本次作业提交的文件名命名格式为：班级+姓名+L0，请各位同学自己通过邮件发送给我：zgcao@hust.edu.cn

L0 作业

1，你对模式识别、机器学习、计算机视觉等有何了解？你对本门课程的学习有何期待？

2，在人工智能领域中你所感兴趣的方向是哪一个？结合你过去的背景和经验说明你会怎样实现在这个专业方向的成功，你认为你在本科阶段所学的知识将会对你取得这个领域的成功有什么帮助。

3，用具体事例描述一门你认为对你来说最重要的课程和一段经历，并解释它们为什么对你来说是至关重要的？

4，你的短期和长期职业目标分别是什么，你渴望的职位是什么，请尽量详述。你计划分哪几步实现这个目标？

Lecture1 作业

1，现在很多火车站都可以持身份证和车票“刷脸”进站，张同学拿着自己的身份证进站时机器却识别错误，从模式识别的角度看机器难以识别他的原因是：

（a）不存在模式；　　（b）没有训练样本；

（c）不同类的模式相似度高；　　（d）同一类的模式差异性大。

2，某研究人员用带有标签的数据库训练其目标识别算法，希望用在自动驾驶中对路况进行场景理解和行人检测，请问这个算法需要去解决哪种类型的学习问题：

（a）有监督学习；　　（b）无监督学习；

（c）半监督学习；　　（d）强化学习。

3，如果一个好的学习算法通过训练样本集在解空间（假设空间）集合中找到一个最优解，使得其对所有训练样本都能够实现正确的模式识别，以下哪种说法正确：

（a）目标函数是已知的；

（b）所有训练样本都是线性可分的；

（c）训练样本集上全部分类正确，不能代表测试时也能正确；

（d）好的学习算法能够容许训练样本存在错误的标签。

4，某个电影网站每名网络用户都可以给电影打分，用户有 id 号，电影也有 id 号，网站已积累了 10 万个用户对所看电影喜爱程度的打分值，并将电影的特征如喜剧片、动作片、爱情片、…、明星甲、明星乙等与观众对这些特征的喜好匹配组合到一起设计了推荐系统，给每部电影一个百分制的评分结果。请用上述这些元素去构造一个学习问题。

5，有一数据集共有 2000 张花卉图片，其中 1400 张是玫瑰，300 张是月季，300 张是蔷薇。某同学设计了一个玫瑰识别算法，从 2000 张图片中识别出 1000 张图片为玫瑰，但实际上其中只有 600 张是玫瑰，另外 300 张是月季、100 张是蔷薇。请计算分类正确率 (Accuracy)、分类错误率 (Error rate)、分类精度 (Precision)、召回率 (Recall)、F1 分数 (F1 Score)。如果该同学的算法把 2000 张图片都识别成玫瑰，请再次计算出上述指标。

Lecture2 作业

1，假设训练样本集为 $\mathrm{D}=\{(\mathbf{x}_1,y_1)=((3,3)^T,1),(\mathbf{x}_2,y_2)=((4,3)^T,1),(\mathbf{x}_3,y_3)=((1,1)^T,-1)\}$，使用感知器算法设计分类面，并判断测试样本$\mathbf{x}=(0,1)^T$属于哪个类别。

2，对于感知器算法（PLA），假设第 t 次迭代时，选择的是第 n 个样本：$\mathrm{sign}(\boldsymbol{w}^T\mathbf{x}_n)\neq y_n$，$\boldsymbol{w}_{t+1}\leftarrow\boldsymbol{w}_t+y_n\mathbf{x}_n$，下述那个式子正确？

（a）$\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n=y_n$

（b）$\mathrm{sign}\left(\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n\right)=y_n$

（c）$y_n\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n\geq y_n\boldsymbol{w}_t^{\boldsymbol{T}}\mathbf{x}_n$

（d）$y_n\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n<y_n\boldsymbol{w}_t^{\boldsymbol{T}}\mathbf{x}_n$

3，证明：针对线性可分训练样本集，PLA 算法中，当$\boldsymbol{w}_0=\mathbf{0}$，在对分错样本进行了 T 次纠正后，下式成立：$\frac{\mathbf{w}_f^T}{\|\mathrm{w}_f\|}\frac{\mathbf{w}_T}{\|\mathbf{w}_T\|}\geq\sqrt{T}\cdot constant$

4，针对线性可分训练样本集，PLA 算法中，假设对分错样本进行了 T 次纠正后得到的分类面不再出现错分状况，定义：$\boldsymbol{R^2}=\boldsymbol{\max_n}\|\boldsymbol{x}_n\|^{\boldsymbol{2}}$，$\rho=\min_n y_n\frac{\mathbf{w}_f^T}{\|\mathrm{w}_f\|}\mathbf{x}_n$，试证明：$\mathrm{T}\leq\frac{\boldsymbol{R}^2}{\rho^2}$

5，假设训练样本集为 $\mathrm{D}=\{(\vec{x}_1,y_1)=((0.2,0.7)^T,1),(\vec{x}_2,y_2)=((0.3,0.3)^T,1),(\vec{x}_3,y_3)=((0.4,0.5)^T,1),\ (\vec{x}_4,y_4)=((0.6,0.5)^T,1),$

$(\vec{x}_5, y_5) = ((0.1,0.4)^T, 1)$, $(\vec{x}_6, y_6) = ((0.4,0.6)^T, -1)$, $(\vec{x}_7, y_7) = ((0.6,0.2)^T, -1)$, $(\vec{x}_8, y_8) = ((0.7,0.4)^T, -1)$, $(\vec{x}_9, y_9) = ((0.8,0.6)^T, -1)$, $(\vec{x}_{10}, y_{10}) = ((0.7,0.5)^T, -1)\}$，用 Pocket 算法设计分类面。（可借助编程实现，迭代次数最多 20 次，需提交每次迭代的结果）

Lecture2 编程作业

1，分别编写 PLA 算法和 Pocket 算法。

2，（a）产生两个都具有 200 个二维向量的数据集$\boldsymbol{X}_1$和$\boldsymbol{X}_2$。数据集$\boldsymbol{X}_1$的样本来自均值向量$\boldsymbol{m}_1 = [-5,0]^T$、协方差矩阵$\boldsymbol{s}_1 = \boldsymbol{I}$的正态分布，属于“+1”类，数据集$\boldsymbol{X}_2$的样本来自均值向量$\boldsymbol{m}_2 = [0,5]^T$、协方差矩阵$\boldsymbol{s}_2 = \boldsymbol{I}$的正态分布，属于“-1”类，其中 I 是一个 2*2 的单位矩阵。产生的数据中 80%用于训练，20%用于测试。

（b）在上述数据集上分别运用 PLA 算法和 Pocket 算法，利用产生的训练样本集得到分类面，算法中用到的各类超参数自定。

（c）分别在训练集和测试集上统计分类正确率。

（d）分别统计两个算法的运行时间

（e）画出数据集和分类面。

3，重复第 2 题的内容，但数据集$\boldsymbol{X}_1$和数据集$\boldsymbol{X}_2$的均值向量分别改为$\boldsymbol{m}_1 = [1,0]^T$和$\boldsymbol{m}_2 = [0,1]^T$，其他不变。

4，改变算法中的各类超参数、样本数量、样本分布等，讨论实验结果。

Lecture3 习题作业

1，假设训练样本集为 $\mathrm{D}=\{(\vec{x}_1,y_1)=((0.2,0.7)^T,1),(\vec{x}_2,y_2)=((0.3,0.3)^T,1),(\vec{x}_3,y_3)=((0.4,0.5)^T,1),(\vec{x}_4,y_4)=((0.6,0.5)^T,1),(\vec{x}_5,y_5)=((0.1,0.4)^T,1),(\vec{x}_6,y_6)=((0.4,0.6)^T,-1),(\vec{x}_7,y_7)=((0.6,0.2)^T,-1),(\vec{x}_8,y_8)=((0.7,0.4)^T,-1),(\vec{x}_9,y_9)=((0.8,0.6)^T,-1),(\vec{x}_{10},y_{10})=((0.7,0.5)^T,-1)\}$，使用线性回归算法（Linear Regression Algorithm），通过广义逆来求解，并设计这两类的分类函数，讨论结果。（可通过编程计算得到广义逆的结果）

2，根据向量或矩阵的计算性质，证明：

$$\|\mathbf{X}\boldsymbol{w}-Y\|^2=\boldsymbol{w}^T\mathbf{X}^T\mathbf{X}\boldsymbol{w}-2\boldsymbol{w}^T\mathbf{X}^T\mathrm{Y}+\mathrm{Y}^T\mathrm{Y}$$

3，总结梯度下降法、随机梯度下降法、Adagrad、RMSProp、动量法（Momentum）和 Adam 等方法权系数更新表达式。

Lecture3 编程作业

1，分别编写一个用广义逆和梯度下降法来求最小误差平方和最佳解的算法

2，（a）产生两个都具有 200 个二维向量的数据集$\boldsymbol{X}_1$和$\boldsymbol{X}_2$。数据集$\boldsymbol{X}_1$的样本来自均值向量$\boldsymbol{m}_1=[-5,0]^T$、协方差矩阵$\boldsymbol{s}_1=\boldsymbol{I}$的正态分布，属于“+1”类，数据集$\boldsymbol{X}_2$的样本来自均值向量$\boldsymbol{m}_2=[0,5]^T$、协方差

矩阵$\boldsymbol{s}_2 = \boldsymbol{I}$的正态分布，属于“-1”类，其中$I$是一个 2*2 的单位矩阵。产生的数据中 80%用于训练，20%用于测试。

（b）在上述数据集上分别第 1 题的两个算法，利用产生的训练样本集得到分类面，算法中用到的各类超参数自定。

（c）分别在训练集和测试集上统计分类正确率。

（d）画出数据集和分类面。

（e）画出损失函数随 epoch 增加的变化曲线。

3，重复第 2 题的内容，但数据集$\boldsymbol{X}_1$和数据集$\boldsymbol{X}_2$的均值向量分别改为$\boldsymbol{m}_1 = [1,0]^T$和$\boldsymbol{m}_2 = [0,1]^T$，其他不变。

4，改变算法中的各类超参数、样本数量、样本分布等，对于梯度下降法还要改变不同的学习率以及不同的 batch size 和不同 epoch 次数，讨论实验结果。

5，单变量函数为$f(x) = x * cos\,(0.25\pi * x)$，分别用梯度下降法、随机梯度下降法、Adagrad、RMSProp、动量法（Momentum）和 Adam 共 6 种方法，编写程序画图呈现x从初始值为-4、迭代 10 次时x及$f(x)$的每次变化情况，这里对所有算法学习率（或初始学习率）均为 0.4，为防止分母为 0 时给的最小量为ε =1e-6，RMSProp 算法的α=0.9，动量法的λ=0.9，Adam 的 beta1=0.9，beta2=0.999，观察不同算法的变化情况体会各自的差异。如果迭代 50 次，并将 Adam 的 beta1 改成 0.99，其他参数不变，观察不同算法的变化结果。尝试调整上述算法的各种参数，体会上述不同方法的特点。

Lecture4 习题作业

1，已知两类样本的数据如下：

$\omega_1: \{(5,37),(7,30),(10,35),(11.5,40),(14,38),(12,31)\}$

$\omega_2: \{(35,21.5),(39,21.7),(34,16),(37,17)\}$

试用 Fisher 判别函数法，求出最佳投影方向 W，及分类阈值 y_0

2，在 Fisher 判别中，用向量梯度的计算法则证明：$\boldsymbol{S_B w = \lambda S_w w}$

Lecture4 编程作业

1，编程实现 Fisher 线性判别算法。

2，（a）产生两个都具有 200 个二维向量的数据集$\boldsymbol{X_1}$和$\boldsymbol{X_2}$。数据集$\boldsymbol{X_1}$的样本来自均值向量$\boldsymbol{m_1} = [-5,0]^T$、协方差矩阵$\boldsymbol{s_1} = \boldsymbol{I}$的正态分布，属于“+1”类，数据集$\boldsymbol{X_2}$的样本来自均值向量$\boldsymbol{m_2} = [0,5]^T$、协方差矩阵$\boldsymbol{s_2} = \boldsymbol{I}$的正态分布，属于“-1”类，其中 I 是一个 2*2 的单位矩阵。产生的数据中 80%用于训练，20%用于测试。

（b）在上述数据集上运用 Fisher 线性判别算法，在产生的训练样本集上得到最佳投影向量，并计算出分类阈值。

（c）在训练集和测试集上分别统计分类正确率。

（d）画出数据集、最佳投影向量和分类阈值。

Lecture5 习题作业

1，有人说当批量大小为 1 时基于随机梯度下降法（Stochastic Gradient Descent，SGD）的逻辑斯蒂回归（Logistic Regression）算法可以被看作为“软性”的感知器算法（PLA），你认同这个说法吗？请给出你的理由。

2，在 Logistic regression 中当标签 y={+1,-1}时常用交叉熵作为损失函数： $L_{in}(\boldsymbol{w}) = \frac{1}{N}\sum_1^N \ln\left(1 + \exp\left(-y_n \boldsymbol{w}^T \mathbf{x}_n\right)\right)$，请推导出该函数的梯度表达式。

3， 为什么在 Logistic Regression 中不用$L_{in}(\boldsymbol{w}) = (\theta(y\boldsymbol{w}^T\mathbf{x}) - 1)^2$作为损失函数，这里假设$\theta(.)$是 *Sigmoid* 函数，标签 y=$\{+1, -1\}$。

Lecture5 编程作业

1，编程实现 Logistic regression 算法。

2，（a）产生两个都具有 200 个二维向量的数据集$\boldsymbol{X}_1$和$\boldsymbol{X}_2$。数据集$\boldsymbol{X}_1$的样本来自均值向量$\boldsymbol{m}_1 = [-5,0]^T$、协方差矩阵$\boldsymbol{s}_1 = \boldsymbol{I}$的正态分布，属于“+1”类，数据集$\boldsymbol{X}_2$的样本来自均值向量$\boldsymbol{m}_2 = [0,5]^T$、协方差矩阵$\boldsymbol{s}_2 = \boldsymbol{I}$的正态分布，属于“-1”类，其中 I 是一个 2*2 的单位矩阵。产生的数据中 80%用于训练，20%用于测试。

（b）在训练集上利用 Logistic regression 算法得到分类面。

（c）利用得到的分类面对测试集样本进行分类，并给出每个样本属于该类别的概率值。

（d）画出数据集和分类面。

（e）画出损失函数随 epoch 增加的变化曲线。

（f）改变算法中的各类超参数、样本数量、样本分布等，对于梯度下降法还要改变不同的学习率以及不同的 batch size 和不同 epoch 次数，讨论实验结果。

Lecture7-8 作业

1，假设两个样本 $\{(\mathbf{v}_1, y_1) = ((v_1, v_2)^T, 1), (\mathbf{v}_2, y_2) = ((-v_1, -v_2)^T, -1)\}$, 假设 H 是这两个样本的最大间隔分类面，写出其表达式。

2，假设三个样本为 $D = \{(\mathbf{x}_1, y_1) = ((3,0)^T, 1), (\mathbf{x}_2, y_2) = ((0,4)^T, 1), (\mathbf{x}_3, y_3) = ((0,0)^T, -1)\}$,计算这三个样本到平面：$x_1 + x_2 = 1$的距离。

3，假设训练样本集为 $D = \{(\mathbf{x}_1, y_1) = ((0,0)^T, -1), (\mathbf{x}_2, y_2) = ((2,2)^T, -1), (\mathbf{x}_3, y_3) = ((2,0)^T, 1), (\mathbf{x}_4, y_4) = ((3,0)^T, 1)\}$,使用 QP 求解器时，$\boldsymbol{a}_n^T$(n=1,2,3,4)分别为多少？

4，假设训练样本集为：$D = \{(\mathbf{x}_1, y_1) = ((1,1)^T, 1), (\mathbf{x}_2, y_2) = ((2,2)^T, 1), (\mathbf{x}_3, y_3) = ((2,0)^T, 1), (\mathbf{x}_4, y_4) = ((0,0)^T, -1), (\mathbf{x}_5, y_5) = ((1,0)^T, -1), (\mathbf{x}_6, y_6) = ((0,1)^T, -1)\}$, 请分别在$y_n(\boldsymbol{w}^T\mathbf{x}_n + b) \geq 1$和$y_n(\boldsymbol{w}^T\mathbf{x}_n + b) \geq 5$的条件下用 Primal SVM 方法来设计最优分类面$\mathbf{g}(\mathbf{x})$，判断两种情况下的分类面是否一致，指出哪些是候选的支撑向量，并回答如何确认哪些是支撑向量。

5，Hinge Loss 是支撑向量机的误差函数，因此，除了用二次规划求解最佳分类面外，也能用梯度下降法求解，（1）请推导梯度并写出算法流程；（2）假设初始增广权向量$\boldsymbol{w} = (0,0,0)^T$，用第 4 题训练样本集去设计分类面，指出哪些向量在边界上？假设它们都是支撑向量的话，请问最佳权系数向量是否是这些支撑向量的线性组合？

6，假如做了非线性变换后的两个训练样本为：$\{(\boldsymbol{Z}_1,+1)=(\boldsymbol{z},1),(\boldsymbol{Z}_2,-1)=(-\boldsymbol{z},-1)\}$，请写出用于设计硬间隔 SVM 时的拉格朗日函数$L(\boldsymbol{w},\mathrm{b},\alpha)$。

7，对于一个单变量w，假设要在$w\geq 1$和$w\leq 3$这两个线性约束条件下，求$\frac{1}{2}w^2$的最小值，请写出其拉格朗日函数$L(w,\alpha)$以及这个最优问题的 KKT 条件。

8，假如做了非线性变换后的两个训练样本为：$\{(\boldsymbol{Z}_1,+1)=(\boldsymbol{z},1),(\boldsymbol{Z}_2,-1)=(-\boldsymbol{z},-1)\}$，在求解硬间隔 SVM 的对偶问题时，假定得到的最佳$\alpha_1>0$，最佳$\alpha_2>0$，请问最佳 b 为多少？

9，假设有 5566 个样本用以训练对偶硬间隔 SVM 时得到 1126 个支撑向量，请问落在分类面边界上的样本数（也就是候选的支撑向量）有可能是：（a）0；（b）1024；（c）1234；（d）9999。

10，如果两个样本$\mathbf{x}$和$\mathbf{x}'$的内积$\mathbf{x}^T\mathbf{x}'=10$，计算其$\phi_2$核函数$K_{\phi_2}(\mathbf{x},\mathbf{x}')$等于多少？

11，假设训练样本集为：$\mathrm{D}=\{(\mathbf{x}_1,y_1)=((2,2)^T,1),(\mathbf{x}_2,y_2)=((-2,-2)^T,1),(\mathbf{x}_3,y_3)=((2,-2)^T,-1),(\mathbf{x}_4,y_4)=((-2,2)^T,-1)\}$，请用 Dual SVM 来设计最优分类面 $\mathrm{g}(\mathbf{x})$，并指出哪些是支撑向量。

Lecture7-8 编程作业：

1，利用二次规划函数，分别编程实现原问题求解的支撑向量机算法（Primal-SVM）、对偶的支撑向量机算法（Dual-SVM）、和核函数的支撑向量机算法（Kernel-SVM）。

2，（a）产生两个都具有 200 个二维向量的数据集$\boldsymbol{X}_1$和$\boldsymbol{X}_2$。数据集$\boldsymbol{X}_1$的样本来自均值向量$\boldsymbol{m}_1 = [-5,0]^T$、协方差矩阵$\boldsymbol{s}_1 = \boldsymbol{I}$的正态分布，属于“+1”类，数据集$\boldsymbol{X}_2$的样本来自均值向量$\boldsymbol{m}_2 = [0,5]^T$、协方差矩阵$\boldsymbol{s}_2 = \boldsymbol{I}$的正态分布，属于“-1”类，其中$I$是一个 2*2 的单位矩阵。产生的数据中 80%用于训练，20%用于测试。

（b）在上述数据集上分别运用 Primal-SVM、Dual-SVM 和 Kernel-SVM 算法，利用产生的训练样本集得到分类面，其中，Kernel-SVM 中的核函数分别采用四次多项式和高斯核函数，算法中用到的各类超参数自定。

（c）分别在训练集和测试集上统计分类正确率。

（d）对于 Dual-SVM 和 Kernel-SVM 算法，指出哪些样本是支撑向量

（e）画出数据集和分类面、间隔面，并标注出哪些样本是支撑向量，观察是否有边界上的向量不是支撑向量的现象。

3，重复第 2 题的内容，但数据集$\boldsymbol{X}_1$和数据集$\boldsymbol{X}_2$的均值向量分别改为$\boldsymbol{m}_1 = [3,0]^T$和$\boldsymbol{m}_2 = [0,3]^T$，其他不变。

4，改变算法中的超参数、样本数量、样本分布等，讨论实验结果。

5，训练集：中国与日本的沿海城市的经纬度坐标向量，中国标签为+1，日本为标签为-1.

测试集：钓鱼岛的经纬度坐标向量

用支撑向量机设计分类器，（1）判断钓鱼岛属于哪一类；（2）增加几个非海边城市的经纬度坐标进行训练，判断这些城市是否影响分类结果，是否为支撑向量。

Lecture 9 习题作业

1，假设有如下训练样本：$\mathbf{x}_1 = (0,0)^T$属于第一类，$\mathbf{x}_2 = (1,1)^T$属于第二类，$\mathbf{x}_3 = (-1,1)^T$属于第三类，请用多类分类中的 OVO（One-versus-one）策略，设计上述三类别的两两分类器，并分析测试样本$\mathbf{x} = (1,-2)^T$属于哪个类别。

2，现有四个样本，假设样本（3，0）和（3，6）属于第一类，样本（0，3）属于第二类，样本（-3，0）属于第三类，请用 Softmax 算法设计出这三个类别的分类器（假设这三个类别的初始权向量均为零向量，迭代步长取 1，需要写出计算过程）。

Lecture 9 编程作业

1，给定 IRIS 数据集，该数据集有三类目标，每个类别有 50 个样本，每个样本有四维特征。实验时每个类别随机选 30 个样本进行训练，另外 20 个样本用于测试。

（a）以感知器算法为基础分类算法，编写一个 OVO 多类分类器算法，对上述数据集进行实验，分析结果。

（b）编写 Softmax 算法实现多类别分类，对上述数据集进行实验，分析结果。

2，给定 MNIST 数据集，该数据集每个样本为 28*28 大小的灰度图像，有 0 到 9 共 10 个类别的手写体数字，其中训练样本 60000，测试样

本 10000，编写 Softmax 算法对该数据集实现分类，权向量初始值由均值为 0、标准差为 0.01 的正态分布产生的随机数得到，统计此时测试集的分类精度（正确分类的样本数/总样本数）。训练时的 batch size 为 256，一共训练 10 遍 epoch，画出训练时的损失函数、训练集上的分类精度和测试集上的分类精度随 epoch 增加的变化曲线。训练完成后，在测试集上随机抽取 10 个样本，观察分类结果。

Lecture10-11 作业

1，假设 $g_0(\vec{x})=1$，以下哪一组（$\alpha_0,\alpha_1,\alpha_2$）允许 $G(\vec{x})=sign(\sum_{t=0}^{2}\alpha_t g_t(\vec{x}))$实现$OR(g_1,g_2)$的功能。（a）（-3，+1，+1）；（b）（-1，+1，+1）；（c）（+1，+1，+1）；（d）（+3，+1，+1）。

2，在 3-5-1 的神经网络中，网络参数有多少？

3，画出（$x+y$）^2z的计算图，当 x=1,y=2,z=-6 时，写出前向传播的数值和反向传播的梯度值。

4，计算 Sigmoid 函数、双曲正切函数和 ReLU 函数的导数函数，分析这三个函数作为激活函数时的优缺点。

5，对于一幅 300*300 大小的彩色（RGB）图像，（1）如果输入端与有 100 个神经元的第一层隐含层用全链接方式（Fully Connected neural Network）连接时，请问这一层会包含多少参数？（2）如果用 100 个 5*5*3 大小的滤波器作卷积操作，那么这一层的参数为多少？如果滤波器移动步长（stride=1）为 1，经过卷积计算后的输出端神经元个数有多少？

6，某一个卷积神经网络结构如下：

（i）输入层 Input 的 RGB 图像大小是 227*227*3。

（ii）第 1 层卷积层 Conv-1 是通过对输入图像用 96 个 11*11*3 大小的滤波器通过步长(stride)为 4，不做边缘填充（padding）得到的。

（iii）接下来是池化层 MaxPool-1，它用 3*3 尺寸、步长为 2 对 Conv-1 做 Max Pooling 操作。

（iv）然后我们对图像进行边缘填充，填充值为 2（如原来图像大小为 7*7 时，做填充值为 2 的填充后，图像大小变为 11*11），用 256 个 5*5 大小的滤波器按步长为 1，做第二次卷积操作，得到 Conv-2 层。

（v）再接一个池化层 MaxPool-2，它用 3*3 尺寸、步长为 2 做一次 Max Pooling 操作。

（vi）MaxPool-2 层输出去接一个有 4096 个神经元的全连接层 FC-1。

（vii）再接一个全连接层 FC-2 实现对 1000 个类别的分类。

请计算：（1）输入层到 Conv-1 层的参数量有多少？（2）经过池化层 MaxPool-1 后的神经元是多少？（3）经过第二次卷积操作后的图像大小为多少？（4）MaxPool-2 层到 FC-1 层的参数量是多少？（5）FC-1 层到 FC-2 层的参数量是多少？

7，有训练样本集为：$D = \{(\vec{x}_1, y_1) = ((1,1)^T, 1), (\vec{x}_2, y_2) = ((-1,-1)^T, 1), (\vec{x}_3, y_3) = ((-1,1)^T, -1), (\vec{x}_4, y_4) = ((1,-1)^T, -1)\}$，假设某神经网络结构为第一层有两个神经元，第二层有三个神经元，第三层有一个神经元，前两层每个神经元的激活函数为ReLU（即$x_d^{(l)} = \max(0, s_d^{(l)})$，这里$s_d^{(l)}$代表第l层第d个神经元的输入，$x_d^{(l)}$代表该神经元的输出），第三层为线性输出，即$\hat{y} = s_1^{(3)}$。误差函数为：$E_{in} = \frac{1}{N}\sum_n (y_n - \hat{y}_n)^2$，学习率为0.01。假设初始权系数矩阵定义如下：

$$\mathbf{w}_0^{(1)} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{w}_0^{(2)} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{w}_0^{(3)} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

其中w的下标0代表迭代次数为0（即初始状态），上标数字分别代表

第1、2、3层。要求将上述训练样本集的样本用反向传播法按顺序进行一轮训练，写出每一次迭代时各层的权系数矩阵，即：t=1时，进入样本$\vec{x}_1$，得到$\mathbf{w}_1^{(1)}$、$\mathbf{w}_1^{(2)}$和$\mathbf{w}_1^{(3)}$；t=2时，进入样本$\vec{x}_2$，得到$\mathbf{w}_2^{(1)}$、$\mathbf{w}_2^{(2)}$和$\mathbf{w}_2^{(3)}$；t=3时，进入样本$\vec{x}_3$，得到$\mathbf{w}_3^{(1)}$、$\mathbf{w}_3^{(2)}$和$\mathbf{w}_3^{(3)}$；t=4时，进入样本$\vec{x}_4$，得到$\mathbf{w}_4^{(1)}$、$\mathbf{w}_4^{(2)}$和$\mathbf{w}_4^{(3)}$

8，写出你对前 16 个学时老师授课的评价和建议。

Lecture10-11 编程作业题

1，IRIS 数据集有三类目标，每个类别有 50 个样本，每个样本有四维特征。自行设计神经网络实现对这三个目标的识别，实验时每个类别随机选 30 个样本进行训练，另外 20 个样本用于测试。希望能通过设计不同的隐含层数、每层的节点数、不同的学习率、不同的激活函数等对实验结果进行讨论。

2，LeNet 网络结构如下：

（i）第 1 层卷积层 Conv-1： 6 个 5*5*1 大小的滤波器， stride=1，padding=2，接 Sigmoid 做激活函数；

（ii）接下来是池化层 AvePool-1，它以 2*2、stride=2 做 Average Pooling 操作；

（iii）第 2 层卷积层 Conv-2： 16 个 5*5*6 大小的滤波器， stride=1，padding=0，接 Sigmoid 做激活函数；

（iv）再接一个池化层 AvePool--2，它以 2*2、stride=2 做 Average Pooling 操作；

（v）对 AvePool--2 层输出做了 Flatten 操作后，与 120 个神经元做全连接，构成 FC-1，Sigmoid 做激活函数；

（vi）再与 84 个神经元做全连接，构成 FC-2，Sigmoid 做激活函数；

（vii）再全连接 10 个神经元输出，用 Softmax 完成 10 个类别的分类。

编写上述网络结构的代码，对 MNIST 数据集实现分类，训练时的 batch size 为 256，一共训练 10 遍 epoch，画出训练时的损失函数、训练集上的分类精度和测试集上的分类精度随 epoch 增加的变化曲线。训练完成后，在测试集上随机抽取 10 个样本，观察分类结果。

1. 对于两分类问题，证明最小风险贝叶斯规则可表示为

若 $l(x)=\dfrac{p(\mathrm{X}|\omega_1)}{p(\mathrm{X}|\omega_2)}>\dfrac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\dfrac{P(\omega_2)}{P(\omega_1)}$，则决策$\mathrm{X}\in\omega_1$；否则$\mathrm{X}\in\omega_2$。

2. 有两类样本 ω1 和 ω2，已知先验概率 P(ω1)=0.2 和 P(ω2)=0.8，类概率密度函数如下：

$$p(x|\omega_1)=\begin{cases}x & 0\le x<1\\ 2-x & 1\le x\le 2\\ 0 & \text{其他}\end{cases}\qquad p(x|\omega_2)=\begin{cases}x-1 & 1\le x<2\\ 3-x & 2\le x\le 3\\ 0 & \text{其他}\end{cases}$$

（1）求贝叶斯最小误判概率准则下的判决域，并判断样本 x=1.5 属于哪一类；

（2）求总错误概率 P(e)；

（3）假设正确判断的损失 $\lambda_{11}=\lambda_{22}=0$，误判损失分别为 λ_{12} 和 λ_{21}，若采用最小损失判决准则，λ_{12} 和 λ_{21} 满足怎样的关系时，会使上述对样本 x=1.5 的判断相反？

3. 已知两类问题，有 $\mu_1=\begin{bmatrix}4\\4\end{bmatrix}$, $\mu_2=\begin{bmatrix}8\\8\end{bmatrix}$, $\sum_1=\begin{bmatrix}2&0\\0&2\end{bmatrix}$, $\sum_2=\begin{bmatrix}2&0\\0&2\end{bmatrix}$，先验概率为 P(ω1) =0.8, P(ω2) = 0.2。

(1) 求两类的贝叶斯决策分界面。

(2) 要求根据 Bayes 决策，对样本 x (3, 3)进行分类。

1. 设 x 服从概率密度函数 $p(x|\theta)=\begin{cases}(\theta+1)x^{(\theta+1)} & (0<x<1)\\ 0 & (otherwise)\end{cases}$，样本 $x_1,\ldots,x_n$ 是从分布 $p(x|\theta)$ 中独立抽取，试用最大似然估计参数 θ。

2.Consider Hidden Markov Model. The hidden states are $\{\omega_1,\omega_2,\omega_3\}$, and the visible states are $\{v_1,v_2,v_3\}$. The transition probabilities are $a_{ij}=\begin{bmatrix}1 & 0 & 0\\ 0.3 & 0.3 & 0.4\\ 0.2 & 0.4 & 0.4\end{bmatrix}$，

$b_{jk}=\begin{bmatrix}1 & 0 & 0\\ 0 & 0.6 & 0.4\\ 0 & 0.2 & 0.8\end{bmatrix}$.

The initial hidden state is ω_2, and initial visible state is v_2. Try to get the probability to generate the particular visible sequence $V^3=\{v_2,v_3,v_1\}$.