

北京邮电大学 2016——2017 学年第二学期

《模式识别》期末考试试题（A 卷）

考 试 注 意 事 项	一、学生参加考试须带学生证或学院证明，未带者不准进入考场。学生必须按照监考教师指定座位就坐。 二、书本、参考资料、书包等物品一律放到考场指定位置。 三、学生不得另行携带、使用稿纸，要遵守《北京邮电大学考场规则》，有考场违纪或作弊行为者，按相应规定严肃处理。 四、学生必须将答题内容做在试题答卷上，做在试题及草稿纸上一律无效
----------------------------	--

一、基本概念题(除了公式，每题答案不超过 20 字，2 分\*10)：

①写出后验概率的表达式，描述它和先验概率的区别。

②写出线性判别函数的表达式，画图指出参数的物理意义，描述它和神经元模型的联系。

③为什么说 SVM 是最优线性分类器，它相比感知器算法的优点是？

④Logistic Regression 的优化表达式，它的功能与 SVM 的作用有什么区别？

⑤单层神经网络有什么局限性?如何将其扩展处理复杂的非线性分类问题?

⑥多层神经网络的主流参数学习算法是什么?试用个公式说明参数学习的原理。

⑦针对图像处理和识别设计的神经网络是什么?为什么它的参数数量远小于全连接网络?

⑧主成分分析获得的特征有什么特点?主成分投影基向量与傅里叶变换基函数有什么区别?

⑨当误差数值的符合什么分布时,我们常用的均方误差是理论最优的?为什么?

⑩当需要衡量两个分布的误差时,一般采用什么物理量作为损失函数?试写出该函数形式。

二、技术应用速答题(每题答案不超过 10 字, 1 分\*10):

①根据用户画像(上网特征)预测用户购买某类商品的概率, 用什么方法?

②为了获得两类分类任务中最靠谱的投影特征, 用什么方法?

③两类特征的类条件密度函数未知, 对测试样本分类并求出后验概率, 用什么方法?

④在训练样本数量较少的两类分类任务中, 一般认为什么方法是最靠谱的?

⑤在高维数据分析任务中, 什么方法是最常用的数据进行降维方法?

⑥系统要对多类样本进行分类, 而手上的分类器只能处理两类, 如何扩展?

⑦当采用线性 SVM 分类器的分类效果较差, 应该首先试验什么改进分类器?

⑧我们常用的正态分布的均值和方差公式, 是用什么算法估计出来的?.

⑨得到一批无标记(类别标签的)数据, 用什么算法可以对其进行自动“分类”

⑩深度学习或者深度神经网络中的“深度”是指

三、综合设计题:假设您是 Facebook 的系统架构师,需求是使用模式识别技术设计一个根据用户上传图片预测出 TA 当前年龄的系统,请您描述主要设计步骤和技术流程框图(设计步骤应包含一般模式识别系统的设计过程,技术流程应包含候选的算法和目标函数,不能加页,10 分)。