1、什么是模式识别：使计算机模仿人的感知能力，从感知数据中提取信息（判别物体和行为）的过程。
2、模式识别包括四个组成部分：数据获取、预处理、特征提取与选择、分类决策。
3、相似性度量满足的条件：a、对称性 b、非负性 c、自身的相似性最大 d、是关于距离的单调函数。
4、做模式识别的前提条件是每个模式类满足紧致性。
5、Bayes 决策的三个前提条件：a、类别数确定 b、各类的先验概率 $P(\omega_i)$ 已知 c、各类的条件概率密度函数 $p(\mathbf{x}|\omega_i)$ 已知。

$$P(\omega_i|\mathbf{x})=\frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_j p(\mathbf{x}|\omega_j)P(\omega_j)}$$

6、Bayes 最小错误率对应于最大后验 $P(\omega_i|\mathbf{x})$
7、损失函数的定义：(N 类问题)做出决策 $D(\mathbf{x})=\omega_i$，但实际上 $\mathbf{x}\in\omega_j$，受到的损失定义为：$\lambda_{i,j}=\lambda(D(\mathbf{x})=\omega_i|\omega_j)\quad i,j=1,2,\cdots,N$

条件期望风险：获得观测值 x 后，决策 D(x)造成的损失对 x 实际所属类别的各种可能的平均，也称为条件风险 R(D(x)|x)。

$$\begin{aligned}R(D(\mathbf{x})|\mathbf{x})&\\&=E\left[\lambda(D(\mathbf{x}),\omega_i)\right]\\&=\sum_i\lambda(D(\mathbf{x})|\omega_i)P(\omega_i|\mathbf{x})\end{aligned}$$

$$\begin{aligned}\hat{D}(\mathbf{x})&=\underset{D}{\operatorname{argmin}}\,R(D(\mathbf{x})|\mathbf{x})\\&=\underset{D}{\operatorname{argmin}}\sum_i\lambda(D(\mathbf{x}),\omega_i)P(\omega_i|\mathbf{x})\end{aligned}$$

8、最小风险决策的计算：a、根据 Bayes 公式计算后验概率 $P(\omega_j|\mathbf{x})$ b、根据后验概率及给定的损失矩阵，算出每个决策的条件风险 $R(\alpha_i|\mathbf{x})$ c、按最小的条件风险进行决策。
9、一元正态分布：

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$E\{x\}=\int_{-\infty}^{\infty}xp(x)dx=\mu$$

$$E\{(x-\mu)^2\}=\int_{-\infty}^{\infty}(x-\mu)^2p(x)dx=\sigma^2$$

10、欧式距离和马氏距离：
$$\begin{aligned}\gamma^2&=M(\mathbf{x},\boldsymbol{\mu})\\&=(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\end{aligned}$$

11、最大似然估计：a、定义最大似然函数 $l(\boldsymbol{\theta})=p(K|\boldsymbol{\theta})=p(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N|\boldsymbol{\theta})$ b、对数化
$$=\prod_{k=1}^{N}p(\mathbf{x}_k|\boldsymbol{\theta})$$

（一般似然函数为指数的形式）$H(\boldsymbol{\theta})=\sum_{k=1}^{N}\ln p(\mathbf{x}_k|\boldsymbol{\theta})$ c、求偏导 $\nabla_{\boldsymbol{\theta}}H(\boldsymbol{\theta})|_{\hat{\theta}_{ML}}=\sum_{k=1}^{N}\nabla_{\boldsymbol{\theta}}\ln p(x_k|\boldsymbol{\theta})|_{\hat{\theta}_{ML}}=0$

12、Fisher 准则：找到一个最合适的投影轴，使两类样本在该轴上投影之间的距离尽可能远，而每一类样本的投影尽可能紧凑，从而使分类效果为最佳。

各类样本均值向量 mi　　$\mathbf{m}_i = \frac{1}{N_i}\sum_{x\in K_i}\mathbf{x} \quad i=1,2$

样本**类内离散度**矩阵 Si 与总类内离散度矩阵 Sw

$\mathbf{S}_i = \sum_{\mathbf{x}\in \chi_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T, \quad i=1,2 \quad \mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2$

样本**类间离散度**矩阵 Sb：$\mathbf{S}_b = (\mathbf{m}_1-\mathbf{m}_2)(\mathbf{m}_1-\mathbf{m}_2)^T$

离散度矩阵在形式上与协方差矩阵很相似，但协方差矩阵是一种总体期望值，而离散矩阵只是表示有限个样本在空间分布的离散程度

$$J_F(w) = \frac{\tilde{S}_b}{\tilde{S}_1 + \tilde{S}_2} = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \qquad \mathbf{w}^* = \underset{\mathbf{w}}{\mathrm{argmax}}\, J_F(\mathbf{w})$$

令 $\mathbf{w}^T S_w \mathbf{w} = \mathrm{c} \neq 0$　定义Lagrange函数: $L(\mathbf{w},\lambda) = \mathbf{w}^T S_b \mathbf{w} - \lambda(\mathbf{w}^T S_w \mathbf{w} - c)$

令: $\frac{\partial L(\mathbf{w},\lambda)}{\partial \mathbf{w}} = S_b\mathbf{w} - \lambda S_w \mathbf{w} = 0 \qquad S_w^{-1} S_b \mathbf{w} = \lambda \mathbf{w} \qquad \lambda\mathbf{w} = S_w^{-1} S_b \mathbf{w} = S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)(\mathbf{m}_1-\mathbf{m}_2)^T\mathbf{w}$

$= S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)R$

$\mathbf{w}^* = \frac{R}{\lambda} S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)$； $S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)$

13、结合 K-L 变换叙述人脸识别的经典流程：（1）、对向量 **x** 用确定的完备正交归一向量系 uj 展开，对应到人脸识别问题上就是一个脸可以由无穷多个本征脸进行重构（2）、为了研究问题的方便我们用有限项估计 X，对应人脸识别是一个脸由有限个本征脸进行重构，假设有 d 个则 $\hat{\mathbf{x}} = \sum_{j=1}^{d} y_j \mathbf{u}_j$

（3）、下面的思路是寻找 uj　即对应的本征脸。

（4）、由于是用 d 个本征脸来重构一张脸 x，所以存在着一定的误差，下面我们求该估计的均方误差：

$$\varepsilon = E\left[(\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}})\right]$$

$$\varepsilon = E\left[\sum_{j=d+1}^{\infty} y_j^2\right] = E\left[\sum_{j=d+1}^{\infty} \mathbf{u}_j^T \mathbf{x}\mathbf{x}^T \mathbf{u}_j\right] \qquad \mathbf{R} = \left[r_{ij} = \mathrm{E}(x_i x_j)\right] = E\left[\mathbf{x}\mathbf{x}^T\right] \ \varepsilon = \sum_{j=d+1}^{\infty} \mathbf{u}_j^T E\left[\mathbf{x}\mathbf{x}^T\right]\mathbf{u}_j = \sum_{j=d+1}^{\infty} \mathbf{u}_j^T \mathbf{R}\mathbf{u}_j$$

用 Lagrange 乘子法：

if $\mathbf{R}\mathbf{u}_j = \lambda_j \mathbf{u}_j$ then $\varepsilon = \sum_{j=d+1}^{\infty} \mathbf{u}_j^T \mathbf{R}\mathbf{u}_j$ 取得极值

以相关矩阵 R 的 d 个本征向量为基向量来展开 **x** 时，其均方误差为

$$\varepsilon = \sum_{j=d+1}^{\infty} \lambda_j$$

当取矩阵 R 的 d 个最大本征值对应的本征向量来展开 **x** 时，其截断均方误差最小。

14、遗传算法的整个过程：a、初始化:设置进化代数计数器 t=0，设置最大进化代数 T，随机生成 M 个个体作为初始群体 P(0)。b、个体评价：计算群体 P(t)中各个个体的适应度。c、选择运算:将选择算子作用于群体。选择的目的是把优化的个体直接遗传到下一代或通过配对交叉产生新的个体再遗传到下一代。选择操作是建立在群体中个体的适应度评估基础上的。d、交叉运算；将交叉算子作用于群体。所谓交叉是指把两个父代个体的部分结构加以替换重组而生成新个体的操作。遗传算法中起核心作用的就是交叉算子。e、变异运算：将变异算子作用于群体。即是对群体中的个体串的某些基因座上的基因值作变动。群体 P(t)经过选择、交叉、变异运算之后得到下一代群体 P(t 1)。f、终止条件判断:若 t=T,则以进化过程中所得到的具有最大适应度个体作为最优解输出，终止计算。（中止条件：达到适应度函数或达到进化规定的代数）

初始参数：适应度函数、变异率、交叉率、代数、初始种群。

遗传算法的优缺点：遗传算法与传统的优化方法（枚举，启发式等）相比较，以生物进化为原型，具有很好的收敛性，在计算精度要求时，计算时间少，鲁棒性高等都是它的优点。在现在的工作中，遗传算法（1972 年提出）已经不能很好的解决大规模计算量问题，它很容易陷入“早熟”。

15、K 均值的迭代思想：a、初始化：选择 c 个代表点 p1, p2, …,pc　b、建立 c 个空聚类列表： K1, K2, …,Kc　c、按照最小距离法则逐个对样本 x 进行分类：d、计算 J 及用各聚类列表(Ki)计算聚类均值(pi) ，作为各聚类新的代表点（更新代表点）e、若 J 不变或代表点未发生变化，则停止。否则转 2。

存在的问题：a、取初始中心对分类有影响　b、类的个数未知

16、例：两类细胞识别问题：正常类(ω1)和类异常(ω2)根据已有知识和经验，两类的先验概率为正常(ω1)： $P(\omega_1)=0.9$ 异常(ω2)： $P(\omega_2)=0.1$ 对某一样本观察值 x，通过计算或查表得到：$p(\mathbf{x}|\omega_1)=0.2$， $p(\mathbf{x}|\omega_2)=0.4$ 如何对细胞 **x** 进行分类？

利用贝叶斯公式计算两类的后验概率：

$$P(\omega_1|\mathbf{x})=\frac{P(\omega_1)p(\mathbf{x}|\omega_1)}{\sum_{j=1}^{2}P(\omega_j)p(\mathbf{x}|\omega_j)}=\frac{0.9\times0.2}{0.9\times0.2+0.1\times0.4}=0.818$$

$$P(\omega_2|\mathbf{x})=\frac{P(\omega_2)p(\mathbf{x}|\omega_2)}{\sum_{j=1}^{2}P(\omega_j)p(\mathbf{x}|\omega_j)}=\frac{0.4\times0.1}{0.2\times0.9+0.4\times0.1}=0.182$$

$$j=\underset{i}{\operatorname{argmax}}\,P(\omega_i|\mathbf{x})=1$$

$$\mathbf{x}\in\omega_1$$

17、OTSU：a、图像有 *L* 阶灰度，*ni* 是灰度为 *i* 的像素数，图像总像素数 $N=n1+n2+\cdots+nL$ *b*、灰度为 *i* 的像素概率：$pi=ni/N$ *c*、*类间方差*： $\sigma_B^2(k)=\omega_1(\mu_1-\mu)^2+\omega_2(\mu_2-\mu)^2$

$$\mu_1=\sum_{i=1}^{k}ip_i,\ \mu_2=\sum_{i=k+1}^{L}ip_i,\ \mu=\sum_{i=1}^{L}ip_i$$

灰度图像阈值：

$$t=\underset{k=1}{\overset{L}{\operatorname{argmax}}}\,\sigma_B^2(k)$$

$$\omega_1=\sum_{i=1}^{k}p_i,\ \omega_2=\sum_{i=k+1}^{L}p_i=1-\omega_1$$