （mean average precision）

讲到mAP，首先要搞清楚平均精度AP（average precision）的概念。AP是针对单类目标检测问题。我们仍以新冠检测为例，比如现有10个人被检出阳性，中间有5个是新冠病人。检测使用的阈值设为37.5°C，我编的数据如下：

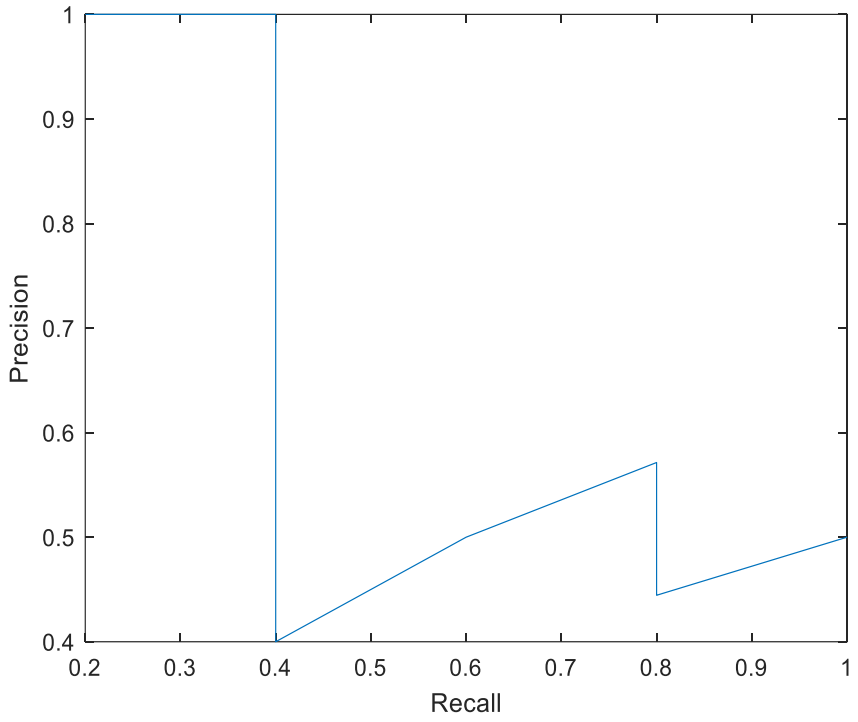
序号	1	2	3	4	5	6	7	8	9	10
真实值	P	P	F	F	F	P	P	F	F	P
体温	42°C	41°C	40°C	39°C	39°C	39°C	38°C	38°C	38°C	38°C
检测是否正确	对	对	错	错	错	对	对	错	错	对

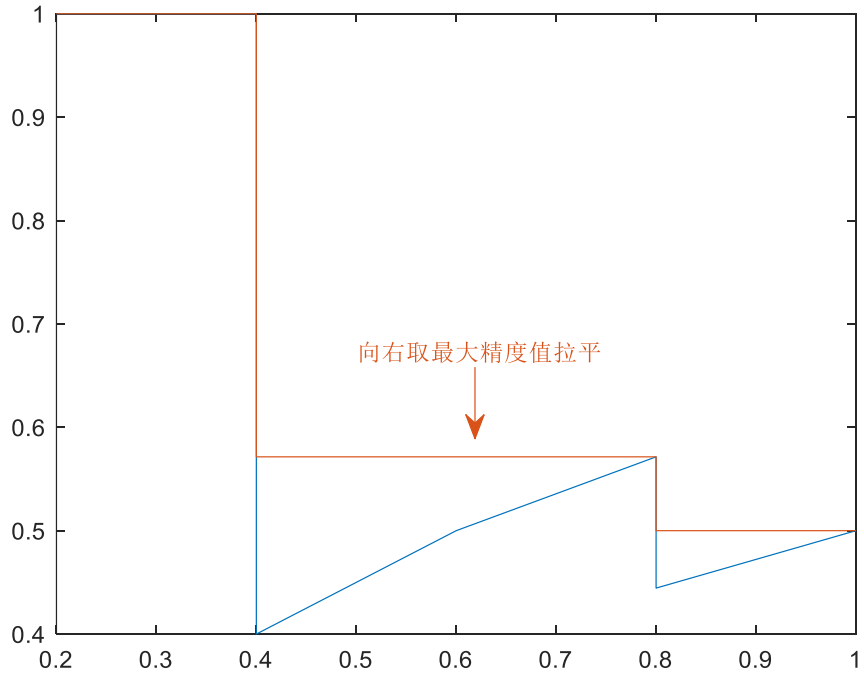
AP （average precision）

计算AP，首先我们先将获得的体温值从高到底进行排序（很关键），观察序号第一的人，检测结果是正确的，此时TP=1，既目前仅一个人且被正确检测出，FP=0，此时Precision=1。因已知新冠病人有5个，所以此时的Recall=1/5 = 0.2。以此类推，前两人时，Precision = 1，Recall = 2/5 = 0.4。前三人时，TP=2，FP=1，Precision=2/3，Recall=2/5 = 0.4。

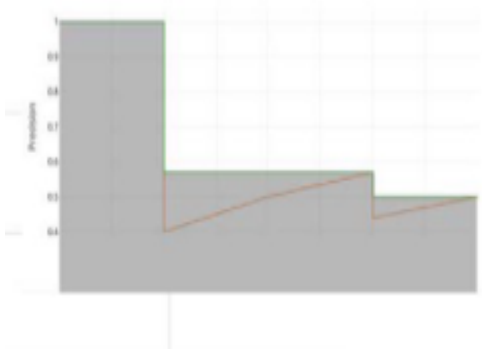
序号	1	2	3	4	5	6	7	8	9	10
Precision	1	1	2/3	1/2	2/5	1/2	4/7	1/2	4/9	1/2
Recall	1/5	2/5	2/5	2/5	2/5	3/5	4/5	4/5	4/5	1

AP （average precision）





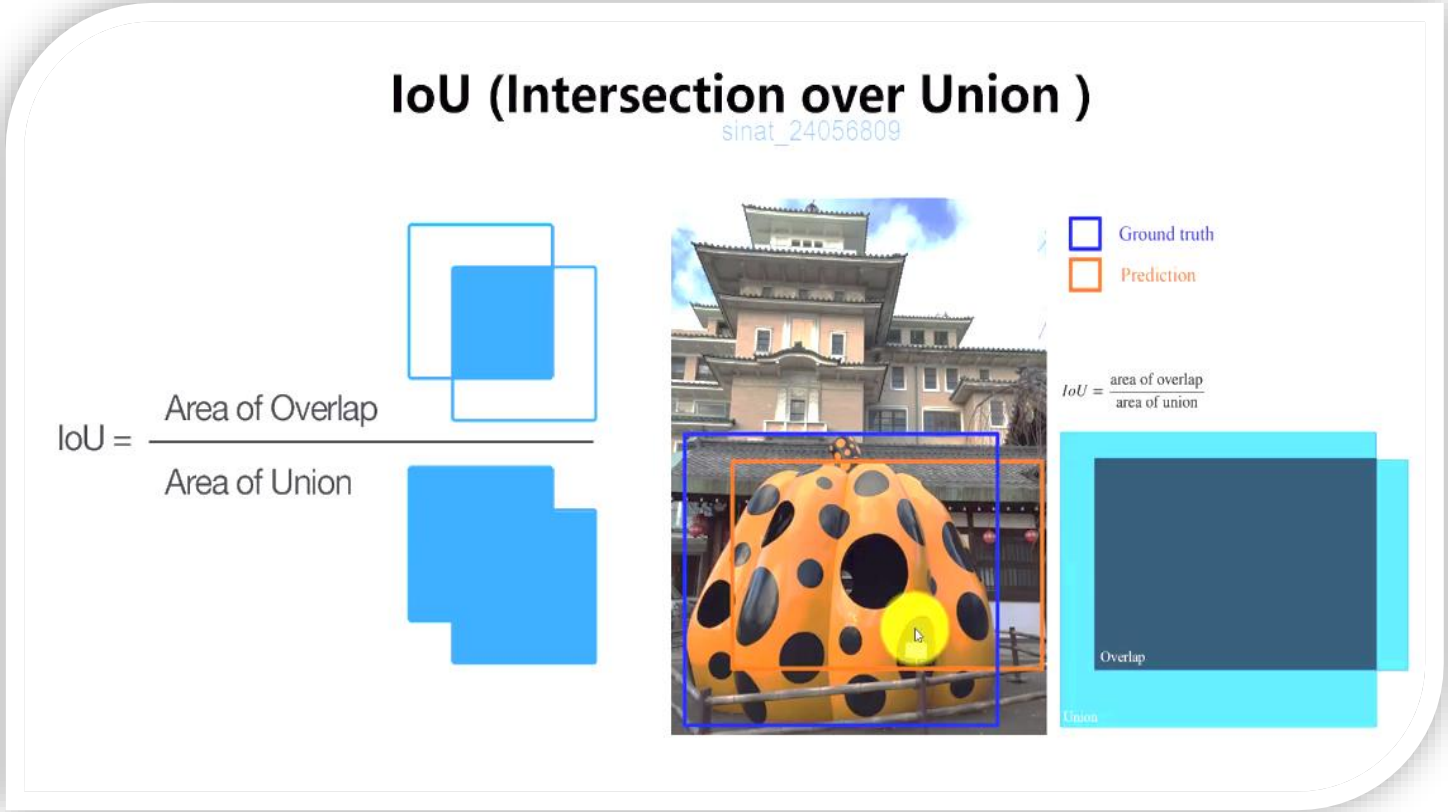
AP 就是右侧橙色折线图的下面积。



mAP （mean average precision）

mAP实际是多类的平均AP，比如我们不单单检测新冠肺炎，还检测流感，那么这两类都会产生AP值，我们把他们平均一下就得到mAP。因为是在37.5°C为阈值得到的mAP，可以记为mAP@37.5°C。当然还可以得到mAP@38°C等。

mAP （mean average precision）



回到图像目标检测问题，如要检测行人、车辆、交通标志、狗等几类，根据不同的IoU评价指标，我们也可以得到mAP@0.5，mAP@0.7，mAP@0.9等。在COCO数据集上，mAP@[.5:.95]是对步长0.05，阈值从0.5到0.95的平均AP值。

$$\mathrm{mAP}_{\mathrm{coco}} = \mathrm{mAP}@[.5:.95] = \frac{\mathrm{mAP}@0.5 + \mathrm{mAP}@0.55 + \cdots + \mathrm{mAP}@0.95}{10}$$

小结

- 模式识别、统计学习的定义与研究对象
- 模式识别与统计学习的分类
- 统计学习三要素
- 过拟合与欠拟合，如何发现与处理
- 混淆矩阵
- Precision, Recall, TPR, FPR与F1 score
- P-R曲线与ROC曲线

小结

- 模式识别、统计学习的定义与研究对象
- 模式识别与统计学习的分类
- 统计学习三要素
- 过拟合与欠拟合，如何发现与处理
- 混淆矩阵
- Precision, Recall, TPR, FPR与Accuracy, F1 score
- P-R曲线与ROC曲线

谢谢各位同学！

群名称：模式识别与统计学习

群　号：497956575