

吴恩达机器学习课程及多版本笔记纲要

Jan 4, 2019 at 6:50 PM

by 深度碎片

[获取最新PDF版本](#)

纲要汇总所有吴恩达机器学习的原视频和笔记视频链接，并对每个视频提供了一句话版的总结

原课程和笔记专辑的视频链接汇总

[吴恩达机器学习原课程视频无字幕 B站](#)

[第一版笔记视频专辑 2018.4.29完结](#)

[探索一句话版的机器学习和深度学习](#)

以下是纲要内容

为什么要做这个视频专辑

2019.1.1 开启图解吴恩达机器学习（直觉版）学习

[为什么要做直觉版笔记，以及如何使用这个版本：视频p1](#)

直觉版笔记 Keynote



ML for everyone.key
Jan 2, 2019

what is machine learning? 什么是机器学习

[原版课程视频](#)

给出了2位教授给出的简明扼要的机器学习的定义（值得反复回味）

[第一版笔记视频](#)

用国际象棋例子理解机器学习的工作原理（现在看来，当时自己理解比较粗糙）

直觉版理解：如何定义机器学习

many tasks cannot be done through programming software explicitly, but can be learnt from data with learning algorithms guided by objective function, and speed up by

optimization function.

许多任务无法通过直接人工手动编程实现，但可以通过数据和学习算法在目标函数的指导下，结合优化算法加速训练，让计算机自己找到解决方案。

直觉版笔记视频

P2 02 为什么要让机器去学习

机器学习的诞生，是人类懒惰的结晶

P3 03 为什么机器必须要学习

机器学习是在模拟人类学习的过程

P4 04 机器学习的流程长什么样子

尝试找到一个统领所有机器学习模型的流程框架

what is supervised learning? 什么是监督学习?

原版课程视频

通过预测房价和肿瘤性质两个例子，分别介绍监督学习算法下的回归与分类方法的模样
明确指出是否有 y ，是决定是否是监督学习的defining factor 决定性因素

第一版笔记视频

紧扣原课程内容，结合2个例子进一步阐述监督学习的定义，以及回归与分类的区别

直觉版：如何理解监督的由来

如何理解监督的学习? regression 和 classification

如果“监督”是因为 y 的存在，那么一种对“监督”的理解可以是： y 作为目标值，是模型要预测的结果，是目标函数的核心组件，指导着模型的每一步更新

what is unsupervised learning? 什么是无监督学习?

原版课程视频

点明区别无监督就是没有 y ，给出了大量实例说明无监督是在做clustering做划分区分归类（不同于分类，分类是有 y 的）。

直觉版笔记视频

紧扣原课程，利用课程中的实例，阐明基于 x （没有 y ）来实现clustering的过程中所实现现实价值。

直觉版：如何理解无监督学习？相似远大于不同

如何理解无监督学习

无监督学习，不知道 y 的存在，所以不追寻 y ；无监督学习只拥有 x ，通过研究 x 数据结构，追寻对 x 做区分分类的目标，遵循统一的机器学习工作流程。

model representation 什么是回归模型？

原版课程视频

用房价预测介绍了线性回归 linear regression的构建过程；将模型 H （hypothesis）理解为 x 到 y 的映射）；单一参数版本uni-variant，和多参数版本multi-variant regression；

第一版笔记

突出阐述对模型 H 的映射理解：将大量廉价知识 X ，映射成昂贵不易获取的知识 Y ，这是监督学习的 H ；将大量廉价知识 X ，映射成不同归类的 X （形成新知识），这是无监督学习的 H

直觉版：机器学习流程图的务实功用

机器学习流程图的务实功用

机器学习流程图可以帮助我们分析所有学习算法和模型，也有助于对各种机器学习领域里的模型算法做归类区分

直觉版：回归模型长什么样子

回归模型长什么样子

套用机器学习流程图，结合吴恩达预测房屋价格的简单案例，了解回归模型的数据，学习算法，目标函数的组成构造

Cost function 损失函数

原课程视频

对损失函数的求解做热身：解读简单线性函数模型的目标函数（损失函数），以及损失函数的目的所在

第一版笔记视频

在原课程基础上，尝试做更细致的损失函数意义的解读

cost function Intuition I 损失函数直觉 I

原课程视频

通过对比平行两张图，参数与模型的关系（左图），参数与损失函数关系（右图），来感知其实训练就是摸索一个最佳参数是的损失值最小

第一版视频笔记

尝试简化左右两张图的理解

cost function Intuition II 损失函数直觉 II

原课程视频

线性函数模型的参数从1个升级到2个，参数与损失函数关系图（右图）成为3D图，映射到2D平面后是contour图。如何在contour图上解读参数与损失值的变化关系。

第一版笔记

尝试用直白语言将左右两张图的内涵和细节说清楚

直觉版：体会损失函数的困境

- 我们已知：损失函数为模型提供评估好坏的标准度量，损失函数的目标是要最小化（这里是0），最小化对应的参数能告诉我们最优模型的样子；
- 我们要尝试体会损失函数追寻最小化目标的艰难：虽然有损失函数定义了参数与函数之间的某种关系（而且案例中的关系非常规则），但我们无法一眼看出存在绝对最小值，但怎样的参数让函数最小，似乎只能一个一个尝试，而参数值可以是任意值，随机搜索非常昂贵
- 以简单案例来看，我们渴望聪明的算法代替随机搜索或者brute force，最好能自动导航

损失函数的困境

Gradient Descent 如何解救Cost Function?

gradient descent 梯度下降

原课程视频

第一版笔记视频

brute force 强行尝试很多值，是非常低效的；异步跟新是怎么回事

gradient descent intuition 梯度下降的直觉

原课程视频

第一版笔记视频

提督下降公式内涵的诠释，learning rate 的作用，什么时候训练会停止，更加清晰和简洁

Gradient Descent For Linear Regression 线性回归的梯度下降

原课程视频

第一版笔记视频

针对简单的线性回归，将梯度下降用于寻找损失函数的最小值；全局和地区最小值，batch, stochastic GD

直觉版：梯度下降如何解救损失函数的最小值搜索

什么是梯度下降（直觉版）

进展战线

吴恩达机器学习所需的所有线性代数

原课程视频 [Matrix Multiplication Properties Inverse and Transpose](#) 笔记视频 [图解深度学习：理解Andrew Ng机器学习所需的所有线性代数知识](#)

如何用vectorization加速多变量函数值求解

原课程： [Multiple Features](#)

笔记： [如何利用vectorization快速求解多变量函数值](#)

如何用vectorization一次性更新多变量函数的所有参数

原课程： [Gradient Descent for Multiple Variables](#)

笔记： [如何用vectorization一次性更新多变量函数的所有参数](#)

如何通过拉伸压缩features来加速梯度下降的训练进度

原课程： [Gradient Descent in Practice I - Feature Scaling](#)

笔记： [如何通过拉伸压缩features来加速梯度下降的训练进度](#)

如何快速识别模型训练是否正常以及如何调节学习率帮助训练正常化

课程：Gradient Descent in Practice II - Learning Rate

笔记：如何快速识别模型训练是否正常以及如何调节学习率帮助训练正常化

图解机器学习：为什么说提炼特征和探寻模型艺术性权重大于科学性

课程：Features and Polynomial Regression

笔记：为什么说提炼特征和探寻模型艺术性权重大于科学性

图解机器学习：Normal Equation如何实现一步求解最优参数及其特点

课程：Normal Equation

笔记：Normal Equation如何实现一步求解最优参数及其特点

图解机器学习：Normal Equation计算中出现non invertible matrix该如何处理以及背后的直观成因有哪些

课程：Noninvertibility

笔记：Normal Equation计算中出现non invertible matrix该如何处理以及背后的直观成因有哪些

如何使用Octave来做机器学习？

吴恩达的编程课：

Basic Operations Control Statements_ for, while, if statements Vectorization Working on and Submitting Programming Exercises

感受：主流语言是python, R, matlab, 非常简单易用，为什么要学Octave？

\- 没有必要学Octave, 如果要给学Octave 找一个理由，那就是Octave无需引入任何lib，操作上是最直截了当，直观简单

图解机器学习：如何使用vectorization来做梯度下降

课程：Vectorization

笔记：如何使用vectorization来做梯度下降

图解机器学习：如何理解classification以及为什么logistic比linear regression更适合做classification

课程： Classification

笔记： 如何理解classification以及为什么logistic比linear regression更适合做classification

图解机器学习： 如何理解logistic regression与linear regression的区别

课程： Hypothesis Representation

笔记： 如何理解logistic regression与linear regression的区别

图解机器学习： 如何理解decision boundary的生成原理和实质内涵

课程： Decision Boundary

笔记： 如何理解decision boundary的生成原理和实质内涵

图解机器学习： 如何理解logistic regression的损失函数的由来和构成

课程： Cost Function

笔记： 如何理解logistic regression的损失函数的由来和构成

图解机器学习： 如何对logistic regression的联合形态的损失函数做梯度下降更新参数

课程： Simplified Cost Function and Gradient Descent

笔记： 如何对logistic regression的联合形态的损失函数做梯度下降更新参数

图解机器学习： 哪些算法比gradient descent 更强大

课程： Advanced Optimization

笔记： 哪些算法比gradient descent 更强大

图解机器学习： 如何利用logistic regression二元分类模型解决多元分类问题

课程： Multiclass Classification_ One-vs-all

笔记： 如何利用logistic regression二元分类模型解决多元分类问题

图解机器学习： 如何粗略理解underfitting, overfitting, 和overfitting的解决方案

课程： The Problem of Overfitting

笔记： 如何粗略理解underfitting, overfitting, 和overfitting的解决方案

图解机器学习： 如何直观理解regularization背后的直觉

课程：Cost Function

笔记：如何直观理解regularization背后的直觉

图解机器学习：regularized linear regression计算中增加的新内容

课程：Regularized Linear Regression

笔记：regularized linear regression计算中增加的新内容

图解机器学习：regularized logistic regression在训练时有何不同？

课程：Regularized Logistic Regression

笔记：regularized logistic regression在训练时有何不同？

图解机器学习：为什么linear logistic regression不足以解决现实世界的问题

课程：Non-linear Hypotheses

笔记：为什么linear logistic regression不足以解决现实世界的问题

图解机器学习：神经网络模型的现状和目标

课程：Neurons and the Brain

笔记：神经网络模型的现状和目标

图解机器学习：跟neural net打个正式的招呼(介绍)

课程：Model Representation I

笔记：跟neural net打个正式的招呼(介绍)

图解机器学习：如何理解neuralnet的智能化精简提炼特征

课程：Model Representation II

笔记：如何理解neuralnet的智能化精简提炼特征

图解机器学习：如何通过XNOR理解neuralnet提炼复杂特征的能力

课程：Examples and Intuitions II

笔记：如何通过XNOR理解neuralnet提炼复杂特征的能力

图解机器学习：为什么说neuralnet做多元分类的本质与多logistic regression解决方案本质相同但模型构架不同

课程：Multiclass Classification

笔记：为什么说neuralnet做多元分类的本质与多logistic regression解决方案本质相同但模型构架不同

图解机器学习：neuralnet做多元分类损失函数代码执行要注意的细节

课程：Cost Function

笔记：neuralnet做多元分类损失函数代码执行要注意的细节

图解机器学习：如何理解neuralnet的backpropagation计算逻辑

课程：Backpropagation Algorithm

笔记：如何理解neuralnet的backpropagation计算逻辑

图解机器学习：如何理解neuralnet的backpropagation中的error的计算逻辑

课程：Backpropagation Intuition

笔记：如何理解neuralnet的backpropagation中的error的计算逻辑

Implementation Note_ Unrolling Parameters

图解机器学习：什么是gradient checking及其注意事项

课程：Gradient Checking

笔记：什么是gradient checking及其注意事项

图解机器学习：参数起点设计的重要性，symmetry breaking

课程：Random Initialization

笔记：参数起点设计的重要性，symmetry breaking

图解机器学习：训练神经网络模型的工作全流程梳理

课程：Putting It Together

笔记：训练神经网络模型的工作全流程梳理

图解机器学习：为什么说诊断模型病症很重要以及如何分割数据来验证模型效果

课程：Autonomous Driving Deciding What to Try Next Evaluating a Hypothesis

笔记：为什么说诊断模型病症很重要以及如何分割数据来验证模型效果

图解机器学习：为什么验证数据让模型评估更客观公正

课程：Model Selection and Train_Validation_Test Sets

笔记：为什么验证数据让模型评估更客观公正

图解机器学习：为什么以及如何用bias variance诊断模型病症

课程：Diagnosing Bias vs. Variance

笔记：为什么以及如何用bias variance诊断模型病症

图解机器学习：为什么说regularizer是模型治疗手段（bias variance诊断方式）

课程：Regularization and Bias_Variance

笔记：为什么说regularizer是模型治疗手段（bias variance诊断方式）

图解机器学习：如何用learning curve动态识别模型的病症和选择合适改进措施

课程：Learning Curves

笔记：如何用learning curve动态识别模型的病症和选择合适改进措施

图解机器学习：机器学习和神经网络模型表现差的病因分类和治疗方案选择

课程：Deciding What to Do Next Revisited

笔记：机器学习和神经网络模型表现差的病因分类和治疗方案选择

图解机器学习：如何用Error analysis 帮助改进模型（垃圾邮件分类案例）

课程：Error Analysis

笔记：如何用Error analysis 帮助改进模型（垃圾邮件分类案例）

图解机器学习：为什么面对skewed 数据precision recall能规避accuracy评估能力的片面性

课程：Error Metrics for Skewed Classes

笔记：为什么面对skewed 数据precision recall能规避accuracy评估能力的片面性

图解机器学习：如何理解precision与recall的博弈及其现实意义，为什么F1能综合precision和recall

课程：trading off precision and recall

笔记：如何理解precision与recall的博弈及其现实意义，为什么F1能综合precision和recall

图解机器学习：怎样的条件下更多数据能够（或不能）帮助改进模型

课程：Data For Machine Learning

笔记：怎样的条件下更多数据能够（或不能）帮助改进模型

图解机器学习：对比logistic regression来理解SVM通道式的decision boundary的由来

课程：Optimization Objective

笔记：SVM公式是如何从logistic regression 变形而来的

课程：Large Margin Intuition

笔记：为什么说SVM是一个维持最大间隔的classifier（decision boundary）

课程：Mathematics Behind Large Margin Classification

笔记：如何利用线性代数帮助理解管道式的decision boundary 的由来

图解机器学习：如何理解kernels通过参照点创造更好更多features来帮助模型实现non-linear能力01

课程：Using An SVM

笔记：如何理解kernels通过参照点创造更好更多features来帮助模型实现non-linear能力01

笔记：何时使用kernel来辅助SVM

图解机器学习：如何理解无监督学习做聚类的K-means算法流程和优化目标的实现

课程：Unsupervised Learning_ Introduction K-Means Algorithm Optimization Objective

笔记：如何理解无监督学习做聚类的K-means算法流程和优化目标的实现

图解机器学习：K-means如何精细处理centroid的random initialization和选择最优cluster数量

课程：Choosing the Number of Clusters

笔记：K-means如何精细处理centroid的random initialization和选择最优cluster数量

图解机器学习：为什么要对数据做降维和压缩，可视化有助于理解压缩后的维度内涵

课程：Motivation I_ Data Compression Motivation II_ Visualization

笔记：为什么要对数据做降维和压缩，可视化有助于理解压缩后的维度内涵

图解机器学习：如何理解PCA是一种projection降维的最优方案以及PCA的算法逻辑

课程：Principal Component Analysis Problem Formulation

课程：Principal Component Analysis Algorithm

笔记：如何理解PCA是一种projection降维的最优方案以及PCA的算法逻辑

图解机器学习：PCA是如何通过K寻找最优的降维幅度的

课程：Choosing the Number of Principal Components

笔记：PCA是如何通过K寻找最优的降维幅度的

图解机器学习：如何理解PCA中的Ureduce, x, z, xapproximation的关系

课程：Reconstruction from Compressed Representation

笔记：如何理解PCA中的U_reduce, x, z, x_approximation的关系

图解机器学习：实际模型训练中如何正确使用PCA和规避常见错误用法

课程：Advice for Applying PCA

笔记：实际模型训练中如何正确使用PCA和规避常见错误用法

图解机器学习：如何理解anomaly detection的本质，工作原理和算法流程

课程：Problem Motivation Gaussian Distribution Algorithm

笔记：如何理解anomaly detection的本质，工作原理和算法流程

图解机器学习：如何理解anomaly detection实际使用的简单流程

课程：Developing and Evaluating an Anomaly Detection System

课程：Anomaly Detection vs. Supervised Learning

课程：Choosing What Features to Use

笔记：如何理解anomaly detection实际使用的简单流程

图解机器学习：如何理解单变量正态分布 对比多变量正态分布运用于anomaly detection的优缺点

课程：Multivariate Gaussian Distribution

课程：Anomaly Detection using the Multivariate Gaussian Distribution

笔记：如何理解单变量正态分布 对比多变量正态分布运用于anomaly detection的优缺点

图解机器学习：为什么要学习推荐系统，简单推荐系统的问题长什么样

课程：Problem Formulation

笔记：为什么要学习推荐系统，简单推荐系统的问题长什么样

图解机器学习：为什么说content based recommender是用电影特征来训练用户参数

课程：Content Based Recommendations

笔记：为什么说content based recommender是用电影特征来训练用户参数

图解机器学习：为什么说collaborative filtering是让电影特征与用户参数不断迭代相互训练相互更新

课程：Collaborative Filtering

笔记：为什么说collaborative filtering是让电影特征与用户参数不断迭代相互训练相互更新

图解机器学习：梳理collaborative filtering合并2个交替训练模型的算法流程和如何利用电影和用户特征来预测

课程：Collaborative Filtering Algorithm

笔记：梳理collaborative filtering合并2个交替训练模型的算法流程和如何利用电影和用户特征来预测

图解机器学习：在collaborative filtering代码执行如何采取vectorization加速计算和训练后快速找相似电影

课程：Vectorization_ Low Rank Matrix Factorization

笔记：在collaborative filtering代码执行如何采取vectorization加速计算和训练后快速找相似电影

图解机器学习：新用户没有给出任何评分时，如何避免训练好的模型代替新用户给所有电影都打0分

课程：Implementational Detail_ Mean Normalization

笔记：新用户没有给出任何评分时，如何避免训练好的模型代替新用户给所有电影都打0分

图解机器学习：面对巨量数据如何高效训练模型和发现问题改进

课程：Learning With Large Datasets

课程：Stochastic Gradient Descent

课程：Mini-Batch Gradient Descent

课程：Stochastic Gradient Descent Convergence

笔记：面对巨量数据如何高效训练模型和发现问题改进

图解机器学习：什么是online learning（一句话）

课程：Online Learning

笔记：什么是online learning（一句话）

图解机器学习：什么是map reduce 和data parallal（一句话）

课程：Map Reduce and Data Parallelism

笔记：什么是map reduce 和data parallal（一句话）

图解机器学习：一个简单的OCR文字识别系统的工作全流程长什么样子

课程：Problem Description and Pipeline

课程：Sliding Windows

笔记：一个简单的OCR文字识别系统的工作全流程长什么样子

图解机器学习：合成新数据的方法及相关挑战和注意事项

课程：Getting Lots of Data and Artificial Data

笔记：图解机器学习：合成新数据的方法及相关挑战和注意事项

图解机器学习：如何找到能最大化改进模型效果的具体工作任务或模型组建

课程：What Part of the Pipeline to Work on Next

笔记：如何找到能最大化改进模型效果的具体工作任务或模型组建

图解吴恩达机器学习课程完结：for who, how to use, why

课程：Summary and Thank you

笔记：图解吴恩达机器学习课程完结：for who, how to use, why