



# 北京中科资环信息技术研究院

## 人工智能人才培养计划

中科资环发(2018)63号

### 第四期-MATLAB 高级编程及机器学习技术应用培训班

各企事业单位:

近年来,随着 AlphaGo 人机大战、ImageNet 挑战赛等热点事件的发生,人工智能再次回到了人们的视野,迎来了新一轮的发展浪潮,各行各业都在寻求新的“风口”和机遇。尤其是在计算机视觉和图像处理领域,各种颠覆性的成果应运而生。因此,为了帮助广大科研人员更加系统地学习机器学习和深度学习的基础理论知识及对应的代码实现方法,北京中科资环信息技术研究院特举办“MATLAB 高级编程及机器学习技术应用”培训班,旨在帮助学员掌握各种经典机器学习算法以及最新的深度神经网络、迁移学习等算法的基本原理及其 MATLAB 编程实现方法。本次培训采用“理论讲解+案例实战+动手实操+讨论互动”相结合的方式,抽丝剥茧、深入浅出分析使用机器学习和深度学习时需要掌握的经验及编程技巧。此外,本次培训还将通过实际案例的形式,介绍如何提炼创新点,以及如何发表高水平论文等相关经验。

一、时间地点: **2018年9月13日—9月17日 四川-成都**  
(时间安排:第一天报到、授课四天)

二、培训目标:

- 1.掌握各种经典机器学习算法的技巧及其 MATLAB 代码实现方法
- 2.掌握最新的深度学习、迁移学习等算法的基本原理及其 MATLAB 代码实现方法
- 3.通过实际案例操作,熟练掌握 MATLAB 神经网络应用方法
- 4.通过实际案例操作,熟练掌握 MATLAB 图像处理技术及深度学习、神经网络在图像处理中的应用
- 5.通过实操培训掌握各种编程技巧
- 6.通过现场指导解决学员实际工作项目中的疑难问题,建立技术服务平台,为后期应用提供技术交流

三、培训对象:

各省市、自治区从事各行各业的数据处理、建模分析等方向相关的企事业单位技术骨干、科研院所研究人员和大专院校相关专业教学人员及在校本科生、硕士和博士等相关人员,以及对机器学习和编程感兴趣的广大爱好者。

主办单位:北京中科资环信息技术研究院  
二零一八年八月三日

协办单位:北京中科硕博研计算技术中心  
二零一八年八月三日





#### 四、主讲专家:

**郁磊博士:** 主要从事 MATLAB 编程、机器学习与数据挖掘、数据可视化和软件开发、生理系统建模与仿真、生物医学信号处理, 具有丰富的实战应用经验, 主编《MATLAB 智能算法 30 个案例分析》、《MATLAB 神经网络 43 个案例分析》相关著作。已发表多篇高水平的国际学术论文。

#### 五、培训费用:

3900 元 (报名费、培训费 (会议费、资料费) 证书费、午餐费) 住宿可统一安排, 费用自理。

**(发票可开具: 培训费、会议费、资料费等, 会场发放会议红头文件, 用于参会人员报销使用)**

#### 六、颁发证书 (参加培训的学员可以获得双证书)

A 类: 国家工业和信息化部人才交流中心颁发的《机器学习开发与应用技术》国家紧缺人才培养工程证书, 该证书直接纳入工信部人才库。

B 类: 《机器学习技术》专业技术培训证书。此证书作为个人学习和知识更新、专业技能提升、单位人才聘用的参考依据。

**注:** 请学员准备电子版身份证、学历证、2 寸蓝底照片提交至会务组。

#### 七、报名办法:

请各单位接此通知后, 尽快确定参加培训人员, 认真填写报名回执并传真 010-53853500 或 E-mail 至: [2010694422@qq.com](mailto:2010694422@qq.com), 会务组收到回执后通知报到相关事项。

#### 八、联系方式:

联系人: 孙畅 手机/微信: 13393121657 电话/传真: 010-53853500 QQ: 2010694422

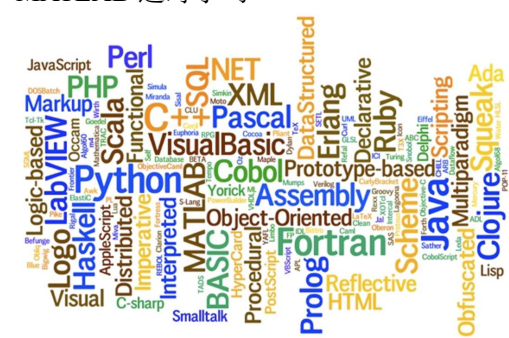
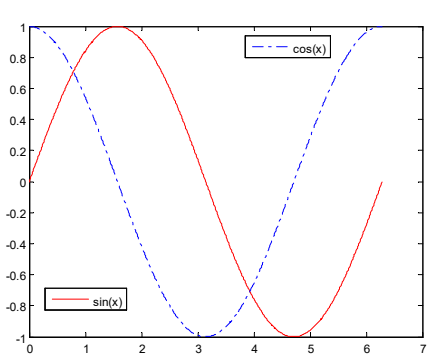
MATLAB 编程与机器学习 QQ 群: [540840947](https://www.qq.com/group/540840947) 加群验证: 【孙畅】邀请

### 报名回执表

单位名称						
发票抬头				发票类型	<input type="checkbox"/> 增值普票 <input type="checkbox"/> 增值专票	
纳税人识别号				发票内容	<input type="checkbox"/> 培训费 <input type="checkbox"/> 会议费	
通讯地址						
学员姓名	性别	部门	工作 (研究) 方向	职务	联系方式	邮箱
希望通过会议解决哪些问题 (请认真填写)		问题一: 问题二: 问题三:				
培训费		万 仟 佰 拾 元		电汇日期		
培训费用: <input type="checkbox"/> 电汇 <input type="checkbox"/> 现场办理费用 (请用“√”标注)						
汇 款 账 户	账户: 北京中科硕博研计算技术中心 开户行: 中国工商银行股份有限公司北京学院路支行 账号: 0200 0255 0920 0106 214 注: 请将银行汇款凭证传真至会务组。					
是否需要住宿: <input type="checkbox"/> 是 <input type="checkbox"/> 否 入住日期__、结束日期__、住宿天数__天 标准间__间、大床__间。				参会单位盖章: 参会学员签名: 会议负责人: 孙 畅 二零一八年 月 日		



# 课程大纲

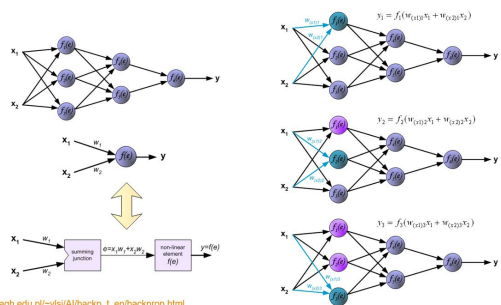
时间	授课方式	课程	主要内容
第一天	09:00 – 10:00		<p>1、到底应该如何学习编程？应该学习哪种编程语言？MATLAB 过时了吗？</p>  <p>2、简单介绍 MATLAB 的安装、版本历史与编程环境（应该安装哪个版本的 MATLAB？在哪些情况下需要时刻关注最新版本的 MATLAB？）</p> <p><a href="#">Neural Network Toolbox Release Notes</a></p> <p>Found 32 notes   Release Range: R2014b to R2017b</p> <p>Sort by:</p> <p>~ R2017b</p> <p>New Features, Bug Fixes, Compatibility Considerations</p> <ul style="list-style-type: none"> <li>Directed Acyclic Graph (DAG) Networks: Create deep learning networks with more complex architecture to improve accuracy and use many popular pretrained models</li> <li>Long Short-Term Memory (LSTM) Networks: Create deep learning networks with the LSTM recurrent neural network topology for time-series classification and prediction</li> <li>Deep Learning Validation: Automatically validate network and stop training when validation metrics stop improving</li> <li>Deep Learning Layer Definition: Define new layers with learnable parameters, and specify loss functions for classification and regression output layers</li> <li>Deep Learning Training Plots: Monitor training progress with plots of accuracy, loss, validation metrics, and more</li> <li>Deep Learning Image Preprocessing: Efficiently resize and augment image data for training</li> <li>Bayesian Optimization of Deep Learning: Find optimal settings for training deep networks (Requires Statistics and Machine Learning Toolbox)</li> <li>GoogLeNet Pretrained Network: Transfer learning with pretrained GoogLeNet convolutional neural network</li> <li>Batch Normalization Layer: Speed up network training and reduce sensitivity to network initialization</li> <li>Deep Learning: New network layers</li> <li>Functionality Being Removed or Changed</li> </ul>
	10:00 – 12:00	<p>课程简介 经验分享</p> <p>相关知识复习 与巩固 理论讲解与案例 演示 实操练习</p>	<p>第一章：MATLAB 快速入门</p> <p>3、MATLAB 基础操作：包括矩阵操作、逻辑与流程控制、函数与脚本文件、基本绘图等（浩如烟海的函数应该怎样记忆？矩阵最常用的五种索引操作是什么？MATLAB 绘图功能真的比较弱吗？如何导出高质量的图像，以供满足 SCI 论文的要求？）</p>  <p>4、文件导入：mat、txt、xls、csv、jpg、wav、avi 等格式（mat 格式的文件是怎么生成的呢？为什么 mat 文件导入到 workspace 中会有各种各样名字的变量？）</p>



	<p>13:30 – 15:30</p>	<p>理论讲解与案例 演示 实操练习</p>	<p>第二章: MATLAB 进阶与提高</p>	<p>1、MATLAB 编程习惯与风格 (Cell 模式和程序发布功能是什么? 为什么代码中需要加入一些空格和空白行? MATLAB 新版本中包含的 Live Script 是什么?)</p> <div style="border: 1px solid #ccc; padding: 5px;"> <p><b>Contents</b></p> <ul style="list-style-type: none"> <li>▪ I. 清空环境变量</li> <li>▪ II. 训练集/测试集产生</li> <li>▪ III. 数据归一化</li> <li>▪ IV. BP神经网络创建、训练及仿真测试</li> <li>▪ V. 性能评价</li> <li>▪ VI. 绘图</li> </ul> <p><b>I. 清空环境变量</b></p> <pre style="font-family: monospace; font-size: 0.9em;">clear all clc</pre> <p><b>II. 训练集/测试集产生</b></p> <p>1. 导入数据</p> <pre style="font-family: monospace; font-size: 0.9em;">load spectra_data.mat</pre> </div> <pre style="font-family: monospace; font-size: 0.8em; margin-top: 10px;">%% I. 清空环境变量 clear all clc  %% II. 训练集/测试集产生 % 1. 导入数据 load spectra_data.mat  %% % 2. 随机产生训练集和测试集 temp = randperm(size(NIR,1)); % 训练集——50个样本 P_train = NIR(temp(1:50),:); T_train = octave(temp(1:50),:); % 测试集——10个样本 P_test = NIR(temp(51:end),:); T_test = octave(temp(51:end),:); N = size(P_test,2);</pre> <p>2、MATLAB 调试技巧 (MATLAB 为什么会给出各种各样的错误信息? 常见的错误信息有哪些? 面对错误信息, 应该是失落还是开心? 如何使用断点调试工具? 应该去哪些网站寻找答案? 怎样能够让别人乐意帮助你解决问题?)</p> <ul style="list-style-type: none"> <li>■ 错误信息的阅读 <ul style="list-style-type: none"> <li>- Index must be a positive integer or logical.</li> <li>- Undefined function or variable "B".</li> <li>- Inner matrix dimensions must agree.</li> <li>- Function definitions are not permitted at the prompt or in scripts.</li> <li>- Index out of bounds because numel(A)=5.</li> <li>- In an assignment A(I) = B, the number of elements in B and I must be the same.</li> <li>- Expression or statement is incorrect--possibly unbalanced ( , {, or [.</li> <li>- Too many input arguments.</li> <li>- .....</li> </ul> </li> </ul> <p>3、向量化编程与内存优化 (怎样提升你的代码效率? MATLAB 的内存管理机制是什么? 为什么你的代码中会出现许多红色的下划波浪线?)</p> <ul style="list-style-type: none"> <li>■ 及时清除不用的变量</li> <li>■ 使用变量前, 预分配内存空间</li> <li>■ 选择恰当的数据类型</li> <li>■ 循环与向量化 <ul style="list-style-type: none"> <li>- 按列优先循环</li> <li>- 循环次数多的变量安排在内层</li> </ul> </li> <li>◆ 给一些函数“瘦身”</li> <li>◆ .....</li> </ul>
	<p>15:30 – 17:30</p>	<p>理论讲解与案例 演示 实操练习</p>	<p>第三章: MATLAB GUI 设计与软件开发</p>	<p>1、利用 GUIDE 向导工具开发 (常用的控件有哪些? 控件的属性怎样设置? 如何编写控件的回调函数?)</p>



				 <p>2、利用 M 语言开发（图形对象句柄是什么？利用 M 语言和 GUIDE 向导工具开发各有什么优缺点？怎样提升软件的交互性？）</p> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>◆窗口的创建</p> <pre>%% 创建窗口Figure HandleFigure = figure('position',[500 300 500 400],...     'name','GUIExample_2',...     'numbertitle','off',...     'menubar','none',...     'toolbar','none',...     'Resize','off');</pre> </div> <div style="width: 45%;"> <p>◆坐标轴的创建</p> <pre>%% 创建Axes HandleAxes = axes('parent',HandlePanel_Plot,...     'position',[0.1 0.1 0.85 0.85],...     'box','on',...     'tick',' ',...     'ytick',' ',...     'uicontextmenu',HandleContextMenu);</pre> </div> </div> <p>3、程序编译与打包发布（怎样生成 exe 可执行文件？将 exe 可执行文件部署到其它计算机上时需要注意哪些问题？怎样加密代码？）</p> 
第二天	09:00 – 10:00	理论讲解与案例演示	第四章：BP 神经网络	1、BP 神经网络的基本原理（人工智能发展过程经历了哪些曲折？人工神经网络的分类有哪些？BP 神经网络的拓扑结构和训练过程是怎样的？什么是梯度下降法？BP 神经网络建模的本质是什么？）
10:00 – 11:00	实操练习			
11:00 – 12:00	知识点总结、延拓课堂讨论			

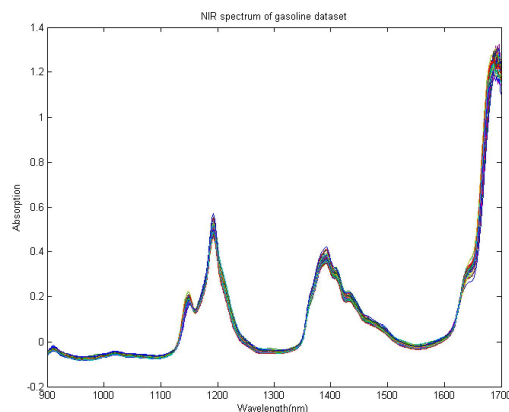


Source: [http://galaxy.agh.edu.pl/~ysi/Al/backp\\_1\\_en/backprop.html](http://galaxy.agh.edu.pl/~ysi/Al/backp_1_en/backprop.html)

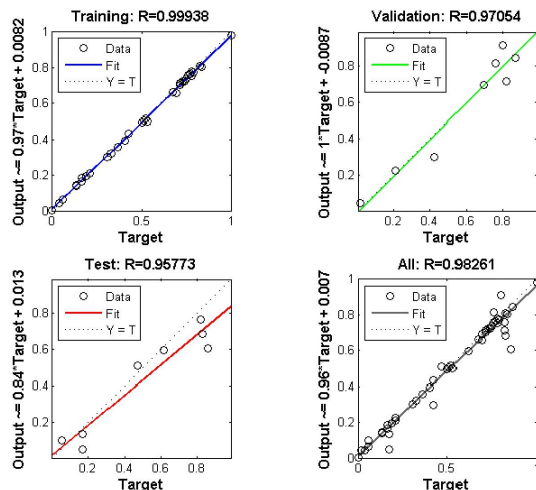
## 2、BP 神经网络的 MATLAB 实现（怎样划分训练集和测试集？为什么需要归一化？归一化是必须的吗？MATLAB 中 BP 神经网络的常用函数有哪些？如何使用？）

- 什么是归一化？
  - 将数据映射到 [0, 1] 或 [-1, 1] 区间或其他区间。
- 为什么要归一化？
  - 输入数据的单位不一样，有些数据的范围可能特别大，导致的结果是神经网络收敛慢，训练时间长。
  - 数据范围大的输入在模式分类中的作用可能会更大，而数据范围小的输入作用可能会更小。
  - 由于神经网络输出的激活函数的值域是有限的，因此需要将网络训练的目标数据映射到激活函数的值域。例如神经网络的输出层若采用 S 形激活函数，由于 S 形函数的值域限制在 (0, 1)，也就是使神经网络的输出只能限制在 (0, 1)，所以训练数据的输出要归一化到 [0, 1] 区间。
  - S 形激活函数在 (0, 1) 区间以外区域很平缓，区分度太小，例如 S 形函数  $f(x)$  在参数  $a=1$  时，有 100 与 105 只相差 0.0067。
- 归一化算法
  - $y = (x - \min) / (\max - \min)$
  - $y = 2 * (x - \min) / (\max - \min) - 1$

## 3、案例实践：近红外光谱汽油辛烷值预测

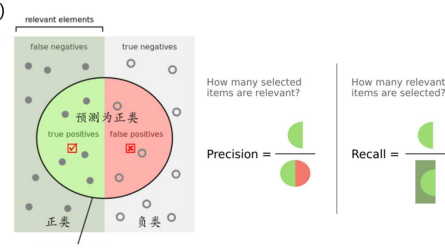
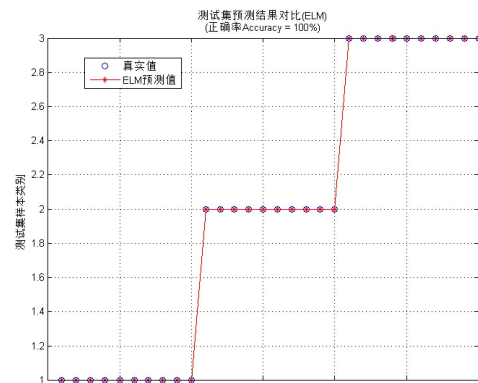
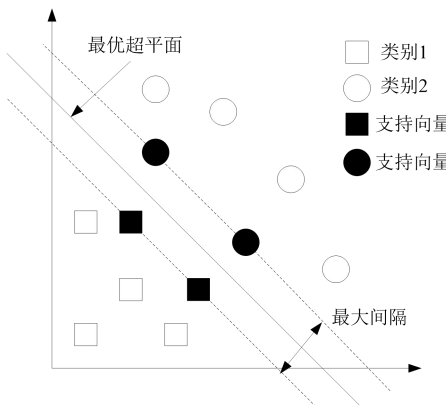


## 4、BP 神经网络参数的优化（隐含层神经元个数、学习率、初始权值和阈值等如何设置？什么是交叉验证？）

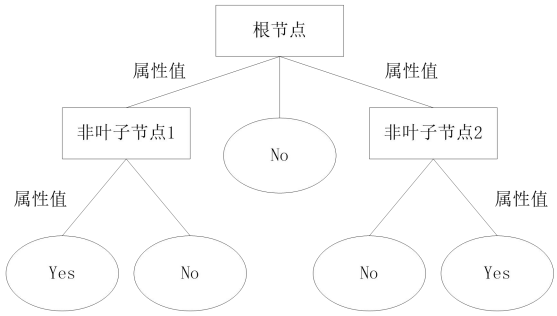
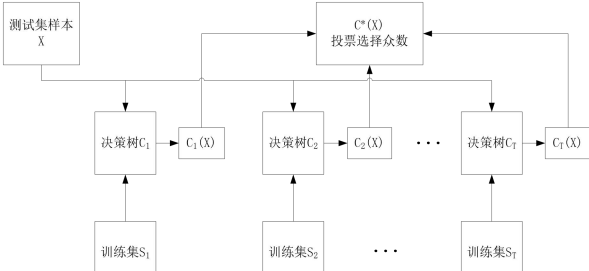
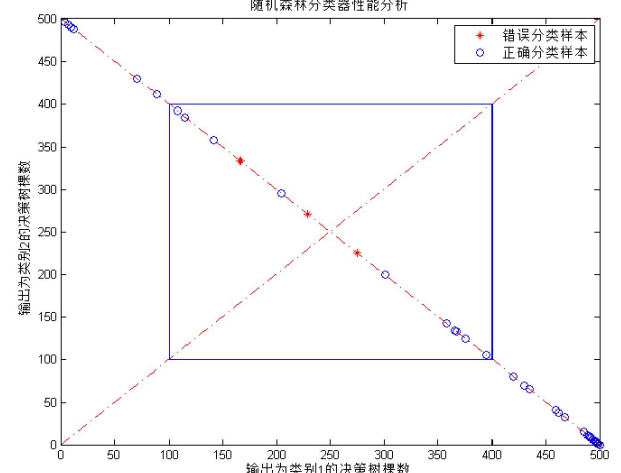


## 5、值得研究的若干问题（欠拟合与过拟合、泛化性能评价指标的设计、样本不平衡问题等）

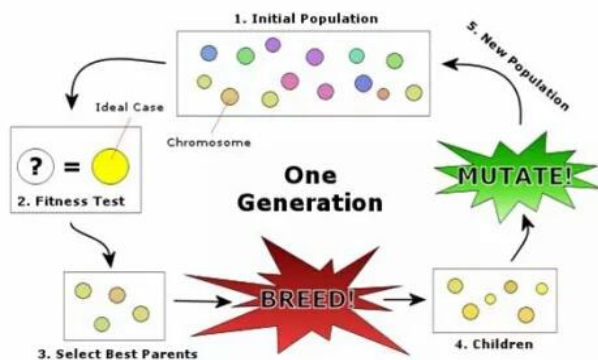


				<p>■ 模型泛化性能评价 (绝对误差、相对误差、均方根误差、准确率、精确率、召回率等)</p>  <p>Precision = <math>\frac{\text{true positives}}{\text{true positives} + \text{false positives}}</math></p> <p>Recall = <math>\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}</math></p>
13:30 – 15:30	理论讲解与案例演示 实操练习	第五章：极限学习机 (Extreme Learning Machine, ELM)	<p>1、ELM 的基本原理 (ELM 的基本算法, “极限”体现在哪些地方?)</p> <p><b>Algorithm ELM:</b> Given a training set <math>\mathcal{N} = \{(x_i, t_i)   x_i \in \mathbf{R}^n, t_i \in \mathbf{R}^m, i = 1, \dots, N\}</math>, activation function <math>g(x)</math>, and hidden node number <math>\tilde{N}</math>,</p> <p>Step 1: Randomly assign input weight <math>w_i</math> and bias <math>b_i</math>, <math>i = 1, \dots, \tilde{N}</math>.</p> <p>Step 2: Calculate the hidden layer output matrix <math>\mathbf{H}</math>.</p> <p>Step 3: Calculate the output weight <math>\beta</math></p> $\beta = \mathbf{H}^\dagger \mathbf{T}, \quad (16)$ <p>where <math>\mathbf{T} = [t_1, \dots, t_N]^T</math>.</p> <p>2、ELM 与 BP 神经网络的区别与联系</p> <p>3、案例实践：鸢尾花种类识别</p> 	<p>1、SVM 的基本原理 (什么是经验误差最小和结构误差最小? SVM 的本质是解决什么问题? SVM 的四种典型结构是什么? 核函数的作用是什么? 什么是支持向量?)</p>  <p>2、SVM 扩展知识 (如何解决多分类问题? SVM 除了建模型之外, 还可以帮助我们做些什么事情?)</p>
15:30 – 17:30	理论讲解与案例演示 实操练习 知识点总结、延拓	第六章：支持向量机 (Support Vector Machine, SVM)		



				<p>3、LibSVM 工具箱的安装与使用</p> <p>4、案例实践一：乳腺癌肿瘤诊断</p> <p>5、案例实践二：混凝土强度预测</p>
<p>第三天</p>	<p>09:00 – 10:30</p>	<p>理论讲解与案例演示 实操练习 知识点总结、延拓 课堂讨论</p>	<p>第七章：决策树与随机森林</p>	<p>1、决策树的基本原理（微软小冰读心术的启示；什么是信息熵和信息增益？ID3算法和C4.5算法的区别与联系）</p>  <p>2、随机森林的基本原理（为什么需要随机森林算法？广义与狭义意义下的“随机森林”分别指的是什么？“随机”体现在哪些地方？随机森林的本质是什么？）</p>  <p>3、案例实践：乳腺癌肿瘤诊断</p> <p>4、知识扩展：决策树除了建模型之外，还可以帮我们做什么事情？怎样解读随机森林的结果？</p> 
	<p>10:30 – 12:00</p>	<p>理论讲解与案例演示 实操练习</p>	<p>第八章：遗传算法 (Genetic Algorithm, GA)</p>	<p>1、遗传算法的基本原理（以遗传算法为代表的群优化算法的基本思想是什么？目前国内外的研究热点在哪些方面？）</p>



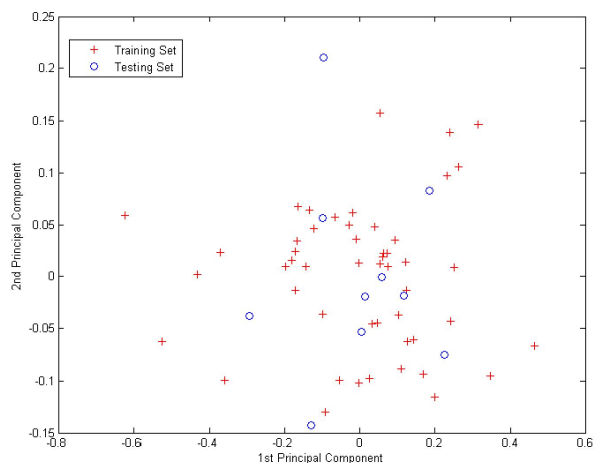


- 2、常见遗传算法工具箱介绍 (GAOT 工具箱的添加与使用方法)
- 3、案例实践一：一元函数的寻优计算 (极大值与极小值)
- 4、案例实践二：多元函数的寻优计算 (遗传算法优化 BP 神经网络的初始权值与阈值)
- 5、案例实践三：离散变量的寻优计算 (基于遗传算法的特征变量筛选)

13:30 – 16:00

理论讲解与案例演示  
实操练习

- 1、变量降维与特征选择在概念上的区分
- 2、主成分分析 (PCA)、偏最小二乘 (PLS) 的基本原理 (PCA 与 PLS 的区别与联系; PCA 除了降维之外, 还可以帮助我们做什么?)

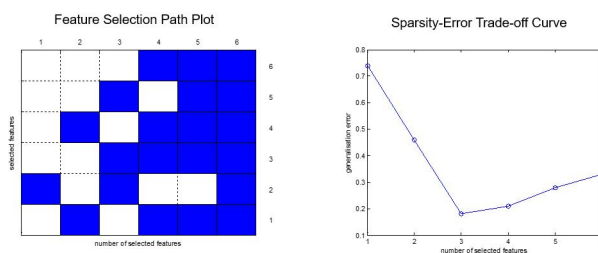


16:00 – 17:30

实操练习  
课堂讨论  
复习与答疑

### 第九章：变量降维与特征选择

- 2、常见的特征选择方法 (优化搜索、Filter 和 Wrapper 等; 前向与后向选择法; 区间法; 无信息变量消除法; 正则稀疏优化方法等)



- 3、案例实践：近红外光谱的降维与波长筛选



第四天

09:00 – 10:00

理论讲解与案例  
演示  
实操练习

第十章：图像处理入门

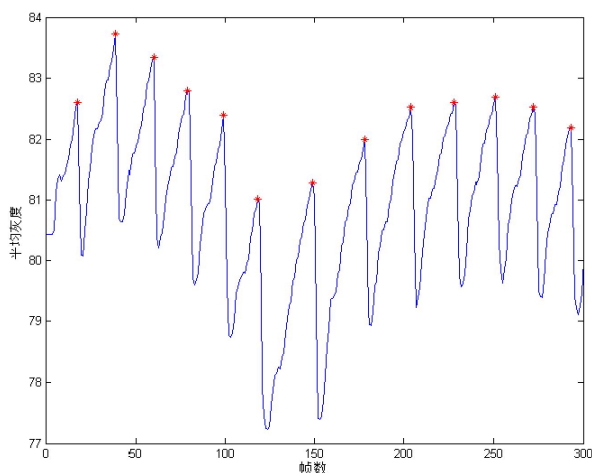
- 1、图像的常见格式及读写（彩色图像、灰度图像、二值图像等）
- 2、图像的基本操作（几何变换、时频域变换等）
- 3、图像直方图（图像的对比度、亮度等改变）
- 4、案例实践：基于手机摄像头的心率计算

### 一、工作原理

将手指覆盖在手机摄像头上，用摄像头照亮指尖皮下毛细血管，当心脏将新鲜的血液压入毛细血管时，亮度（红色的深度）会有轻微变化，通过摄像头监测这一有规律变化的间隔，利用图像处理分析技术，即可计算出心率。

### 二、实验方法

- 手机型号：小米（Android 4.2）
- 视频采集帧频：15.17帧/秒
- 数据采集方法：受试者呈坐姿，保持安静状态，将右手手指轻轻覆盖在手机摄像头上，左手操作手机进入摄像模式，录制30秒左右的视频并保存为3gp格式。
- 利用格式转换工具将视频文件转换为avi格式视频，以便于MATLAB软件的读取与分析。

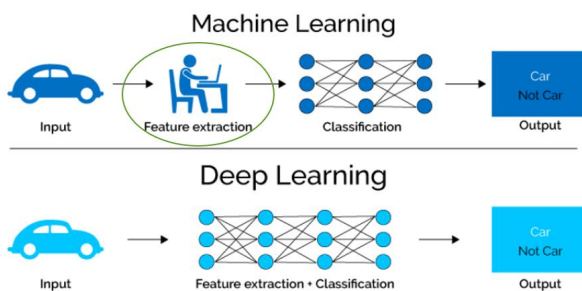


10:00 – 12:00

理论讲解与案例  
演示  
实操练习

第十一章：深度学习入门

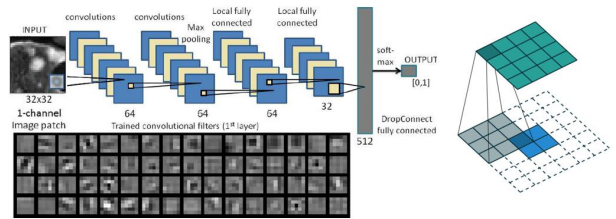
- 1、深度学习与传统机器学习的区别与联系（神经网络的隐含层数越多越好吗？深度学习与传统机器学习的本质区别是什么？）



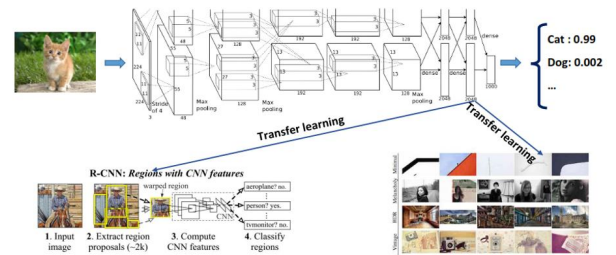
- 2、深度学习开源工具箱简介（Python & MATLAB）



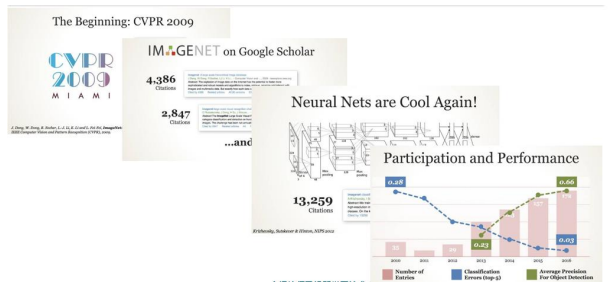
3、卷积神经网络的基本原理(什么是卷积核? CNN的典型拓扑结构是怎样的? CNN的权值共享机制是什么? CNN提取的特征是怎样的?)



4、迁移学习算法的基本原理(为什么需要迁移学习? 为什么可以迁移学习? 迁移学习的基本思想是什么?)



5、专题讨论: 数据 VS.模型, 孰更重要?(模型泛化性能不好, 究竟是孰之过?)

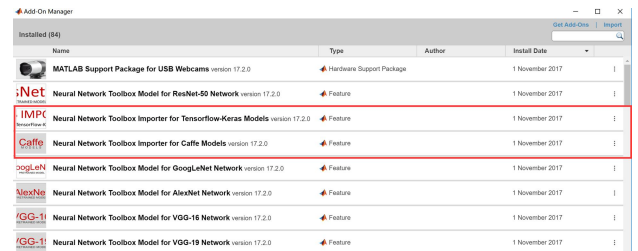


13:30 – 15:30

理论讲解与案例演示  
实操练习

第十二章: 深度学习在图像处理中的应用案例

### 1、MATLAB 2017b 深度学习工具箱功能介绍



### 2、案例一: 11 行代码实现深度学习物体识别



			 <pre data-bbox="877 313 1141 504">= webcam; % Connect to the camera alexnet; % Load the neural net  :true icture = camera.snapshot; % T icture = imresize(figure, [227, 227]); % R  nel = classify(nnnet, picture); % C  ge(figure); % Show the picture le(char(label)); % Show the label nnow;</pre> <p data-bbox="877 526 1492 683">3、案例二：利用卷积神经网络抽取抽象特征 4、案例三：自定义卷积神经网络拓扑结构 5、案例四：基于卷积神经网络的手写数字识别 6、案例五：基于卷积神经网络实现迁移学习</p>
15:30 – 17:30	课堂讨论 复习与答疑	第十三章：讨论与答疑	<ol data-bbox="877 694 1492 1176" style="list-style-type: none"><li>1. 如何查阅文献资料？（你会使用 Google Scholar、Sci-Hub、ResearchGate 吗？应该去哪些地方查找与论文配套的数据和代码？）</li><li>2. 如何高效率撰写专业论文？（SCI 不同分区的论文差别在哪些地方？你知道你的论文为什么显得很单薄吗？）</li><li>3. 从审稿人的角度看，SCI 期刊论文需要具备哪些要素？（审稿人关注的点有哪些？如何回应审稿人提出的意见？）</li><li>4. 如何提炼与挖掘创新点？（如果在算法层面上难以做出原创性的工作，如何结合自己的实际问题提炼与挖掘创新点？）</li></ol>