

## 机器学习应用-20级大作业

本次作业分为必选题和选做题，考试要求如下：

1. **最终的大作业以ppt形式提交**，ppt最后一页请注明**组内每位同学的贡献度（用百分比表示）**
2. **必选开放问答题**：在回答过程中，请区分出每位同学的回答，如若看不出每位同学的回答，则组内每位同学在该项的得分均视为0分；
2. **必选专业问答题**：在回答过程中，每组同学可以凝练成一个回答，不用区分多位同学的回答；
3. **必选双创题**：请每组同学参考双创题的具体要求答题
4. **必选算法题**：每道题请用2-4页进行展示，内容可以自行安排
5. **选做题**：若选择做该部分，只能选1道题做，每道题除了对应的ppt以外，还需提供对应的论文。**特别说明，选做题不做，不会导致挂科**



**必选开放问答题（每道题均要回答，总页数不超过4页ppt）：**

1. 您为什么要学习机器学习这门课程？
2. 您对该课程最满意和最不满意的地方是什么？
2. 您认为还有哪些问题急需在今后的教学中改进和提高？
4. 总体上谈谈您对该课程的看法





**必选专业问答题（每道题均要回答，总页数不超过6页ppt）：**

1. 谈谈你对梯度和梯度下降算法的理解
2. 为什么要使用交叉验证
3. 一般的机器学习应用问题通常包括哪些步骤？
4. 对于图像分类任务，请描述用机器学习算法（非深度学习算法）解决此类任务的具体流程（必要时可画图描述）



## 必选双创题：

请各组以“AI赋能行业应用”为主题，写一份关于此内容的创新创业大纲。内容包含以下几个部分：

1. 明确所要解决的问题是什么（1页ppt呈现）
2. 行业背景调研（1页ppt呈现）
3. 市场需求及代表性的企业（1页ppt呈现）
4. 拟采用的AI技术及对应的解决此类问题的技术路线（2页ppt呈现）
5. 团队成员分工及各自的优势（1页ppt呈现）
6. 商业模式（1页ppt呈现）
7. 其他（可选项，自由发挥，1页ppt呈现）





## 必选算法题：

1. 附件heart.xls是一组包含病人是否患心脏病的数据集，其中age:年龄，sex:性别，cp:胸痛类型，trestbps:休息时血压，chol:胆固醇，fbs:血糖，restecg:心电图，thalach:最大心率，exang:运动后心绞痛，oldpeak:运动后ST段压低，slope:运动高峰期ST段的斜率，ca:主动脉荧光造影染色数，thal:缺陷种类，target:0代表无心脏病，1代表有心脏病。请用ML方法建模实现判断病人是否患有心脏病，并验证该模型的有效性。同时建模判断该病人在患有心脏病的条件下，判别其胸痛属于哪一类别，并验证该模型的有效性。



## 必选算法题：

2. 请分别使用“gist特征+SVM”、“gist特征+MLP”、“gist特征+决策树”、“gist特征+贝叶斯分类器”、“LBP特征+SVM”、“LBP特征+MLP”、“LBP特征+决策树”、“LBP特征+贝叶斯分类器”和CNN九种方法对NWPU-RESISC45数据集进行图像分类，评价准则和相关内容请参考文献1，CNN网络结构可自行决定。

## 参考文献：

1. Gong Cheng, Junwei Han, Xiaoqiang Lu. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE, vol. 105, no. 10, pp. 1865-1883, 2017.





**选做题（若选择做该部分，只能选1道题做，每道题除了对应的ppt以外，还需提供对应的论文）：**

1. 请读以下论文，试比较ResNet和DenseNet在数据集SIPAKMED上的分类效果，评价规则可参考文献3。同时使用无监督或者半监督CNN方法在该数据集上进行实验，并撰写相关论文（包含7部分内容：中文摘要、英文摘要、引言、方法原理、实验、结论和参考文献，字数不限）

参考文献：

1. Kaiming He, Xiangyu Zhang, Shaoqing Ren et al. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016
2. Gao Huang, Zhuang Liu, Laurens Van Der Maaten, et al. Densely Connected Convolutional Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017
3. Marina E. Plissiti, Panagiotis Dimitrakopoulos, Giorgos Sfikas, et al. SIPAKMED: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images, IEEE International Conference on Image Processing (ICIP), 2018.
4. <https://www.cs.uoi.gr/~marina/sipakmed.html>



**选做题（若选择做该部分，只能选1道题做，每道题除了对应的ppt以外，还需提供对应的论文）：**

2. colorectal cancer (CRC)数据集包含10万张HE染色的结直肠图像块，训练集记为NCT-CRC-HE-100K，测试集包含7180张图像块，记为CRC-VAL-HE-7K。请使用CNN方法以及基于注意力机制的CNN方法完成对该数据集的图像分类和图像检索，并撰写相关论文（包含7部分内容：中文摘要、英文摘要、引言、方法原理、实验、结论和参考文献，字数不限）。

参考文献：

1. Tingying Peng, Melanie Boxberg, Wilko Weichert, et al. Multi-task Learning of a Deep K-Nearest Neighbour Network for Histopathological Image Classification and Retrieval. MICCAI 2019.





**选做题（若选择做该部分，只能选1道题做，每道题除了对应的ppt以外，还需提供对应的论文）：**

3. 目标检测是计算机视觉和机器学习领域的热点问题。请根据附件里的图像，使用目标检测算法实现对细胞的检测，并撰写相关论文（包含7部分内容：中文摘要、英文摘要、引言、方法原理、实验、结论和参考文献，字数不限）。

