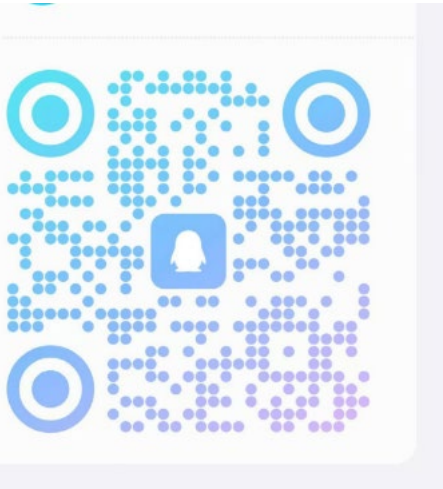
Machine Learning
机器学习

Lecture1：开学篇

李洁
nijanice@163.com

倪张凯
zkni@tongji.edu.cn



QQ群：559174662

人工智能

利用计算机解决人类通过直觉可以解决的问题

（如：自然语言理解，图像识别，语音识别等）

机器学习

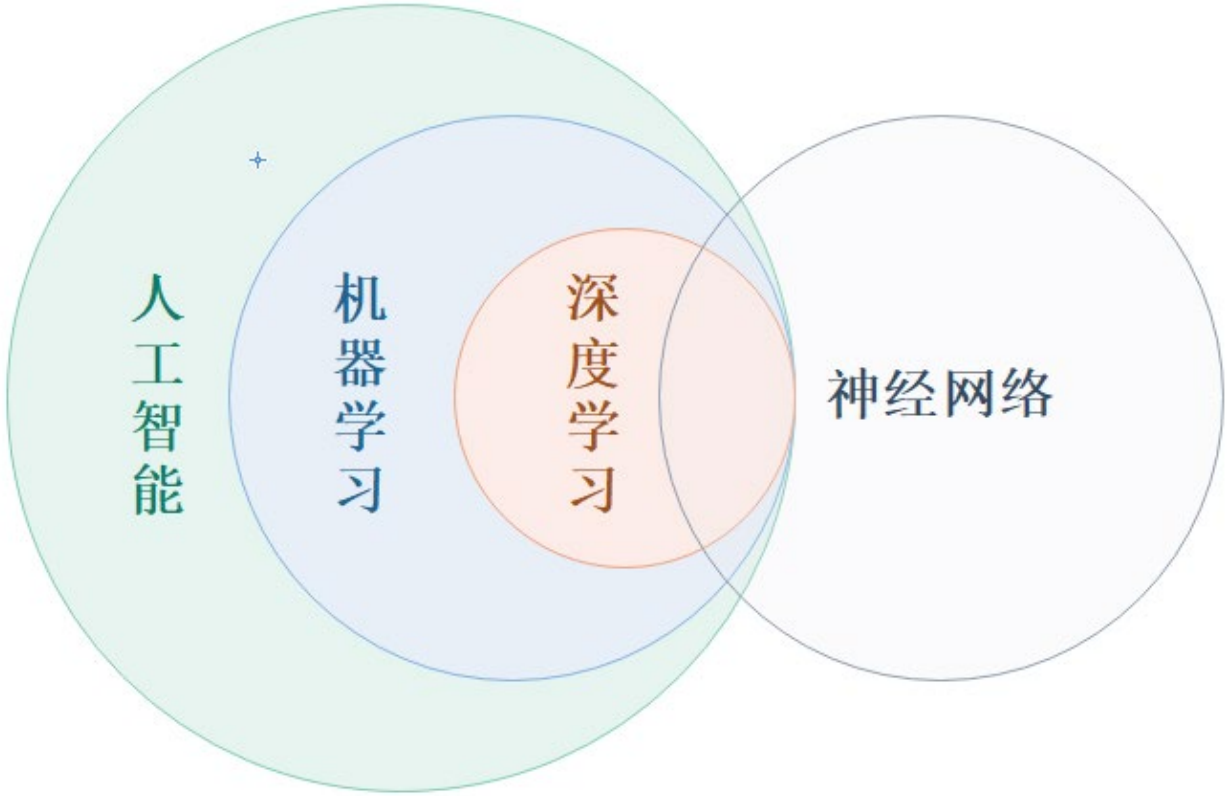
通过赋予机器学习的能力使其完成一系列功能

深度学习

自动将简单的特征组合成更加复杂的特征，并用这些特征解决问题

神经网络

最初是指大脑神经元，人工智能受神经网络的启发，发展出了人工神经网络。



人工智能的方法
Methodologies of Artificial Intelligence

人工智能的方法
Methodologies of Artificial Intelligence

- Rule-based

  Implemented by direct programing

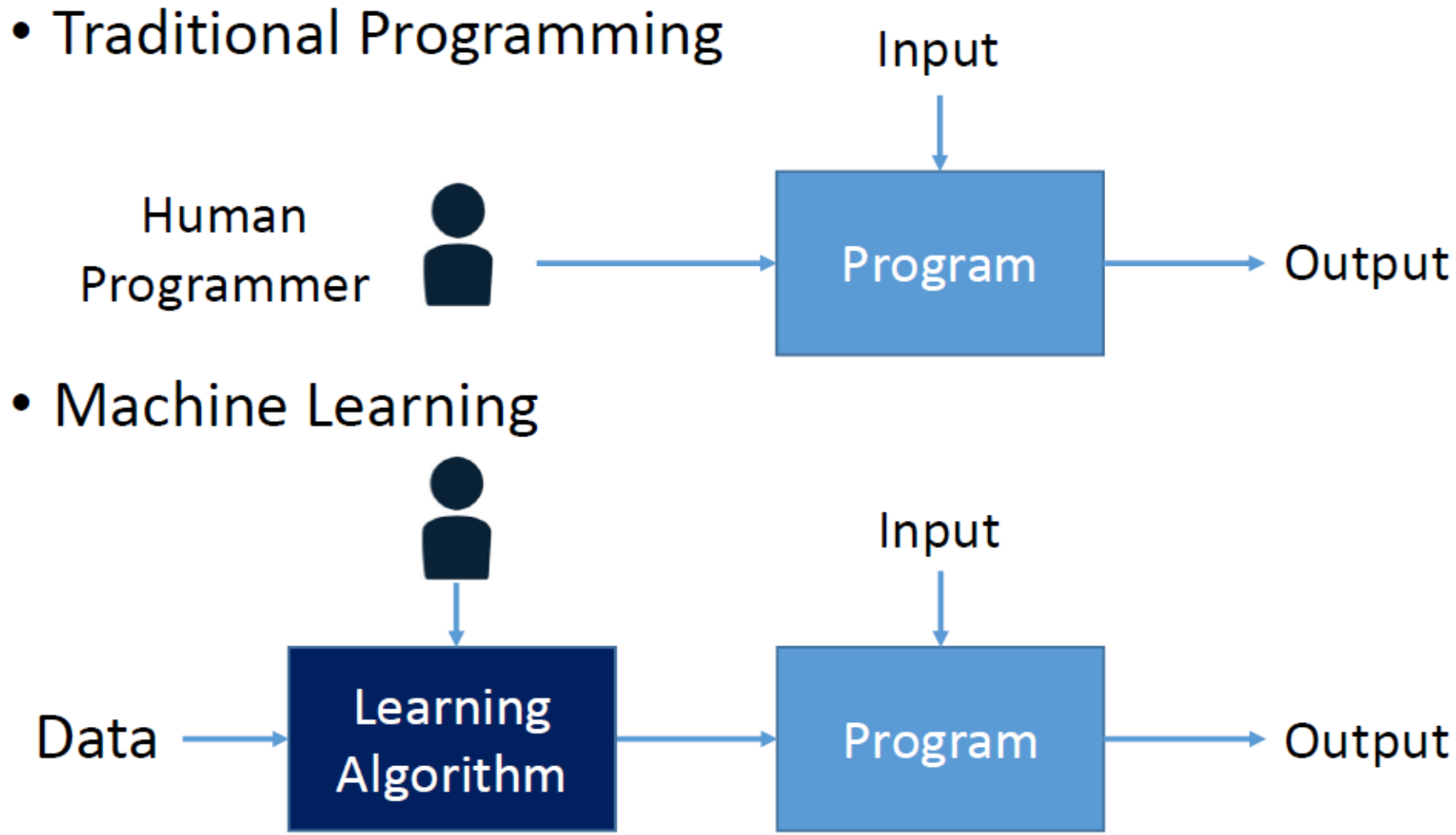
  Inspired by human heuristics

人工智能的方法
Methodologies of Artificial Intelligence

- Rule-based
  - Implemented by direct programing
  - Inspired by human heuristics

- Data-based
  - Expert systems
    - Experts or statisticians create rules of predicting or decision making based on the data
  - Machine learning
    - Direct making prediction or decisions based on the data
    - Data Science

机器学习和传统编程的不同
Traditional Programming vs. Machine Learning



机器学习适用于…
Learning is used when

- Develop systems that are too difficult/expensive to construct manually because they require specific detailed skills or knowledge tuned to a specific task （Speech / face recognition）
- Even if we had a good idea about how to do it, the program might be horrendously complicated. (Robot arm, autonomous helicopter, handwriting recognition, most of natural language processing, game of Go)
- Human expertise does not exist (navigating on Mars)
- Rapid decisions that humans cannot do(high-frequency trading)
- Solution changes in time (routing on a computer network)

机器学习适用于…
Learning is used when

- Develop systems that can automatically adapt and customize themselves to individual users in a massive scale.
  - Personalized news or mail filter
  - Personalized tutoring
  - Product recommendation

- Discover new knowledge from large databases (data mining).
  - Market basket analysis (e.g. web click data)
  - Medical text mining (e.g. migraines to calcium channel blockers to magnesium)

机器学习定义
What is Machine Learning?

- Learning is any process by which a system improves performance from experience

--- Herbert Simon

Turing Award (1975)

artificial intelligence, the psychology of human cognition

Nobel Prize in Economics (1978)

decision-making process within economic organizations

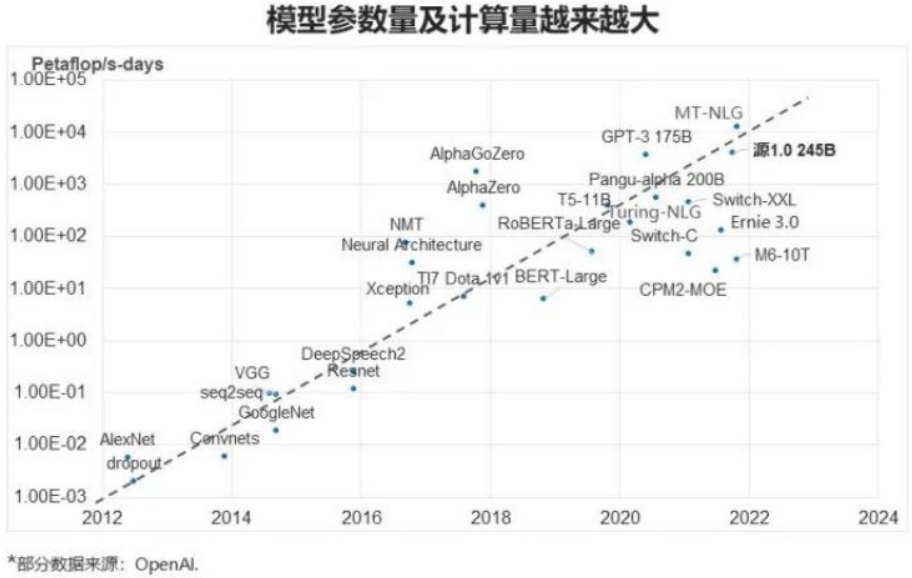
机器学习定义
What is Machine Learning?

- A more mathematical definition by Tom Mitchell

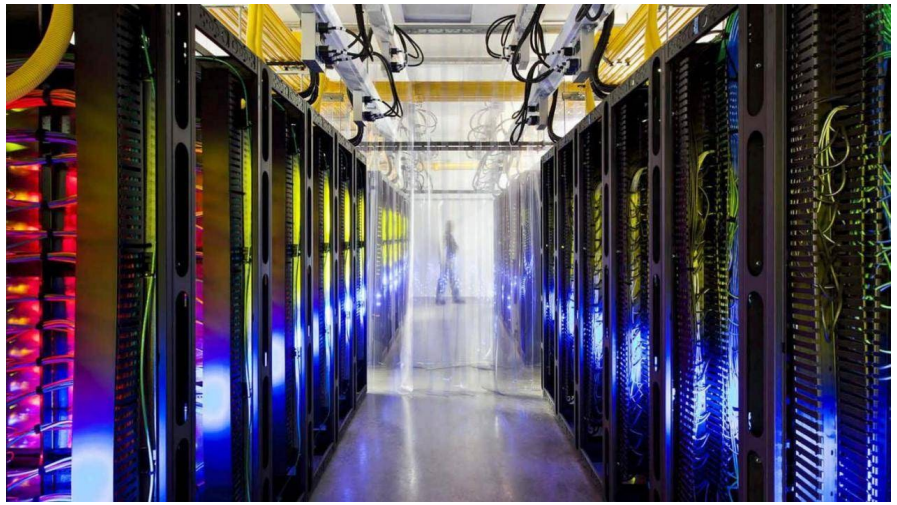
- Machine learning is the study of algorithms that
    - improvement their performance *P*
    - at some task *T*
    - based on experience *E*
    - with non-explicit programming
- A well-defined learning task is given by <*P, T, E*>

机器学习条件
Why Study Machine Learning? The Time is Ripe

- Many basic effective and efficient algorithms available.
- Large amounts of on-line data available.
- Large amounts of computational resources available.



一万亿参数神经网络



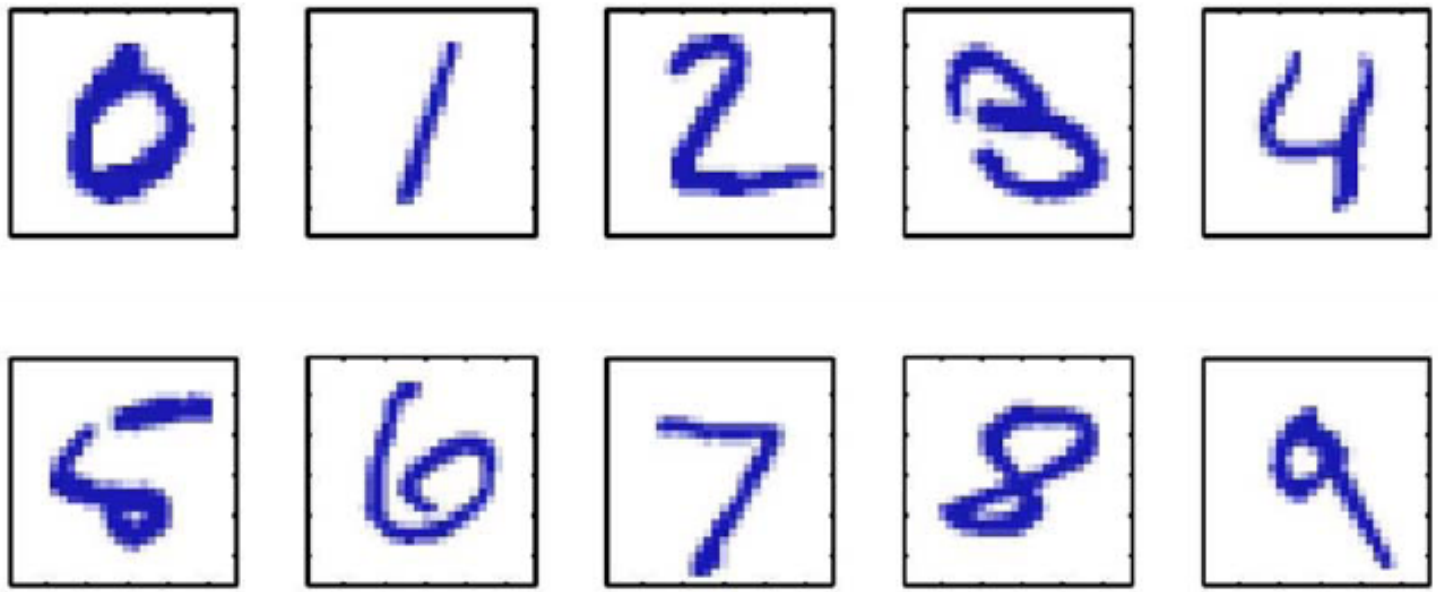
模拟训练一次仅电费大约1000万美元

学什么
What is Learning?

- Machine learning is the study of algorithms that improvement their performance *P* at some task *T* based on experience *E* with non-explicit programming

- What is the task?
  - Classification
  - Regression
  - Clustering
  - Dimensionality reduction

Example 1: hand-written digit recognition



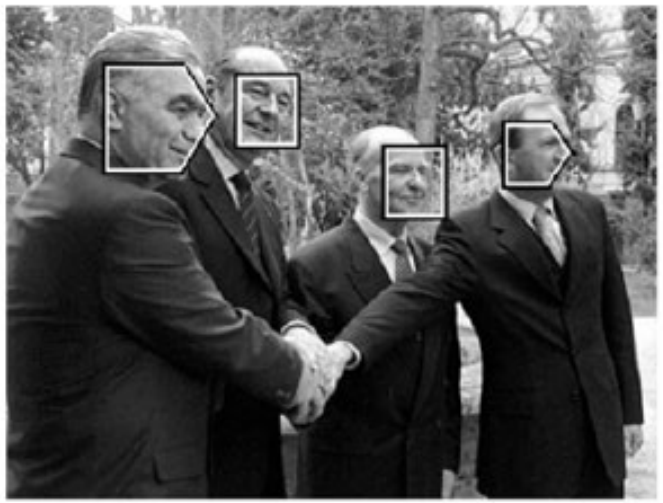
Images are 28 x 28 pixels

Represent input image as a vector $\mathbf{x} \in \mathbb{R}^{784}$

Learn a classifier $f(\mathbf{x})$ such that,

$$f : \mathbf{x} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Example 2: Face detection



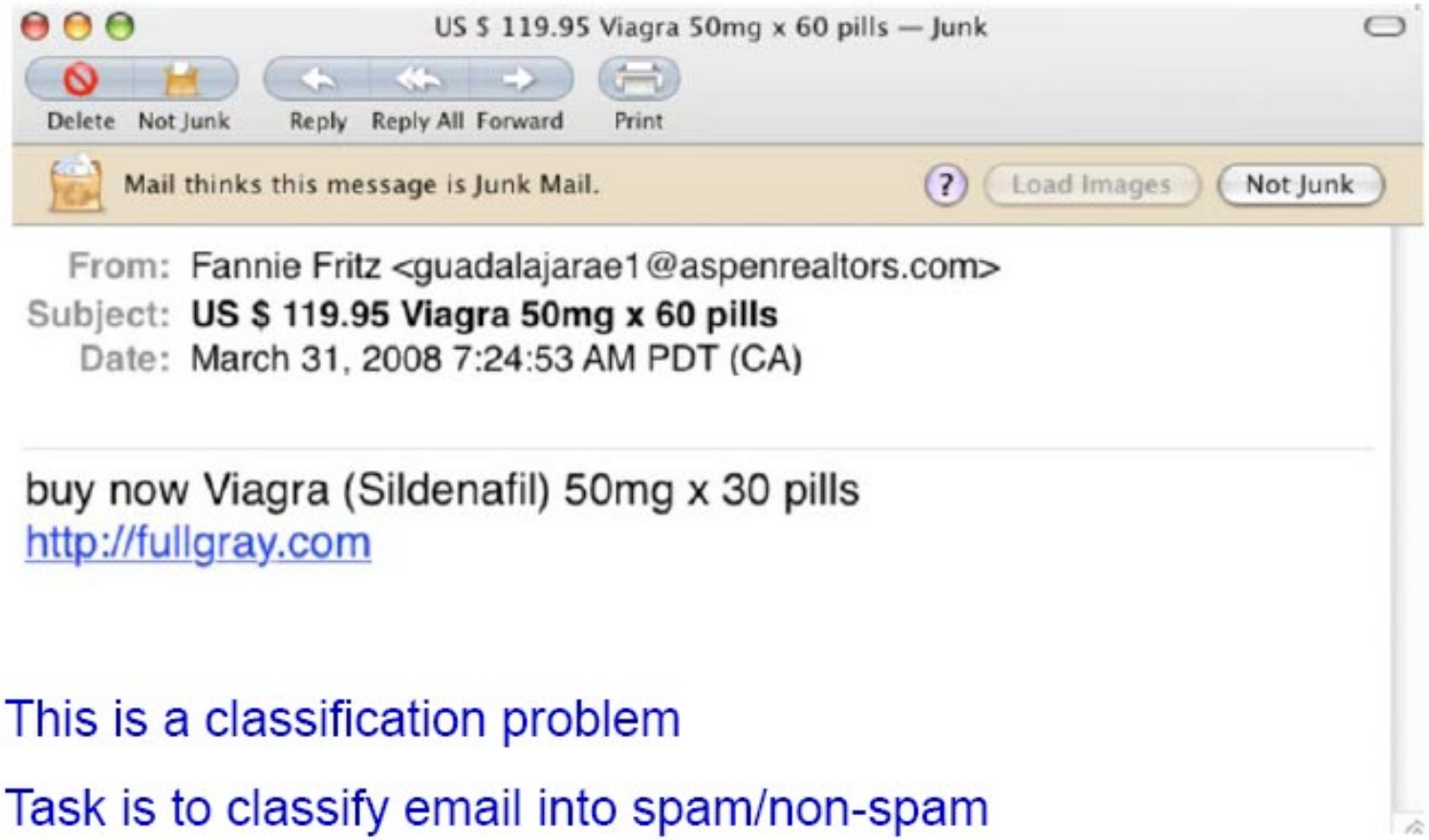
- Again, a supervised classification problem
- Need to classify an image window into three classes:
  - non-face
  - frontal-face
  - profile-face

Classifier is learnt from labelled data

Training data for frontal faces

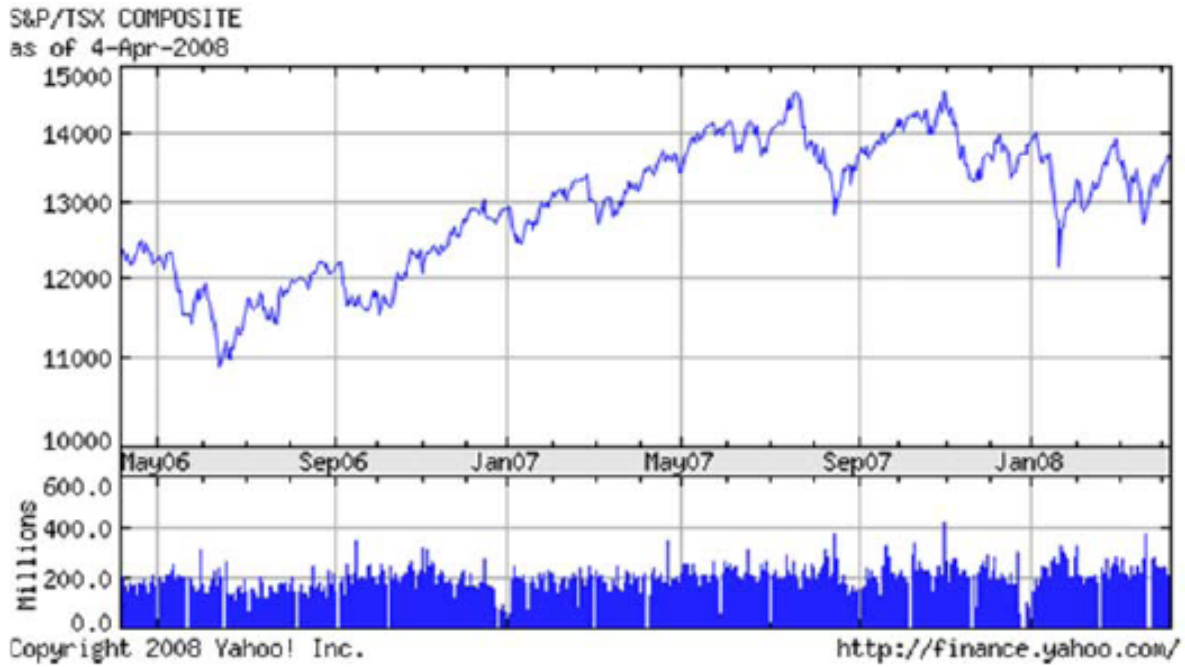
- 5000 faces
  - All near frontal
  - Age, race, gender, lighting
- 10^8 non faces
- faces are normalized
  - scale, translation

Example 3: Spam detection



- This is a classification problem
- Task is to classify email into spam/non-spam
- Data x_i is word count, e.g. of viagra, outperform, "you may be surprized to be contacted" …
- Requires a learning system as "enemy" keeps innovating

Example 4: Stock price prediction



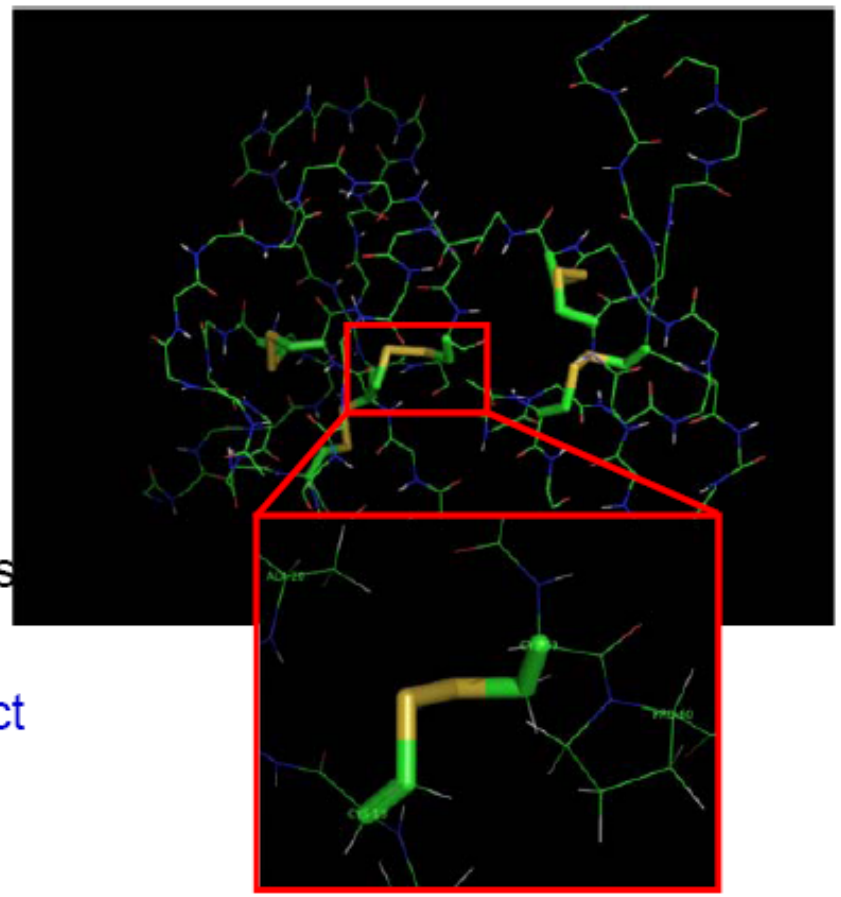
- Task is to predict stock price at future date
- This is a regression task, as the output is continuous

Example 5: Computational biology

x

y

AVITGACERDLQCG
KGTCCAVSLWIKSV
RVCTPVGTSGEDCH
PASHKIPFSGQRMH
HTCPCAPNLACVQT
SPKKFKCLSK



Protein Structure and Disulfide Bridges

Regression task: given sequence predict 3D structure

Protein: 1IMT

Example 6: s: Recommender systems

People who bought Hastie …

Frequently Bought Together

Customers buy this book with Pattern Recognition and Machine Learning (Information Science and Statistics) (Information Science and Statistics) by Christopher M. Bishop

Price For Both: £104.95

Add both to Basket

Customers Who Bought This Item Also Bought

Page 1

Pattern Recognition and Machine Learning (Infor...
by Christopher M. Bishop
★★★★☆ (4) £48.96

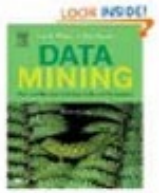
Show related items

MACHINE LEARNING (Mcgraw-Hill International Edit) by Thom M. Mitchell
★★★★★ (3) £42.74

Show related items

Pattern Classification, Second Edition: 1 (A Wi...
by Richard O. Duda
★★★★★ (1) £78.38

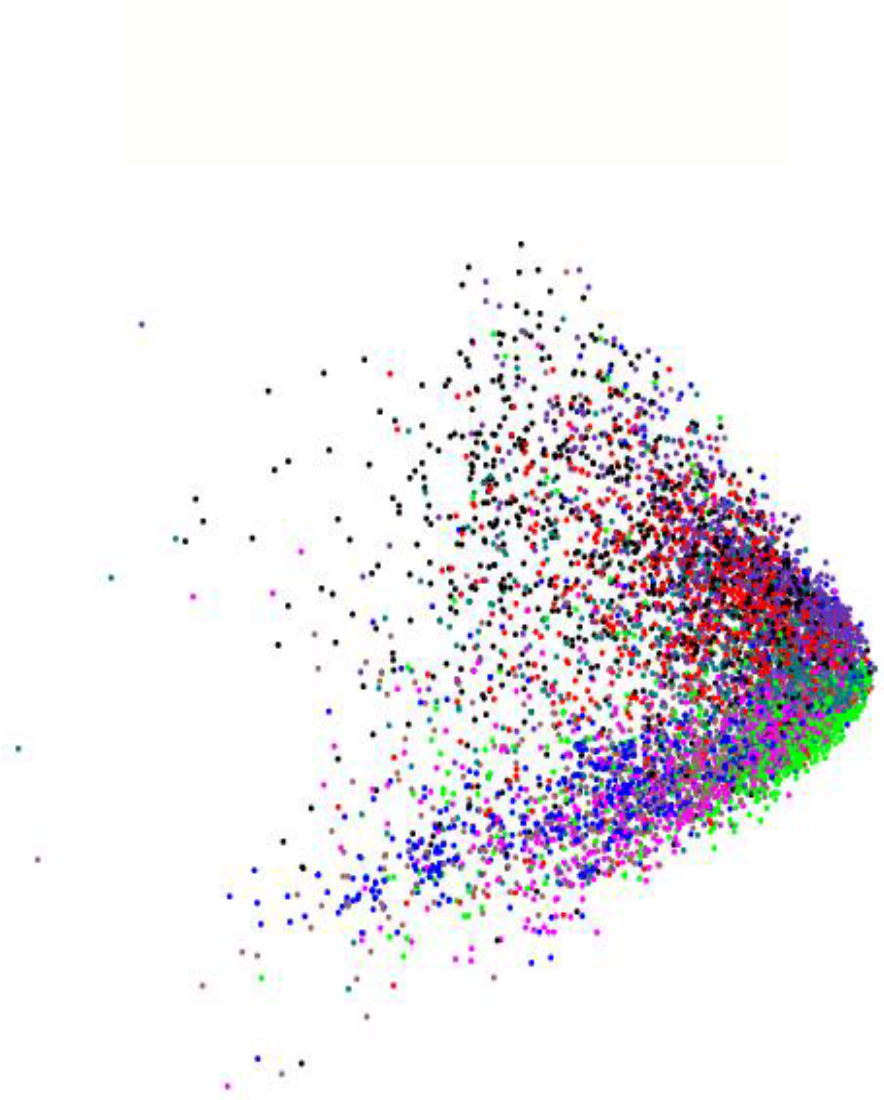
Show related items



Data Mining: Practical Machine Learning Tools a...
by Ian H. Witten
★★★★★ (1) £37.04

Show related items

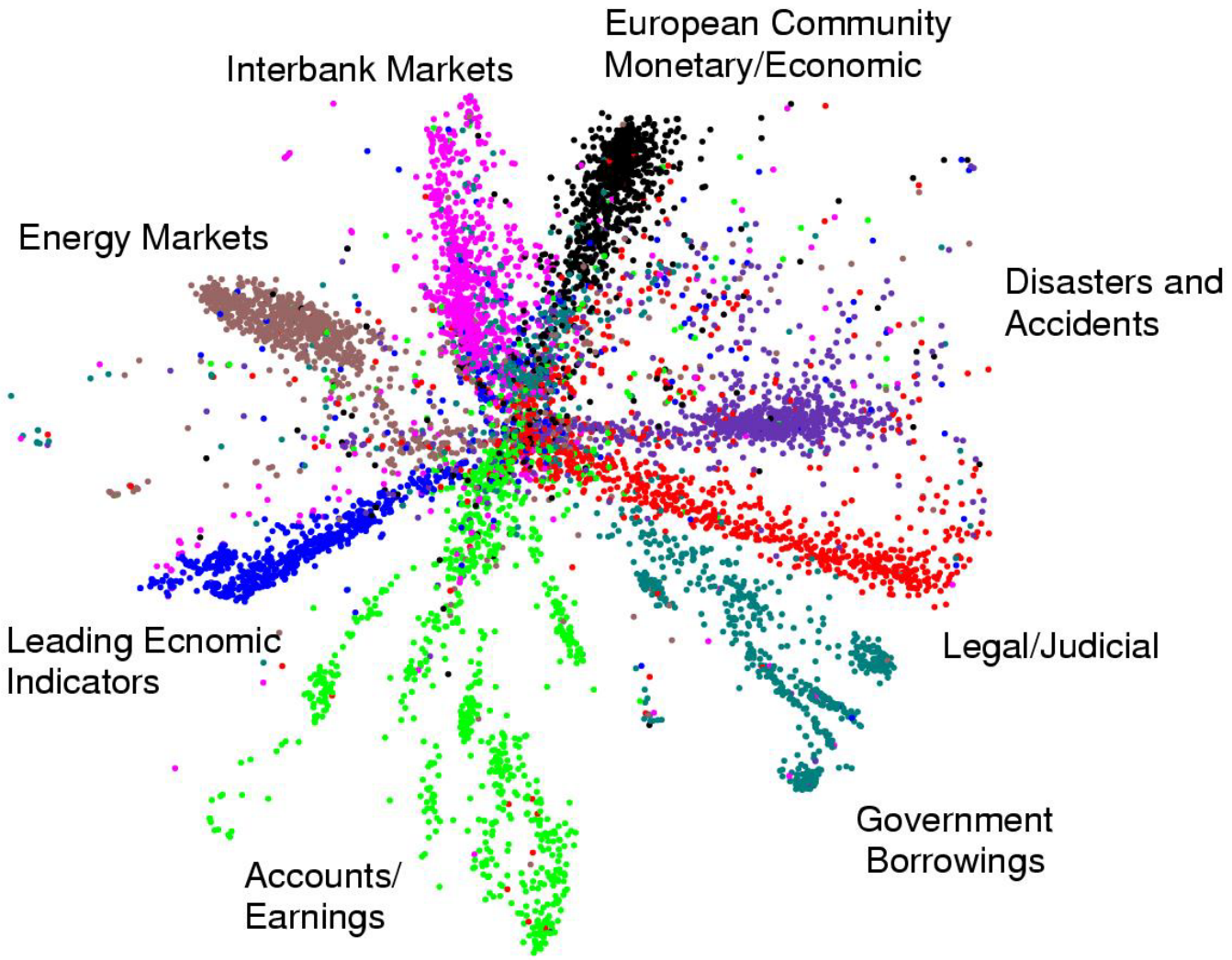
Example 7: Displaying the structure of a set of documents using Latent Semantic Analysis (a form of PCA)



Each document is converted to a vector of word counts. This vector is then mapped to two coordinates and displayed as a colored dot. The colors represent the hand-labeled classes.

When the documents are laid out in 2-D, the classes are not used. So we can judge how good the algorithm is by seeing if the classes are separated.

Example 7: Displaying the structure of a set of documents using Latent Semantic Analysis (a form of PCA)



学习任务类型
Types of learning task

- Supervised learning
  - infer a function from labeled training data.
- Unsupervised learning
  - try to find hidden structure in unlabeled training data
  - clustering
- Reinforcement learning
  - To learn a policy of taking actions in a dynamic environment and acquire rewards

学习任务类型
Types of learning task

- Supervised learning
  – infer a function from labeled training data.

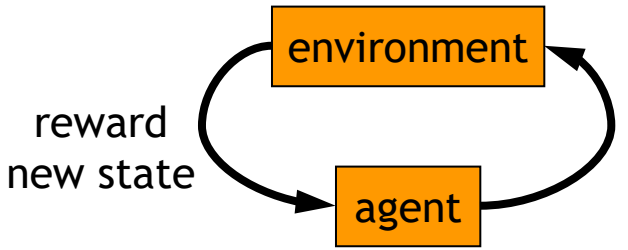
学习任务类型
Types of learning task

- Supervised learning
  - infer a function from labeled training data.

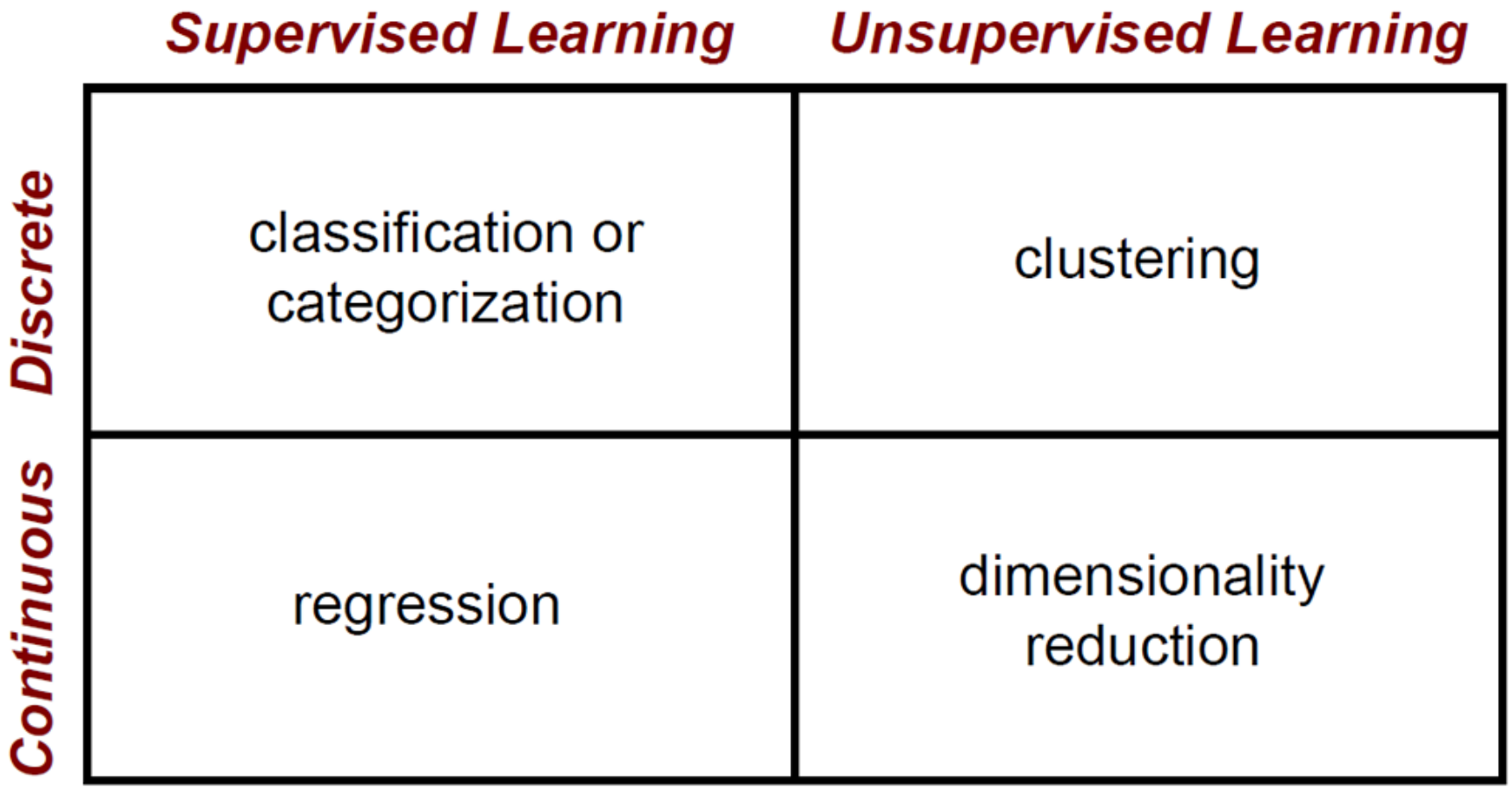
- Unsupervised learning
  - try to find hidden structure in unlabeled training data
  - clustering

学习任务类型
Types of learning task

- Supervised learning
  - infer a function from labeled training data.
- Unsupervised learning
  - try to find hidden structure in unlabeled training data
  - clustering
- Reinforcement learning
  - To learn a policy of taking actions in a dynamic environment and acquire rewards



学习任务类型
Types of learning task



机器学习发展历史
History of Machine Learning

- 1950s
  - Samuel's checker player
  - Selfridge's Pandemonium
- 1960s:
  - Neural networks: Perceptron
  - Pattern recognition
  - Learning in the limit theory
  - Minsky and Papert prove limitations of Perceptron
- 1970s:
  - Symbolic concept induction
  - Winston's arch learner
  - Expert systems and the knowledge acquisition bottleneck
  - Quinlan's ID3
  - Mathematical discovery with AM

机器学习发展历史
History of Machine Learning

- 1980s:
  - Advanced decision tree and rule learning
  - Explanation-based Learning (EBL)
  - Learning and planning and problem solving
  - Utility problem
  - Analogy
  - Cognitive architectures
  - Resurgence of neural networks
  - Valiant's PAC (Probably approximately correct) Learning Theory
- 1990s
  - Data mining
  - Adaptive software agents and web applications
  - Text learning
  - Reinforcement learning (RL)
  - Inductive Logic Programming (ILP)
  - Ensembles: Bagging, Boosting, and Stacking
  - Bayes Net learning
  - Support vector machines
  - Kernel methods

机器学习发展历史
History of Machine Learning

- 2000s
  - Graphical models (Bayesian networks and Markov random fields)
  - Variational inference
  - Statistical relational learning
  - Transfer learning
  - Sequence labeling
  - Collective classification and structured outputs
  - Computer Systems Applications
    - Compilers
    - Debugging
    - Graphics
    - Security (intrusion, virus, and worm detection)
  - Email management
  - Personalized assistants that learn
  - Learning in robotics and vision

机器学习发展历史
History of Machine Learning

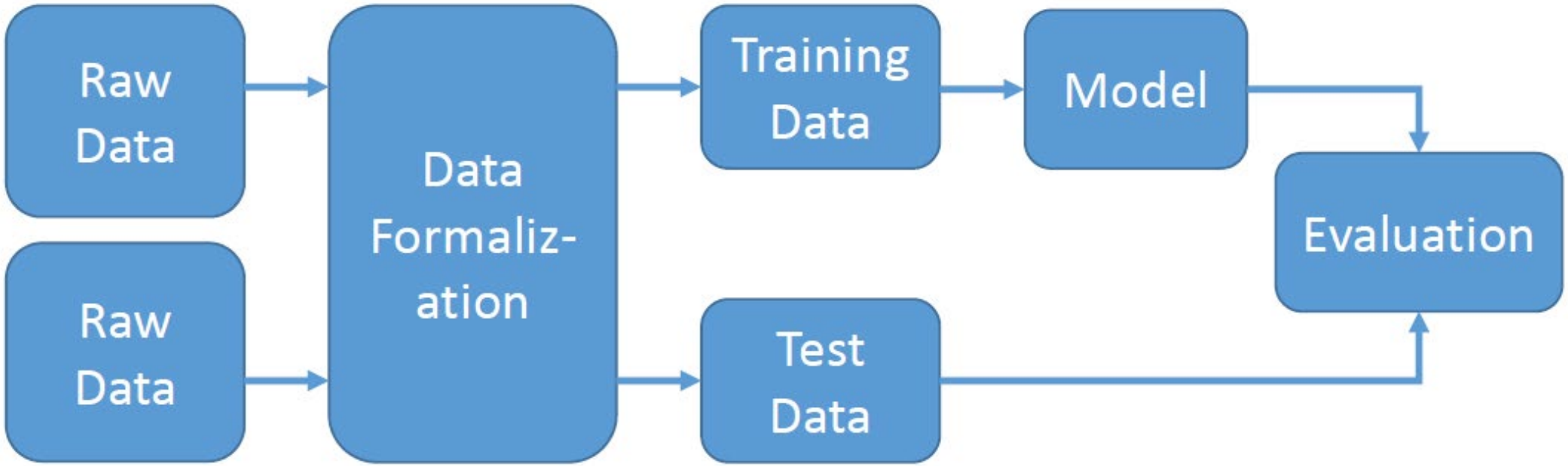
- 2010s
  - Deep learning (DNN, CNN, RNN)
  - ImageNet Challenge
  - Learning from big data
  - Learning with GPUs or HPC
  - Generative Adversarial Networks (GANs)
  - Reinforcement Learning Breakthroughs (AlphaGo)
  - Multi-task & lifelong learning
  - Deep reinforcement learning
  - Natural Language Processing (Transformer, BERT、GPT)

机器学习发展历史
History of Machine Learning

- 2020s…
  - Self-supervised Learning
  - Multimodal Learning (CLIP)
  - Transformer Models (GPT-3、BERT)
  - Large Language Models (LLMs)
  - Federated Learning
  - Quantum Machine Learning
  - Ethics and Fairness in AI
  - AI in Healthcare
  - AI for Climate Change

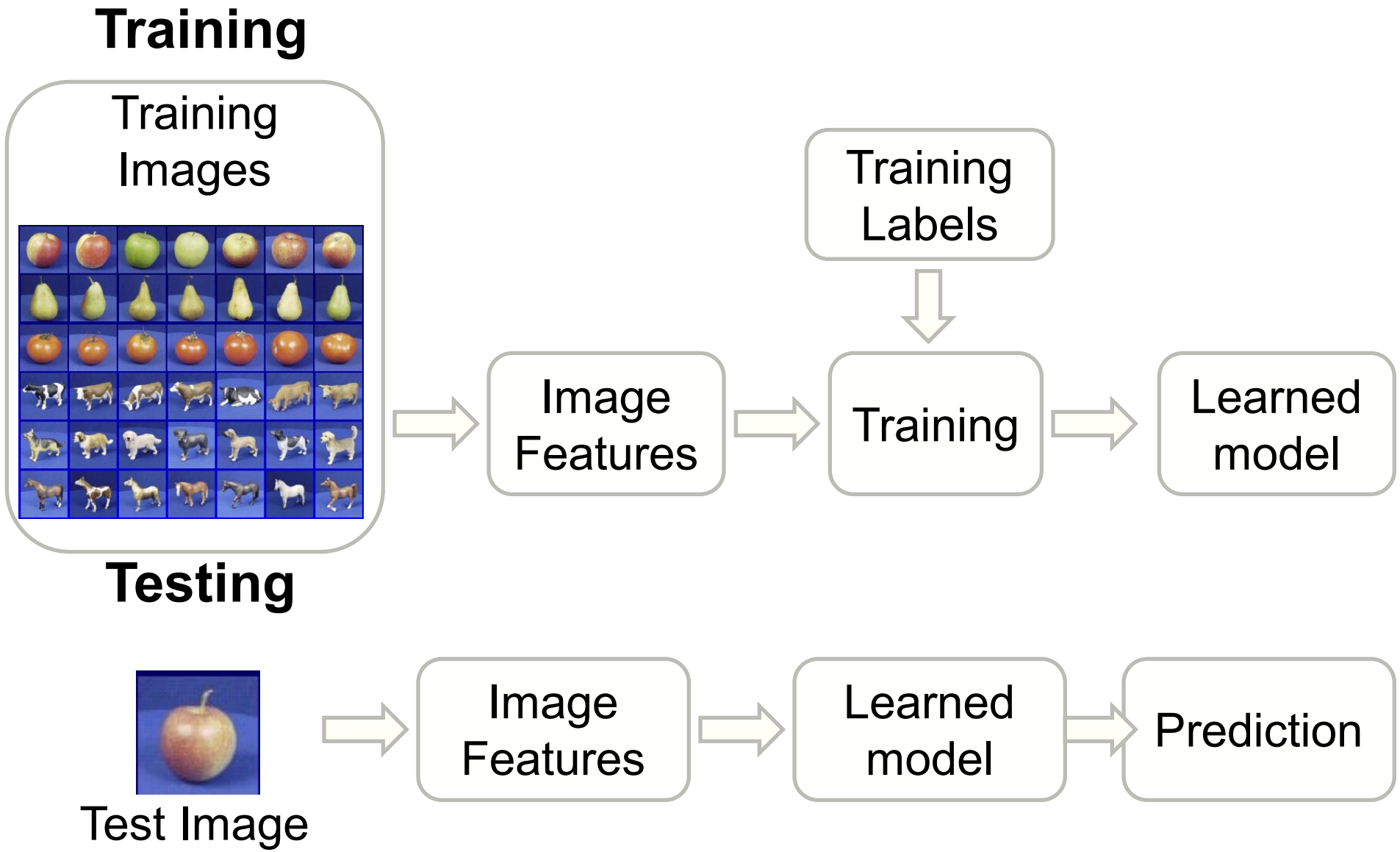
  …

机器学习的一般过程
Machine Learning Process



- Basic assumption: there exist the same patterns across training and test data

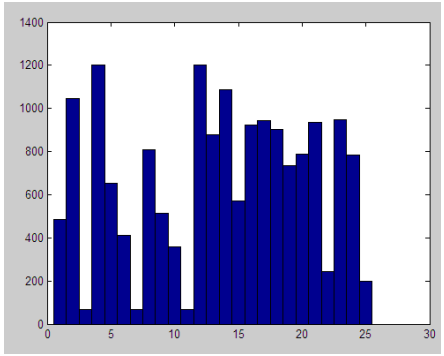
图像分类问题举例
Image Classification Example



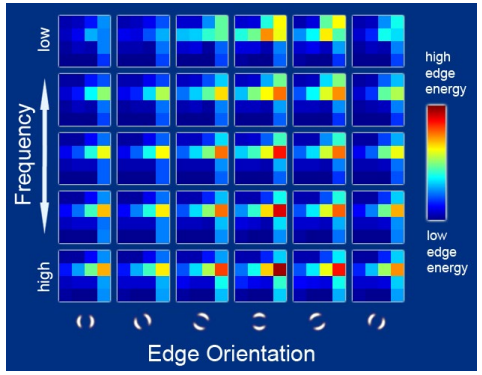
图像分类问题举例
Image Classification Example

Features

- Raw pixels



- Histogram



- GIST descriptors

- …

图像分类问题举例
Image Classification Example

- Apply a prediction function to a feature representation of the image to get the desired output:

f() = “apple”

f() = “tomato”

f() = “cow”

图像分类问题举例
Image Classification Example

$$y = f(\mathbf{x})$$

Output label

prediction function

Input instance

- **Training:** given a *training set* of labeled examples $\{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}) \ldots, (\boldsymbol{x}^{(\mathrm{N})}, \boldsymbol{y}^{(\mathrm{N})})\}$, estimate the prediction function f by minimizing the prediction error on the training set

- **Testing:** apply f to a never before seen test example **x** and output the predicted value y = f(**x**)

监督学习
Supervised Learning

- Given the training dataset of (data,label) pairs,

$$D = \{(x^{(i)}, y^{(i)})\}_{i=1.2,\ldots,N}$$

let the machine learn a function from data to label

$$y^{(i)} \approx f_\theta(x^{(i)})$$

- Function set $\{f_\theta(x^{(i)})\}$ is called hypothesis space
- Learning is referred to as updating the parameter θ to make the prediction closed to the corresponding label

监督学习
Supervised Learning

- Given the training dataset of (data,label) pairs,

$$D = \{(x^{(i)}, y^{(i)})\}_{i=1.2,\ldots,N}$$

$x^{(i)}$= input data(features) of i^{th} training example
$y^{(i)}$= output data(label) of i^{th} training example

let the machine learn a function from data to label

$$y^{(i)} \approx f_\theta(x^{(i)})$$

- Function set $\{f_\theta(x^{(i)})\}$ is called hypothesis space
- Learning is referred to as updating the parameter θ to make the prediction closed to the corresponding label

监督学习
Supervised Learning

- Given the training dataset of (data,label) pairs,

$$D = \{(x^{(i)}, y^{(i)})\}_{i=1.2,\ldots,N}$$

$x^{(i)}$= input data(features) of i^{th} training example
$y^{(i)}$= output data(label) of i^{th} training example

let the machine learn a function from data to label

$$y^{(i)} \approx f_\theta(x^{(i)})$$

- Function set $\{f_\theta(x^{(i)})\}$ is called hypothesis space
- Learning is referred to as updating the parameter θ to make the prediction closed to the corresponding label

1. What is the learning objective?
2. How to update the parameters?

问题1：学习目标
What is the Learning Objective?

- Make the prediction closed to the corresponding label

(Expected Risk/Error Minimization)

Cannot be directly calculated!

问题1：学习目标
What is the Learning Objective?

- Make the prediction closed to the corresponding label

$$\min_{\theta} \quad \frac{1}{N}\sum_{i=1}^{N} L(y^{(i)}, f_\theta(x^{(i)}))$$

(Empirical Risk Minimization, ERM)

Loss functior $L(y^{(i)}, f_\theta(x^{(i)}))$ measures the error between the label and prediction for single sample.

The definition of loss function depends on the data and task

损失函数
Loss function

0-1 Loss Function

$$L\left(y^{(i)}, f\left(x^{(i)}\right)\right) = \begin{cases} 1, & \text{if } y^{(i)} \neq f\left(x^{(i)}\right) \\ 0, & \text{if } y^{(i)} = f\left(x^{(i)}\right) \end{cases}$$

Mean Squared Error, MSE

$$L\left(y^{(i)}, f\left(x^{(i)}\right)\right) = \left(y^{(i)} - f\left(x^{(i)}\right)\right)^2$$

Absolute Loss Function

$$L\left(y^{(i)}, f\left(x^{(i)}\right)\right) = \left|y^{(i)} - f\left(x^{(i)}\right)\right|$$

Logarithmic Loss Function (Cross-Entropy Loss Function)

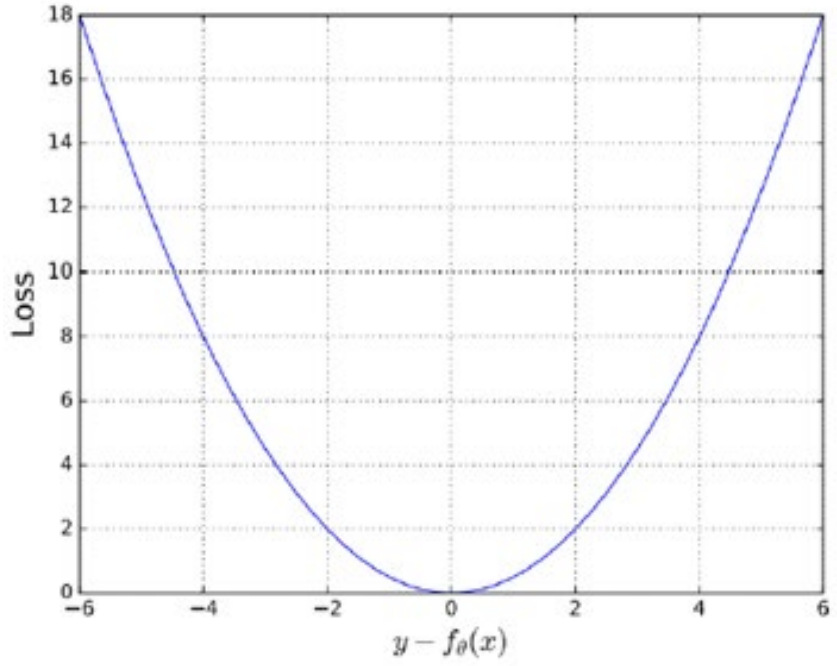
$$L\left(y^{(i)}, p^{(i)}\right) = -\left[y^{(i)} \log\left(p^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - p^{(i)}\right) \right]$$

$y^{(i)} \in \{0,1\}$, $p^{(i)} = f\left(x^{(i)}\right)$ is the predicted probability that the i th sample belongs to the positive class (usually denoted as class 1))

损失函数
Loss function

Most popular loss function: squared loss

$$L(y^{(i)}, f_\theta(x^{(i)})) = \frac{1}{2}(y^{(i)} - f_\theta(x^{(i)}))^2$$



*Penalty much more on larger distances

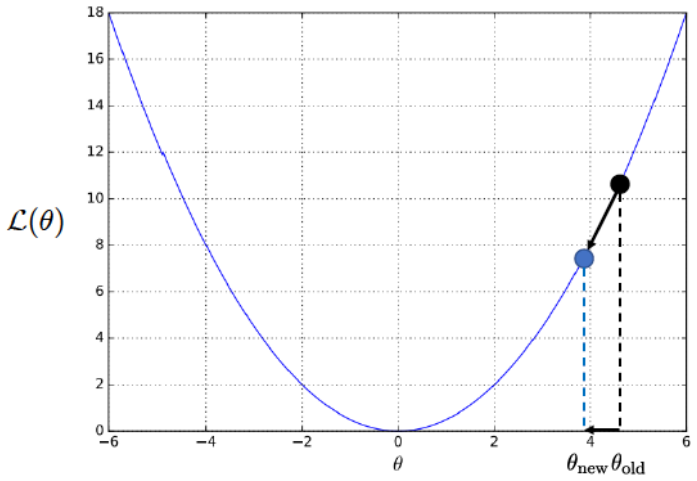
*Accept small distance (error)
.Observation
noise etc.
.Generalization

问题2：如何更新参数
How to update the parameters?

Given θ, we have code that can compute

- $J(\theta)$
- $\frac{\partial J(\theta)}{\partial \theta_i}$ (for θ_i j=0，1，...,n)

Gradient learning method:



$$\theta_{\text{new}} \leftarrow \theta_{\text{old}} - \eta \frac{\partial \mathcal{L}(\theta)}{\partial \theta}$$

Other Optimization algorithms:

- Newton method
- Quasi-Newton method
- Conjugate gradient method

超参数
Hyperparameters

Hyperparameters are settings in machine learning models that are set before training begins and are not learned from the data. They can significantly affect model performance.

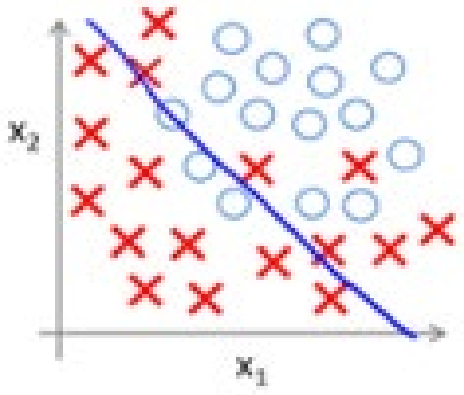
Key hyperparameters include:

- **Learning Rate**: Controls how quickly the model learns.
- **Regularization Coefficient：** Prevents overfitting by penalizing complex models.
- **Batch Size**: Number of training samples used in one update.
- **Number of Epochs**: How many times the entire dataset is passed through the model.
- **Model Architecture**: Includes the number of layers and neurons in neural networks.
- **Optimizer**: Algorithm used to minimize the loss function.
- …

超参数优化方法
Hyperparameter Optimization Methods

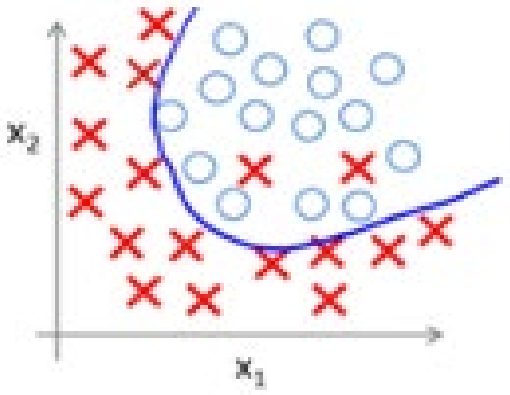
- **Grid Search**
  Exhaustively explores all possible hyperparameter combinations
- **Random Search**
  Randomly samples hyperparameter combinations
- **Bayesian Optimization**
  Builds a surrogate model to guide hyperparameter selection
- **Evolutionary Algorithms**
  Mimics natural selection to optimize hyperparameters
- **Gradient-based Optimization**
  Uses gradient information to optimize differentiable hyperparameters
- **Hyperparameter Optimization Libraries**
  Tools like Optuna, Hyperopt, and Ray Tune support various optimization methods

模型选择
Model Selection

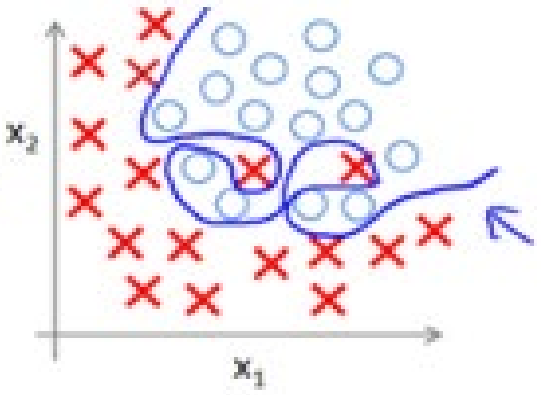


$f_\theta(x)$
$= g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

g =sigmoid function



$f_\theta(x)$
$= g(\theta_0 + \theta_1 x_1 + \theta_2 x_2$
$+ \theta_3 {x_1}^2 + \theta_4 {x_2}^2 + \theta_5 x_1 x_2$



$f_\theta(x)$
$= g(\theta_0 + \theta_1 x_1 + \theta_2 {x_1}^2$
$+ \theta_3 {x_1}^2 x_2 + \theta_4 {x_1}^2 {x_2}^2$
$+ \theta_5 {x_1}^2 {x_2}^3 + + \theta_6 {x_1}^3 x_2 + \cdots)$

Linear model: underfitting Quadratic model: well fitting Sth.order model: overfitting

- Underfitting occurs when a statistical model or machine learning algorithm cannot capture the underlying trend of the data.
- Overfitting occurs when a statistical model describes random error or noise instead of the underlying relationship

模型选择
Model Selection

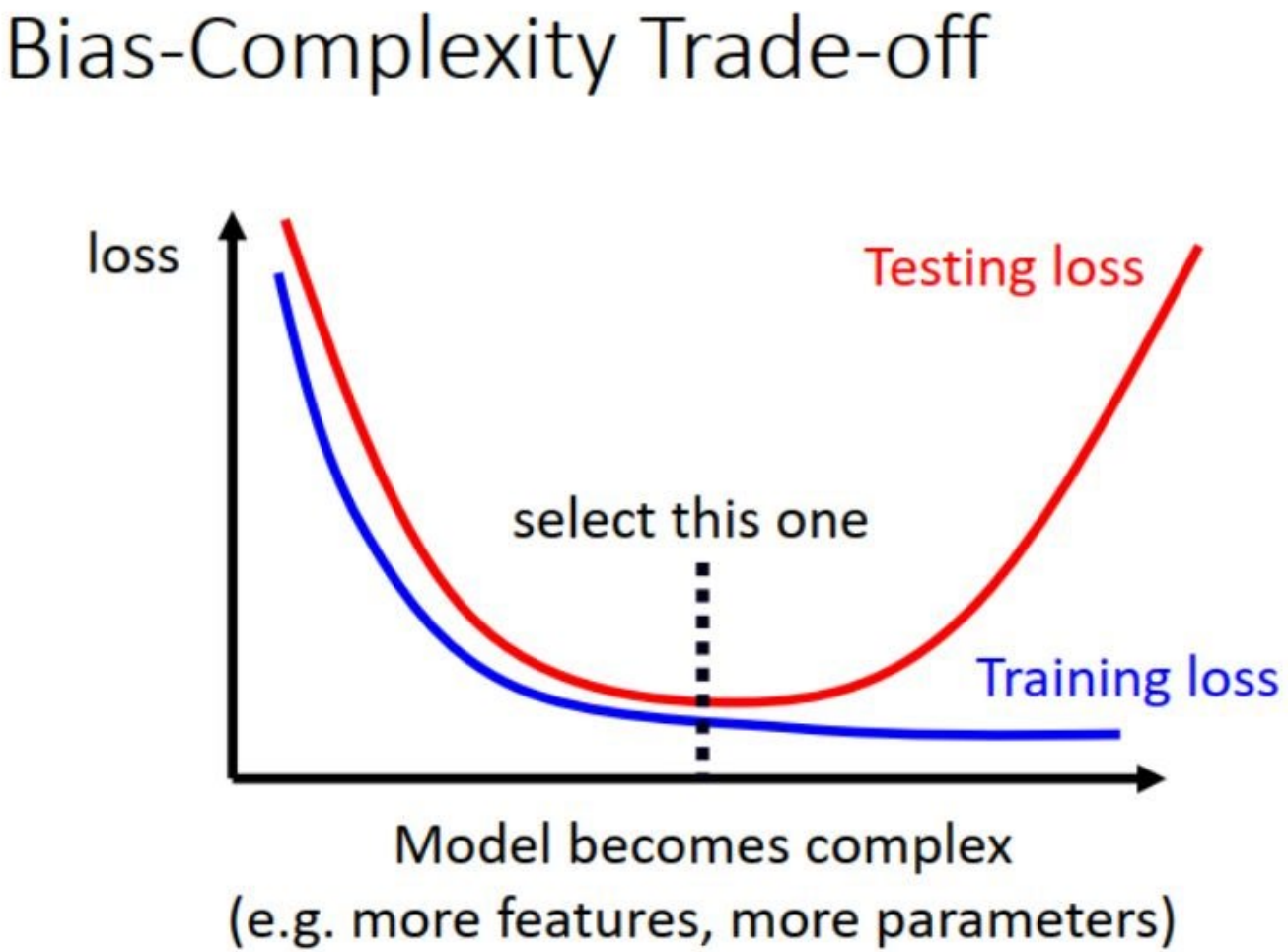
Training data

Test data

overfitting

underfitting

模型选择
Model Selection



奥卡姆剃刀原则
Principle of Occam's razor

Among competing hypotheses, the one with the fewest assumptions should be selected.

Recall the function set $\{f_\theta(x^{(i)})\}$ is called hypothesisspace

$$\min_\theta \frac{1}{N}\sum_{i=1}^{N} L(y^{(i)}, f_\theta(x^{(i)})) + \lambda\Omega(\theta)$$

Original loss　　Penalty on assumptions

Structural Risk Minimization, SRM

正则化方法
Regularization

L2-Norm (Ridge):

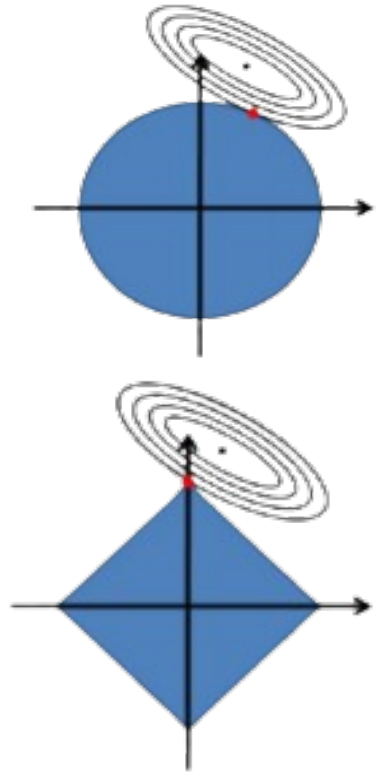
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y^{(i)}, f_{\theta}(x^{(i)})) + \lambda \|\theta\|^2$$

L1-Norm (LASSO):

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y^{(i)}, f_{\theta}(x^{(i)})) + \lambda |\theta|$$

Elastic Net：

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y^{(i)}, f_{\theta}(x^{(i)})) + \lambda |\theta| \quad + \quad (1-\lambda) \|\theta\|^2$$



损失函数&代价函数&目标函数
Loss function &Cost function &Objective function

Loss Function:

$$L(y^{(i)}, f_\theta(x^{(i)}))$$

Cost Function:

$$\frac{1}{N}\sum_{i=1}^{N} L(y^{(i)}, f_\theta(x^{(i)}))$$

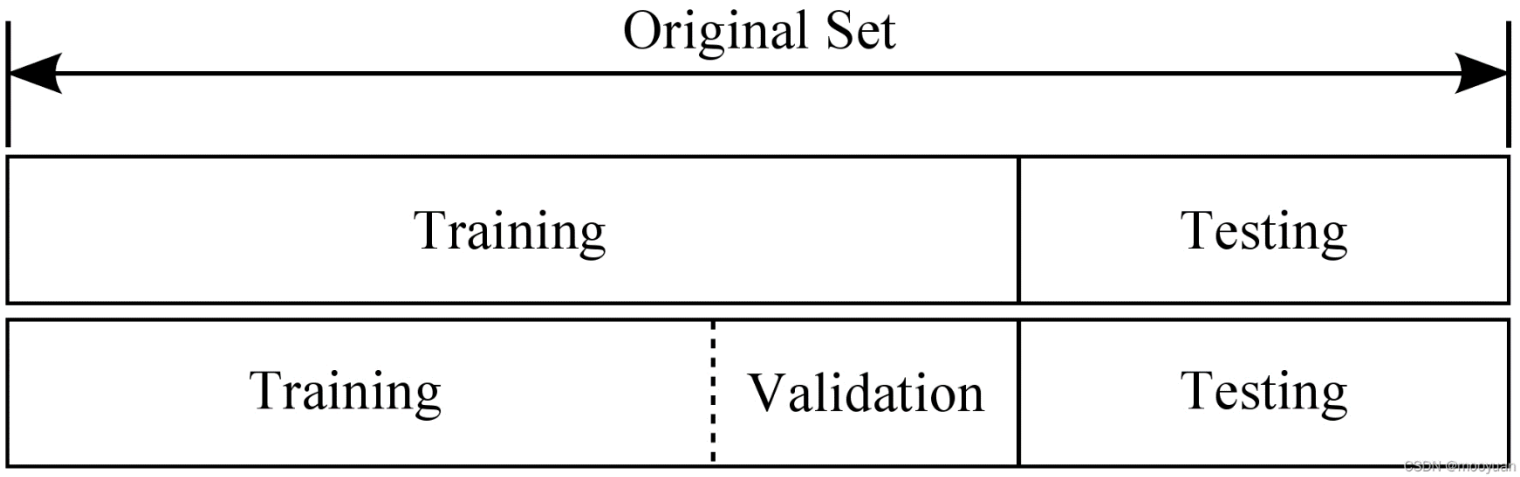
Objective Function:

The function that the training process aims to minimize or maximize.

$$\frac{1}{N}\sum_{i=1}^{N} L(y^{(i)}, f_\theta(x^{(i)})) + \lambda\|\theta\|^2$$

$$\frac{1}{N}\sum_{i=1}^{N} L(y^{(i)}, f_\theta(x^{(i)})) + \lambda|\theta|$$

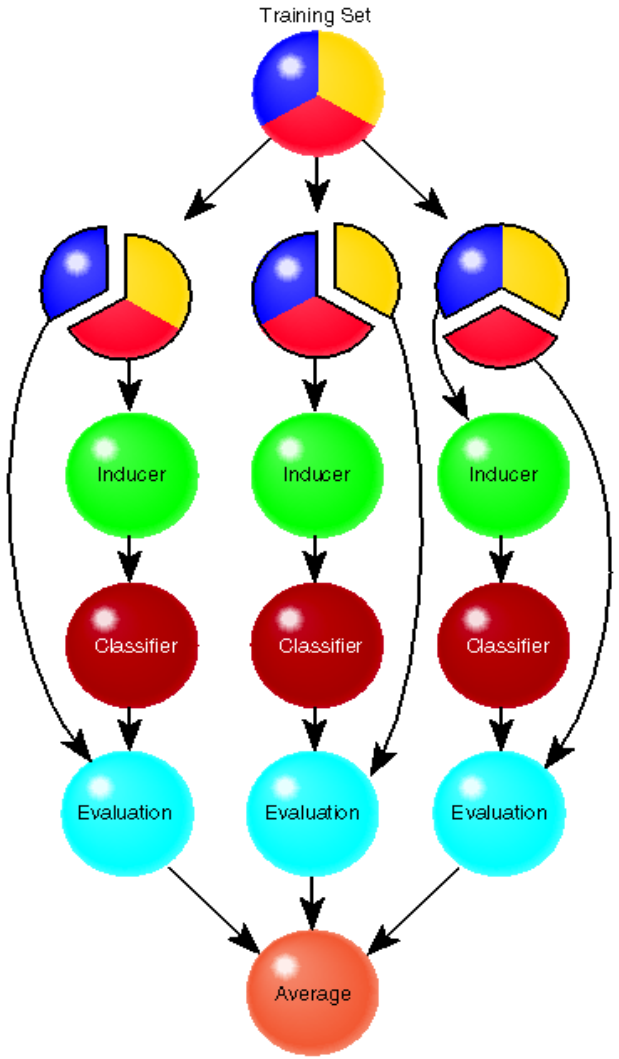
留出法
hold-out for model selection



training set 60% validation set 20% testing set 20%

Split training data into
training set and validation set
for model selection

交叉验证
Cross validation for model selection



K-fold cross validation

CV(data S, alg L, int k)
  Divide S into k disjoint sets { S_1, S_2, …, S_k }
  For i = 1..k do
    Run L on $S_{-i} = S - S_i$
    obtain $L(S_{-i}) = h_i$
    Evaluate h_i on S_i
      $err_{S_i}(h_i) = 1/|S_i| \sum_{\langle x,y \rangle \in S_i} I(h_i(x) \neq y)$
  Return Average $1/k \sum_i err_{S_i}(h_i)$

leave-one-out cross validation (K=|S|)

偏差和方差
Bias &variance

The true value of y is given by a function f(x)plus some noise ϵ: $y = f(\mathrm{x}) + \epsilon$

偏差和方差
Bias &variance

The true value of y is given by a function f(x)plus some noise ϵ: $y = f(\mathrm{x}) + \epsilon$

The mean squared error (MSE) of the model $\hat{f}(\mathrm{x})$:

$= E\left[(\hat{f}(\mathrm{x}) - y)^2\right]$

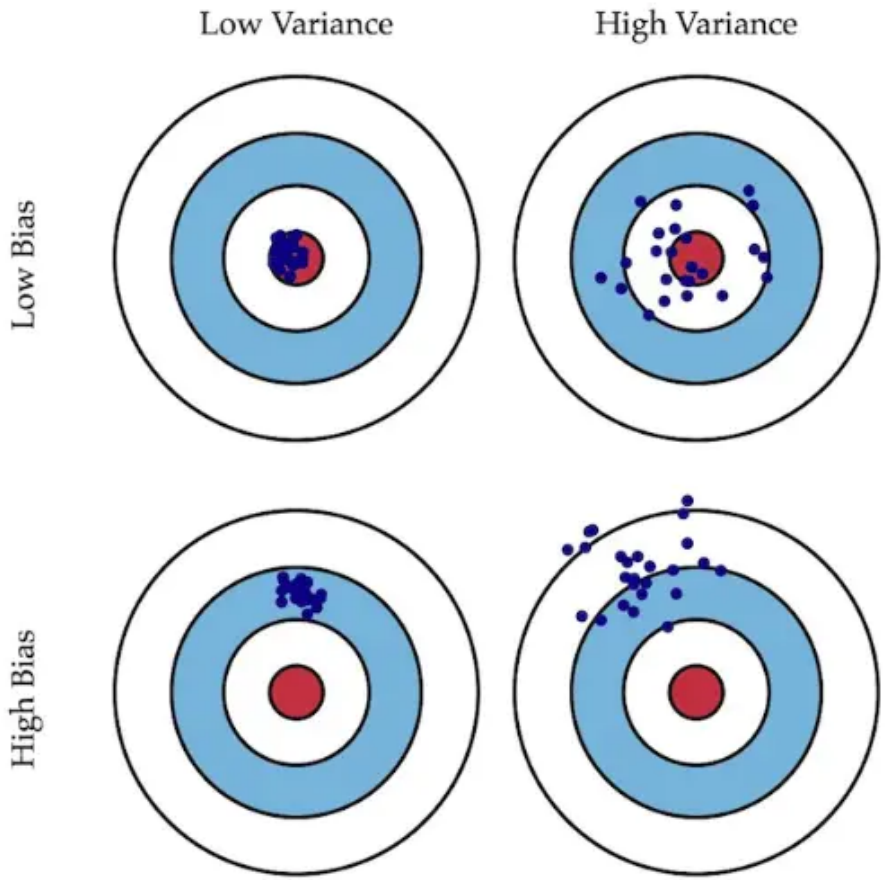
偏差和方差
Bias &variance

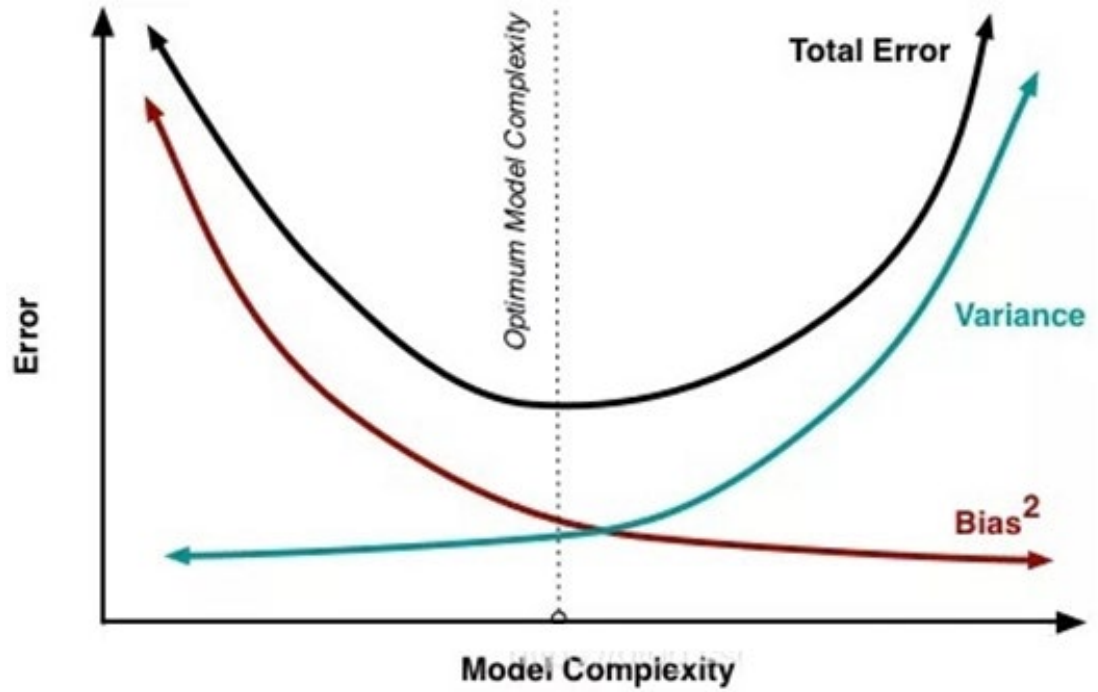
The true value of y is given by a function f(x)plus some noise ϵ: $y = f(\mathrm{x}) + \epsilon$

The mean squared error (MSE) of the model $\hat{f}(\mathrm{x})$: $\epsilon \sim \mathrm{N}(0, \sigma^2)$

$$= E\left[(\hat{f}(\mathrm{x}) - y)^2\right]$$
$$= E\left[(\hat{f}(\mathrm{x}) - f(\mathrm{x}) - \epsilon)^2\right]$$
$$= E\left[(\hat{f}(\mathrm{x}) - f(\mathrm{x}))^2\right] - 2E\left[\left(\hat{f}(\mathrm{x}) - f(\mathrm{x})\right)\epsilon\right] + E[\epsilon^2)]$$
$$= E\left[(\hat{f}(\mathrm{x}) - f(\mathrm{x}))^2\right] + E[\epsilon^2)]$$
$$= E\left[(\hat{f}(\mathrm{x}) - f(\mathrm{x}))^2\right] + \sigma^2$$
$$= E\left[(\hat{f}(\mathrm{x}) - E\left[\hat{f}(\mathrm{x})\right] + E\left[\hat{f}(\mathrm{x})\right] - f(\mathrm{x}))^2\right] + \sigma^2$$
$$= E\left[(\hat{f}(\mathrm{x}) - E\left[\hat{f}(\mathrm{x})\right])^2\right] + E\left[(E\left[\hat{f}(\mathrm{x})\right] - f(\mathrm{x}))^2\right] + 2E\left[\left(\hat{f}(\mathrm{x}) - E\left[\hat{f}(\mathrm{x})\right]\right)\left(E\left[\hat{f}(\mathrm{x})\right] - f(\mathrm{x})\right)\right] + \sigma^2$$
$$= E\left[(\hat{f}(\mathrm{x}) - E\left[\hat{f}(\mathrm{x})\right])^2\right] + E\left[(E\left[\hat{f}(\mathrm{x})\right] - f(\mathrm{x}))^2\right] + \sigma^2$$
$$= \boldsymbol{Variance} + \boldsymbol{Bias^2} + \boldsymbol{\sigma^2}$$

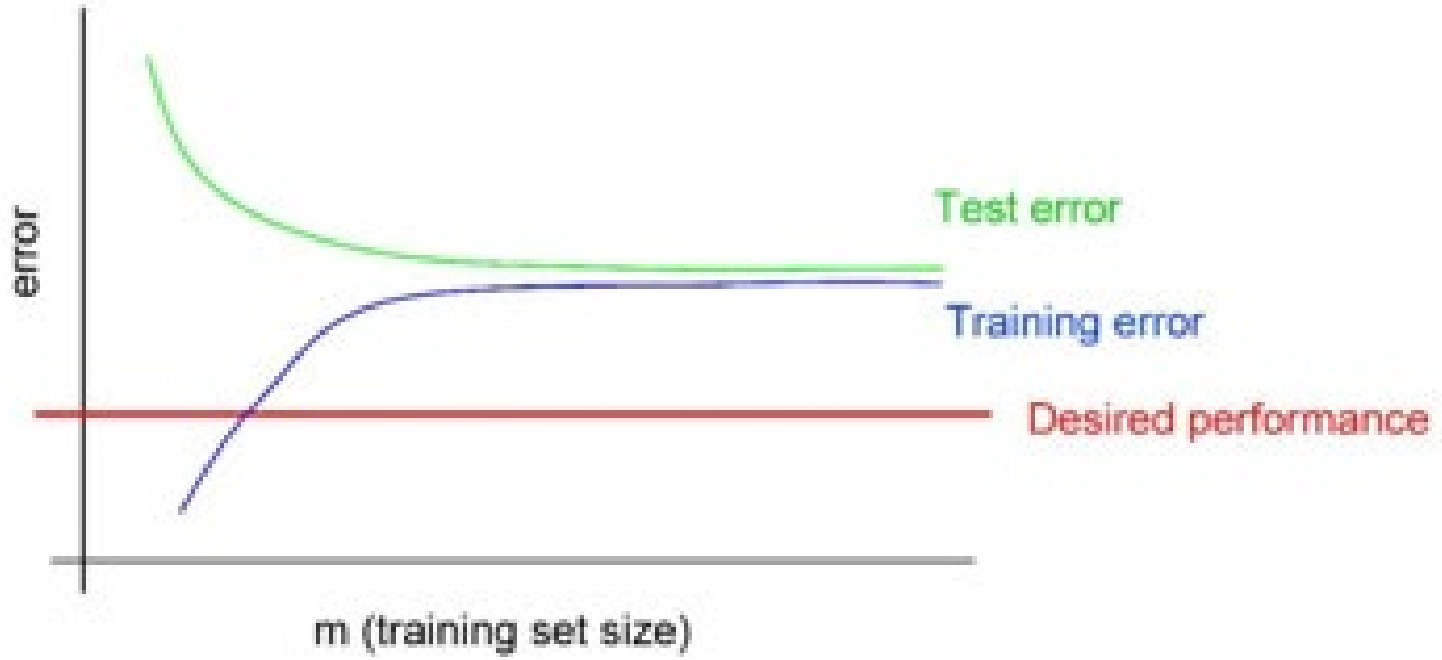
偏差和方差
Bias &variance





偏差和方差
Bias &variance

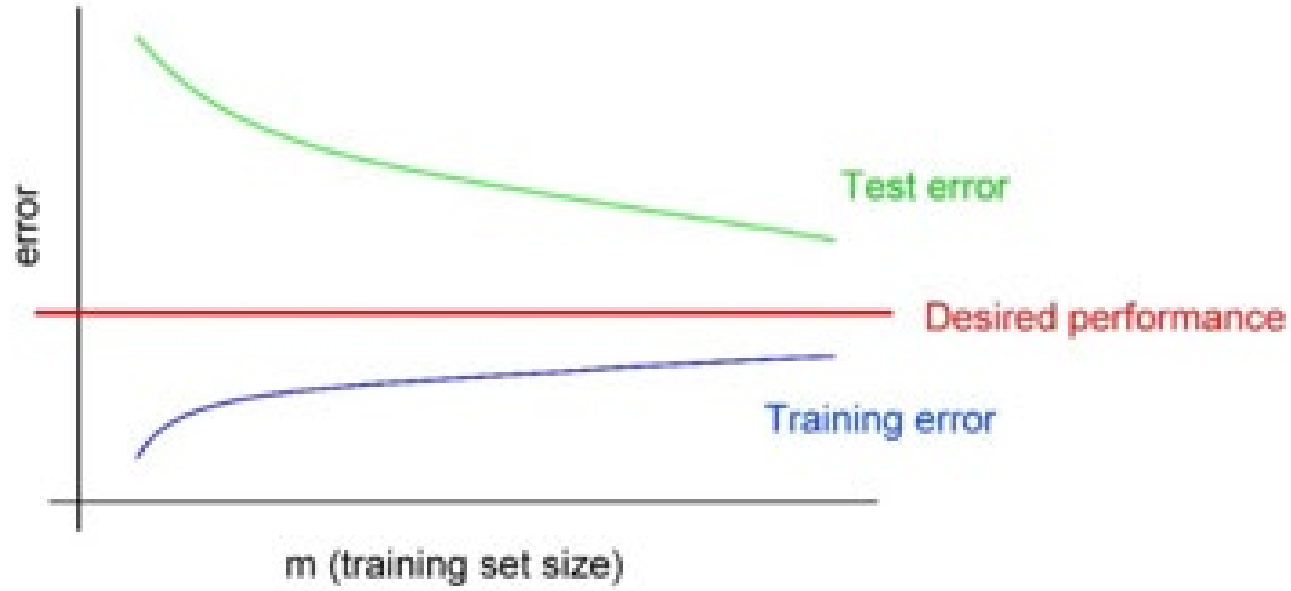
Typical learning curve for high bias:



- Even training error is unacceptably high.
- Small gap between training and test error.

偏差和方差
Bias &variance

Typical learning curve for high variance:



- Test error still decreasing as m increases. Suggests larger training set will help.
- Large gap between training and test error.

如何应对欠拟合和过拟合
How to fix underfitting & overfitting

Fixes to underfitting try:

Fixes to overfitting try:

如何应对欠拟合和过拟合
How to fix underfitting & overfitting

Fixes to underfitting try:

- Increase model complexity
- Try a larger set of features
- Reduce regularization

Fixes to overfitting try:

- Reduce model complexity
- Try getting more training examples
- Increase regularization
- Try a smaller set of features
- Ensemble learning

评价标准
Evaluation metrics

- For regression tasks:
  - Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean absolute Error (MAE), R-Squared (R2)

- For classification tasks:
  - Accuracy, True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN), …
- For multi-classification tasks:
  - Take each individual class as the positive class, and the rest as the negative class one by one

回归评价标准
Regression Evaluation metrics

MSE(Mean Squared Error)
均方误差

$$\frac{1}{N}\sum_{i=1}^{N}(y^{(i)}-f(x^{(i)}))^2$$

RMSE（Root Mean Squared Error）
均方根误差

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y^{(i)}-f(x^{(i)}))^2}$$

MAE (Mean absolute Error)
平均绝对误差

$$\frac{1}{N}\sum_{i=1}^{N}\left|(y^{(i)}-f(x^{(i)}))^2\right|$$

R-Squared（r2score）R方/决定系数

$$= 1-\frac{\sum_{i=1}^{N}(y^{(i)}-f(x^{(i)}))^2}{\sum_{i=1}^{N}(y^{(i)}-\bar{y}))^2}$$

$$= 1-\frac{MSE}{Var}$$

分类性能评价
Classification Evaluation Metrics

Confusion matrix

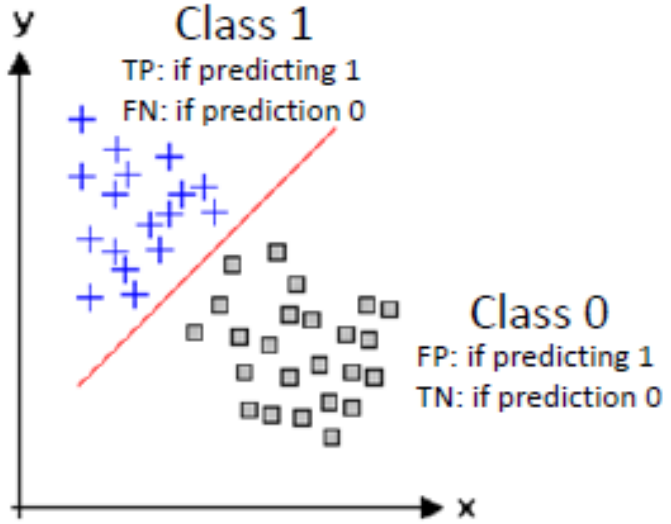
Actual classes \ Predicted classes	1	0
1	True Positive	False Negative
0	False Positive	True Negative

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

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

$$\text{Recall} = \frac{TP}{TP + FN}$$

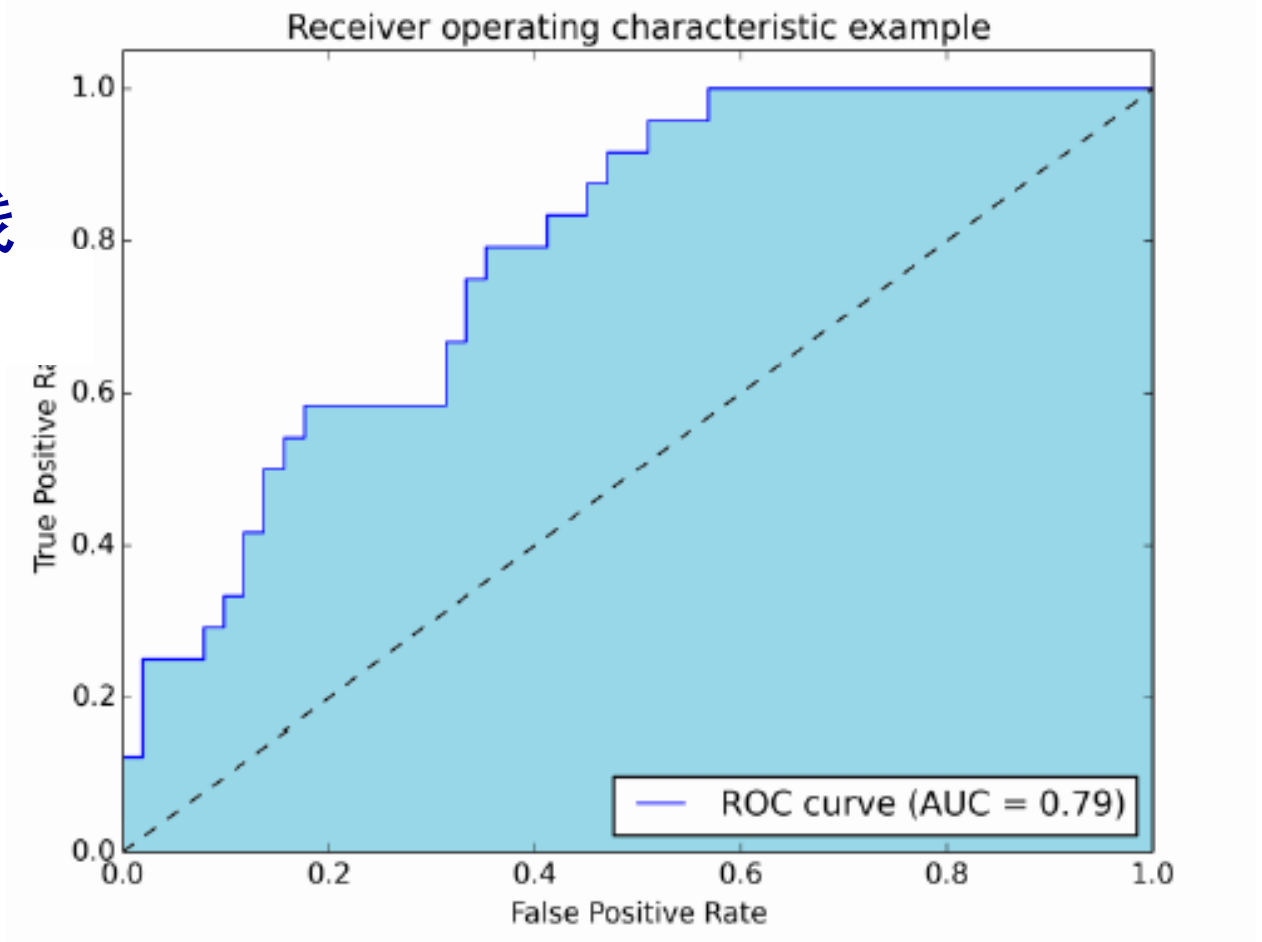
$$\text{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$



分类性能评价
Classification Evaluation Metrics

- Ranking-based measure: Area Under ROC Curve (AUC)

分类性能评价： ROC曲线
Receiver Operating Characteristic



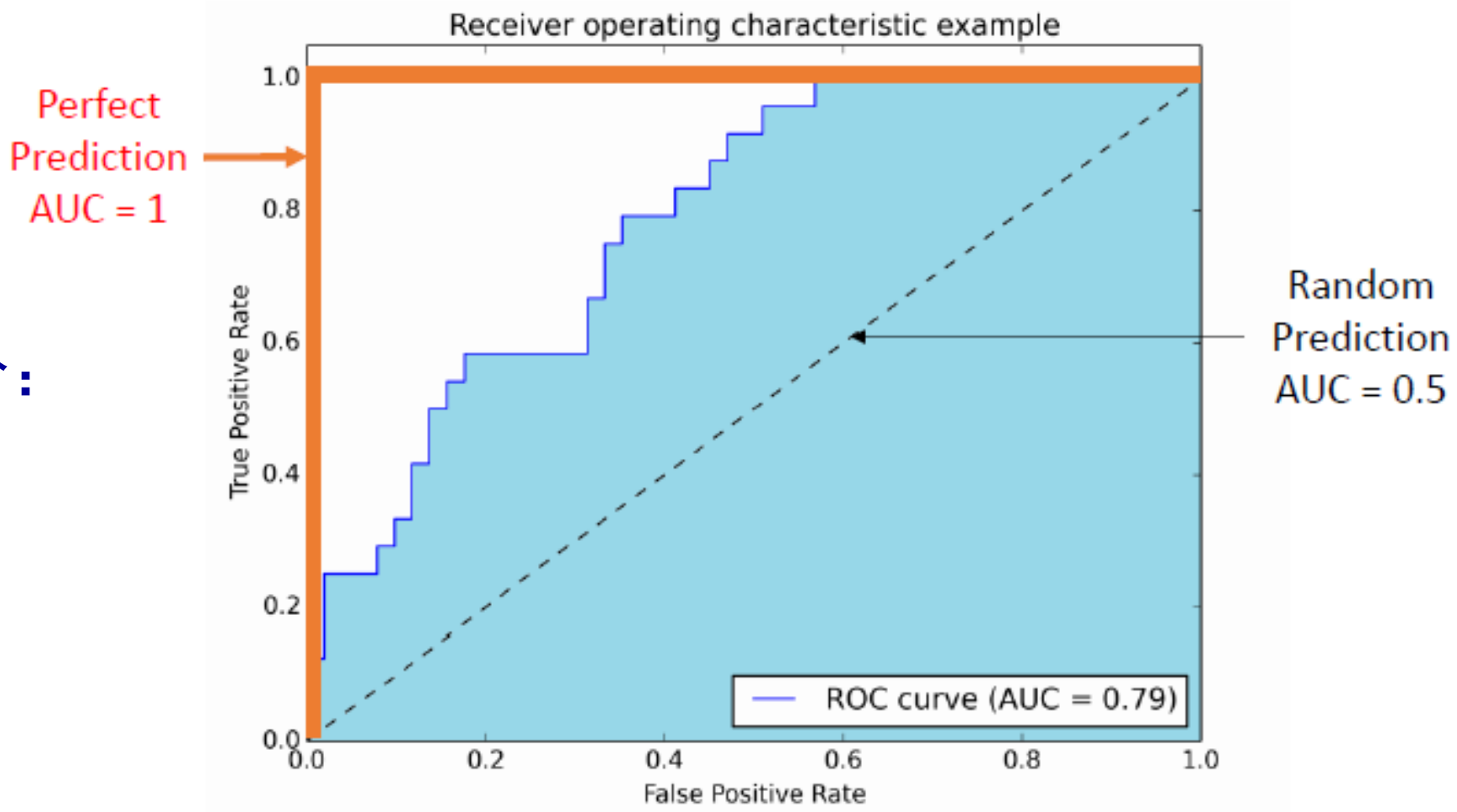
True Positive Rate

$$\mathrm{TPR} = \frac{TP}{TP + FN}$$

False Positive Rate $\mathrm{FPR} = \frac{FP}{TN + FP}$

分类性能评价
Classification Evaluation Metrics

- Ranking-based measure: Area Under ROC Curve (AUC)



分类性能评价：
AUC 面积
Area under
Curve

监督学习
Supervised Learning

- Given the training dataset of (data,label) pairs,

$$D = \{(x^{(i)}, y^{(i)})\}_{i=1.2,\ldots,N}$$

$x^{(i)}$= input data(features) of i^{th} training example
$y^{(i)}$= output data(label) of i^{th} training example

let the machine learn a function from data to label

$$y^{(i)} \approx f_\theta(x^{(i)})$$

- Function set $\{f_\theta(x^{(i)})\}$ is called hypothesis space
- Learning is referred to as updating the parameter θ to make the prediction closed to the corresponding label

机器学习和假设空间
Machine Learning & Hypothesis Space

- One way to think about a supervised learning machine is as a device that explores a “hypothesis space”.

- The art of supervised machine learning is in:
  - Deciding how to represent the inputs and outputs
  - Selecting a hypothesis space that is powerful enough to represent the relationship between inputs and outputs but simple enough to be searched.

- **Different learning methods** assume different hypothesis spaces (representation languages) and/or employ different search techniques.

函数表征
Various Function Representations

- Numerical functions
  - Linear regression
  - Neural networks
  - Support vector machines
- Symbolic functions
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic
- Instance-based functions
  - Nearest-neighbor
  - Case-based
- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

搜索算法
Various Search Algorithms

- Gradient descent
  - Perceptron
  - Backpropagation
- Dynamic Programming
  - HMM Learning
  - PCFG Learning
- Divide and Conquer
  - Decision tree induction
  - Rule learning
- Evolutionary Computation
  - Genetic Algorithms (GAs)
  - Genetic Programming (GP)
  - Neuro-evolution

课程安排
Course Arrangement

- Course content：
  - Linear models (regression)
  - Linear models (classification);
  - Decision tree ;
  - Artificial neural network;
  - Deep learning;
  - SVM;
  - Bayesian Learning
  - Ensemble learning;
  - Clustering
  - Dimension reduction& feature selection
  - Frontier Introduction

- lab assignments（3人1组）：regression; classification; clustering

课程考核
Course Assessment

- **Class attendance & Quize 15%**
  - Class quizzes **MUST** be handed to the teacher after each class.
- **Project (3 times) 45%**
  - Submission Instructions:
    - Submit to: canvas
    - Subject: **Student ID_Name_MLProject_X** (replace X with 1, 2, or 3)
    - File Name: **Student ID_Name_MLProject_X.zip** (replace X with 1, 2, or 3)
  - **Projects Exemption**: Excellent academic paper achievements
- **Final Exam 40%**

参考书

- 周志华. “机器学习”. 清华大学出版社, 2016
- 李航，统计学习方法，清华大学出版社，2019
- TOM M MICHELLE. Machine Learning[M]. New York: McGraw-Hill Companies,Inc,1997.
- Andrew Ng. Machine Learning[EB/OL]. StanfordUniversity,2014.https://www.coursera.org/course/ml
- Bishop 的Pattern Recognization and Maching Learning，简称为PRML。中文译本为《模式识别与机器学习》

周志华，南京大学计算机科学与技术系主任 、人工智能学院院长。
代表作：《机器学习》（西瓜书）

吴恩达（Andrew Ng），斯坦福大学副教授，前“百度大脑”的负责人与百度首席科学家。

李航，现任字节跳动科技有限公司人工智能实验室总监，北京大学、南京大学客座教授，IEEE 会士，ACM 杰出科学家，CCF 高级员。
代表作：《统计学习方法》

Christopher M. Bishop，是机器学习和统计学领域的著名学者之一。
《Pattern Recognition and Machine Learning》被广泛视为机器学习领域的经典教材。

预备知识

- 线性代数
- 微积分
- 数学优化
- 概率论
- 信息论

https://nndl.github.io/

《数学基础》

Resources: Datasets

- 通用性分析：
  - UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
  - UCI KDD Archive: http://kdd.ics.uci.edu/summary.data.application.html
  - Statlib: http://lib.stat.cmu.edu/
  - Delve: http://www.cs.utoronto.ca/~delve/
- 探索性分析：Kaggle数据集：https://www.kaggle.com/datasets
- 深度学习：
  - Deeplearning.net - 用于对深度学习算法进行基准测试的最新数据集列表
  - DeepLearning4J.org - 用于深度学习研究的高质量数据集的最新列表
  - https://paperswithcode.com/datasets

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

Resources: Conferences

- International Conference on Machine Learning (ICML)
- Neural Information Processing Systems (NIPS→NeurIPS)
- Computational Learning Theory (COLT)
- European Conference on Machine Learning (ECML)
- Asian Conference on Machine Learning (ACML)
- IEEE Conference on Computer Vision and Pattern Recognition（CVPR）
- AAAI Conference on Artificial Intelligence，AAAI）:
- 1（International Joint Conference on Artificial Intelligence，IJCAI）

- 中国机器学习大会(CCML)
- 机器学习及其应用 (MLA)

问题

- 试分析什么因素会导致模型出现图中所示的高偏差和高方差情况.