深 圳 大 学 实 验 报 告

课程名称 机器学习

项目名称 实验四：SVM 算法

学　　院 计算机与软件学院

专　　业 软件工程

指导教师 赖志辉

报 告 人 郑杨 学号 2020151002

实验时间 2022 年 4 月 27 日至 2022 年 4 月 28 日

实验报告提交时间 2022 年 4 月 28 日

教务处制

目录

一、实验目的与要求

1、分别给出经典的/ 软间隔/核-SVM 的优化问题并推导其求解优化过程，实现经典的 SVM 算法进行图像识别；在二维平面对二类问题给出 support vector 的一个示例。
2、用 PCA、LDA 算法提取前 10,20,30,...,160 维的图像特征，然后再用 SVM 进行识别，并比较识别率。

二、实验内容与方法

1、硬间隔 SVM 的基本型推导
2、硬间隔 SVM 的对偶型推导
3、硬间隔核化 SVM 的推导
4、软间隔核化 SVM 的推导
5、经典 SVM 在二维平面上的实验示例与对人脸图像进行识别的实验
6、使用 PCA、LDA 降维后再使用 SVM 分类进行人脸识别实验

三、实验步骤与过程

1. 硬间隔 SVM 的基本型

1.1 线性二分类问题

二分类问题的训练集定义为：

$$D \coloneqq \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\}$$

其中，$x_i \in \mathbb{R}^d$ 为第 i 个训练样本向量（d 维列向量）；$y_i \in \{-1,1\}$ 表示第 i 个训练样本的标签；m 表示训练样本总数。
给定训练集 D 之后，二分类问题的目标就是找到一个假设函数 h：

$$h: \mathbb{R}^d \to \{1,-1\}$$

给定一个向量 x_i，$h(x_i) \in \{1,-1\}$ 称为模型预测值，那么模型训练的目标就是使得预测值 $h(x_i)$ 与真实值 y_i 一致，即 $h(x_i) = y_i$。那么模型的优化目标为：

$$h(x_i) = \begin{cases} 1, if\ (y_i = 1) \\ -1, if\ (y_i = -1) \end{cases}$$

对于最简单的线性二分类模型而言，有：

$$h(x_i) \coloneqq sign(w^T x_i + b)$$

那么模型的优化目标为：

$$sign(w^T x_i + b) = \begin{cases} 1, if\ (y_i = 1) \\ -1, if\ (y_i = -1) \end{cases}$$

这个目标等价于：

$$\begin{cases} w^T x_i + b > 0, if\ (y_i = 1) \\ w^T x_i + b < 0, if\ (y_i = -1) \end{cases}$$

进一步可以简化为：

$$y_i(w^T x_i + b) > 0$$

不难理解，线性可分二分类模型的几何意义就是在空间 $\mathbb{R}^d$ 中找到一个 $d-1$ 维超平面将正样本分开，如图 1 所示。以二维平面上的样本点为例，超平面就是一条直线。

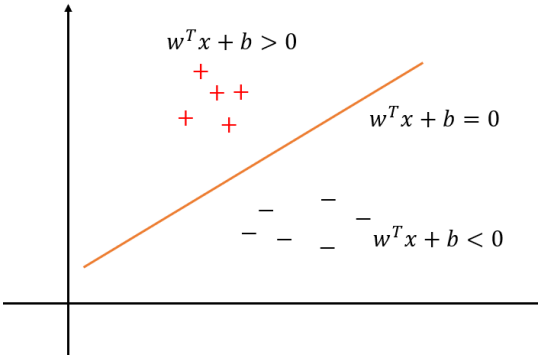

图 1：超平面划分

1.2 最大间隔分类器

不难想到，有无数个超平面可以对正负样本点进行划分，如图 2 所示。对于不同分类器而言，**归纳偏好**不同使得确定出来的超平面不同。

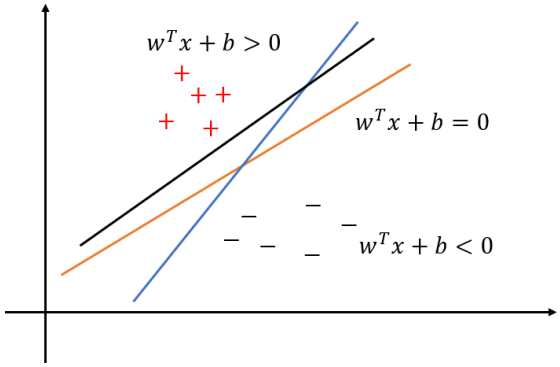

图 2：无数个超平面

硬间隔 SVM 的归纳偏好是找到离正负样本点都尽可能远的划分超平面。原因在于，由于我们在实际应用时关注的是模型的泛化误差，也就是模型对于未知样本点的预测误差，而不太注重训练误差（对于训练样本的预测误差），由于训练集可能存在噪声样本点，使用其他模型可能使得过拟合训练样本。而硬间隔 SVM 的偏好是尽可能远离样本点，这样做使得模型的泛化性与鲁棒性较强。

那么该如何对**离正负样本都比较远**进行数学形式化表达呢，硬间隔 SVM 使用间隔 $\gamma \in \mathbb{R}$ 对超平面与正负样本之间的最小距离进行描述。间隔 γ 的定义为训练集中与超平面最近的样本与超平面的距离的**两倍**：

$$\gamma := 2 \min_i \frac{1}{\|w\|} |w^T x_i + b|$$

由于**划分超平面离正负样本尽可能远**等价于**正负样本与划分超平面的距离尽可能远**，也就是让离超平面最近的样本与超平面的距离尽可能大，也就是让 γ 尽可能大。于是模型的优化问题可写为：

$$\max_{w,b} \min_i \frac{2}{\|w\|} |w^T x_i + b|$$
$$s.t. y_i (w^T x_i + b) > 0, i = 1, 2, ..., m$$

其中，第一行表示优化目标，第二行表示优化目标需要满足的约束（超平面需要把正负样本划分开）。

需要注意的是，最终确定的划分超平面恰好在位于最近的正负样本正中间，如图 3 所示，如果该超平面往右下角移动，那么负样本与超平面距离最近，最终的间隔会减小，往左上角移动也是同样的道理。

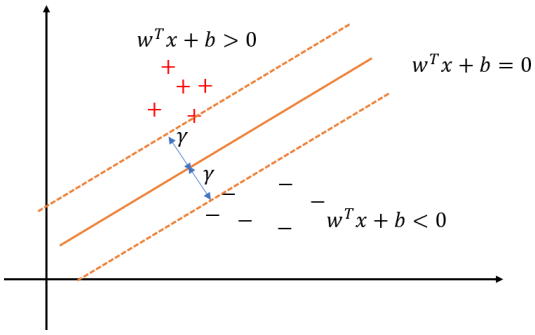

图 3：超平面处于中间

1.3 简化优化问题

原优化问题如下，过于复杂，我们想要对它进行简化。

$$\max_{w,b} \min_i \frac{2}{\|w\|} |w^T x_i + b|$$
$$s.t. y_i (w^T x_i + b) > 0, i = 1, 2, ..., m$$

我们发现，如果 w^*, b^* 是该优化问题的解，那么对于任意 $k > 0$， kw^*, kb^* 仍为该优化问题的解，也就是说 w, b 的缩放不影响解。为了简化问题，我们可以通过调整 w, b 使得：

$$\min_i |w^T x_i + b| = 1$$

故优化目标可以简化为：

$$\max_{w,b}\frac{2}{\|w\|}=\min_{w,b}\frac{1}{2}\|w\|=\min_{w,b}\frac{1}{2}\|w\|^2=\min_{w,b}\frac{1}{2}w^Tw$$

对于约束条件，由于 $y_i(w^Tx_i+b)>0$ 且 $y_i\in\{-1,1\}$，那么有：

$$y_i(w^Tx_i+b)=|w^Tx_i+b|$$

故原约束条件可以变为：

$$\min_i y_i(w^Tx_i+b)=1$$

即：

$$y_i(w^Tx_i+b)\geq 1$$

故原优化问题就简化为：

$$\min_{w,b}\frac{1}{2}w^Tw$$
$$s.t.y_i(w^Tx_i+b)\geq 1,i=1,2,...,m$$

这就是**硬间隔 SVM 的基本型。**

1.4 二次规划问题

对于形如

$$f(x):=x^TAx+B^Tx+C$$

的函数称为二次函数，其中 $A\in\mathbb{R}^{n\times n},B\in\mathbb{R}^n,C\in\mathbb{R}$。

当 $A\geq 0$（A 为半正定矩阵）时，f 为凸函数；当 $A<0$（A 为负定矩阵）时，f 为凹函数。

二次规划是指目标函数为二次函数，约束是线性不等式约束的一类优化问题，如下：

$$\min_u\frac{1}{2}u^TQu+t^Tu$$
$$s.t.c_i^Tu\leq d_i,i=1,2,...,m$$

事实上也支持等式约束，等式约束可以转化为两个不等式约束。

当二次规划的目标函数为凸函数时（Q 为半正定矩阵时），称为凸二次规划。

对于凸二次规划问题，若约束条件确定的可行域不为空，且目标函数在此可行域有下界，那么该问题有全局最小值。

我们可以证明，**硬间隔 SVM** 的优化问题本质上就是一个**凸二次优化问题**，证明如下：

硬间隔 SVM 的优化问题为：

$$\min_{w,b} \frac{1}{2} w^T w$$
$$s.t. y_i(w^T x_i + b) \geq 1, i = 1, 2, ..., m$$

我们定义：

$$u := \begin{bmatrix} w \in \mathbb{R}^d \\ b \in \mathbb{R} \end{bmatrix} \in \mathbb{R}^{d+1}$$
$$Q := \begin{bmatrix} I \in \mathbb{R}^{d \times d} & \mathbf{0} \in \mathbb{R}^{d \times 1} \\ \mathbf{0} \in \mathbb{R}^{1 \times d} & 0 \in \mathbb{R} \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)}$$

$$t := \mathbf{0} \in \mathbb{R}^{d \times 1}$$
$$c_i := -y_i \begin{bmatrix} x_i \in \mathbb{R}^d \\ 1 \in \mathbb{R} \end{bmatrix} \in \mathbb{R}^{d+1}, i = 1, 2, ..., m$$
$$d_i := -1 \in \mathbb{R}, i = 1, 2, ..., m$$

那么**硬间隔 SVM 的基本型**可以写为二次规划的形式，又因为此时 Q 为一个半正定矩阵，故问题变为凸二次规划问题，包含 $d+1$ 个优化变量， m 项不等式约束。

2. 硬间隔 SVM 的对偶型

在此之前我们已经得出了硬间隔 SVM 的基本型，但是基本型是一个带不等式约束的凸二次优化问题，比较复杂，我们想要继续简化该问题。

2.1 拉格朗日乘子法求解带约束优化问题的一般流程

- **把约束优化问题写成基本形式**

基本形式具有 m 个不等式约束和 n 个等式约束。这个形式也称为主问题和基本型。

$$\min_u f(u)$$
$$s.t. g_i(u) \leq 0, i = 1, 2, ..., m$$
$$h_j(u) = 0, j = 1, 2, ..., n$$

- **构建拉格朗日函数**

拉格朗日乘子法的本质就是把约束优化问题转化为无约束的优化问题，使求解更加简单。若无法转化为无约束优化问题，就尽可能简化优化项。

构造拉格朗日函数为：

$$L(u, \alpha, \beta) := f(u) + \sum_{i=1}^{m} \alpha_i g_i(u) + \sum_{j=1}^{n} \beta_j h_j(u)$$

那么基本型可以等价为：

$$\min_{u} \max_{\alpha,\beta} L(u,\alpha,\beta)$$
$$s.t.\alpha_i \geq 0, i = 1,2,...,m$$

可以证明该优化问题与基本型等价，由于篇幅问题这里就不作证明了。

- **交换 min 和 max 的顺序得到对偶问题并求解**

由于转化后的问题内层仍然是一个带不等式约束的优化问题，我们希望将其转化为其对偶问题进行求解：

$$\max_{\alpha,\beta} \min_{u} L(u,\alpha,\beta)$$
$$s.t.\alpha_i \geq 0, i = 1,2,...,m$$

这样内层问题就是一个无约束的优化问题，可以直接求偏导解决。由对偶问题以及硬间隔 SVM 满足 Slater 条件，可以证明对偶问题和原问题的解相等。

求解对偶问题得到解 (α^*, β^*)

- **利用 KKT 条件得到主问题的最优解**

KKT 条件指的是优化问题在最优值处时的解 (u^*, α^*, β^*) 必须满足的条件，包括：

- 主问题可行：满足主问题的约束 $g_i(u^*) \leq 0, h_i(u^*) = 0$
- 对偶问题可行：满足对偶问题的约束 $\alpha_i^* \geq 0$
- 主问题最优：$\frac{\partial L}{\partial u^*} = 0$
- 互补松弛：$\alpha_i^* g_i(u^*) = 0$

KKT 条件一个重要用途是基于其中的**主变量最优**和**互补松弛**得到主问题的最优解 u^*

2.2 硬间隔 SVM 的对偶型

回顾一下基本型：

$$\min_{w,b} \frac{1}{2} w^T w$$
$$s.t. y_i(w^T x_i + b) \geq 1, i = 1,2,...,m$$

- **把约束优化问题写成基本形式**

$$\min_{w,b} \frac{1}{2} w^T w$$
$$s.t. 1 - y_i(w^T x_i + b) \leq 0, i = 1,2,...,m$$

其中，优化变量 $u = \begin{bmatrix} w \\ b \end{bmatrix} \in \mathbb{R}^{d+1}$，包含 m 项不等式约束。

- **构建拉格朗日函数**

$$L(w,b,\alpha) := \frac{1}{2}w^T w + \sum_{i=1}^{m}\alpha_i(1-y_i(w^T x_i + b))$$

得到主问题的等价形式：

$$\min_{w,b}\max_{\alpha}\frac{1}{2}w^T w + \sum_{i=1}^{m}\alpha_i(1-y_i(w^T x_i + b))$$
$$s.t.\alpha_i \geq 0, i = 1,2,...,m$$

- **交换 min 和 max 的顺序得到对偶问题并求解**

得到对偶问题：

$$\max_{\alpha}\min_{w,b}\frac{1}{2}w^T w + \sum_{i=1}^{m}\alpha_i(1-y_i(w^T x_i + b))$$
$$s.t.\alpha_i \geq 0, i = 1,2,...,m$$

接下来求解该对偶问题，可以看到内层优化问题是关于 w,b 的无约束优化问题，我们可以通过令偏导为零的方式得到 w,b 的最优解。

首先计算 w 的最优值，令：

$$\frac{\partial L}{\partial w} = \frac{\partial}{\partial w}(\frac{1}{2}w^T w + \sum_{i=1}^{m}\alpha_i(1-y_i(w^T x_i + b)))$$
$$= w - \sum_{i=1}^{m}\alpha_i y_i x_i$$
$$= 0$$

得：$w^* = \sum_{i=1}^{m}\alpha_i y_i x_i$

接下来求解 b 的最优解，令

$$\frac{\partial L}{\partial b} = \frac{\partial}{\partial b}(\frac{1}{2}w^T w + \sum_{i=1}^{m}\alpha_i(1-y_i(w^T x_i + b)))$$
$$= -\sum_{i=1}^{m}\alpha_i y_i$$
$$= 0$$

得：$\sum_{i=1}^{m}\alpha_i y_i = 0$

可以看到，这里并不能直接得出 b^*，在下一步应用 KKT 条件的互补松弛可以计算得到。

将得到的两个等式代入对偶问题的内层表达式可以将对偶问题变为：

$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{m}\alpha_i$$
$$s.t. \alpha_i \geq 0, i = 1,2,...,m$$
$$\sum_{i=1}^{m}\alpha_i y_i = 0$$

该问题等价于：

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{m}\alpha_i$$
$$s.t. \alpha_i \geq 0, i = 1,2,...,m$$
$$\sum_{i=1}^{m}\alpha_i y_i = 0$$

可以证明，最终的对偶问题是一个凸二次优化问题。这里就不作证明了。

- **利用 KKT 条件得到主问题的最优解**

➢ 主问题可行：满足主问题的约束 $g_i(u^*) = 1 - y_i(w^{*T} x_i + b^*) \leq 0$

➢ 对偶问题可行：满足对偶问题的约束 $\alpha_i^* \geq 0$

➢ 主问题最优：$w^* = \sum_{i=1}^{m}\alpha_i y_i x_i$ ， $\sum_{i=1}^{m}\alpha_i y_i = 0$

➢ 互补松弛：$\alpha_i^* g_i(u^*) = \alpha_i^*(1 - y_i(w^{*T} x_i + b^*)) = 0$

由对偶问题可行条件 $\alpha_i^* \geq 0$，我们可以根据 α_i^* 的取值去划分训练集中的所有样本。

当 $\alpha_i^* > 0$ 时，由互补松弛条件有 $1 - y_i(w^{*T} x_i + b^*) = 0$，即 $y_i(w^{*T} x_i + b^*) = 1$，该样本是距离超平面最近的样本，位于最大间隔的边界上。该样本称为支持向量。

当 $\alpha_i^* = 0$ 时，有 $y_i(w^{*T} x_i + b^*) \geq 1$，则该样本不一定是距离超平面最近的样本。

令

$$SV := \{i \mid \alpha_i > 0, i = 1,2,...,m\}$$

表示所有支持向量编号的集合，则有：

$$w^* = \sum_{i=1}^{m}\alpha_i^* y_i x_i = \sum_{i \notin SV} 0 y_i x_i + \sum_{i \in SV}\alpha_i^* y_i x_i = \sum_{i \in SV}\alpha_i^* y_i x_i$$

也就是说，w^* 仅由支持向量决定。

对于 b^* 的取值，对于某一支持向量 x_s 及其标签 y_s。由于 $y_s(w^{*T} x_s + b^*) = 1$ 得：

$$b^* = \frac{1}{y_s} - w^{*T} x_s = y_s - w^{*T} x_s = y_s - \sum_{i \in SV}\alpha_i^* y_i x_i^T x_s$$

理论上只需要一个支持向量就可以表示出b^*，但实际上通常用所有支持向量求解之后取平均值。

因此，在求解参数w^*, b^*之后，就可以把硬间隔 SVM 的假设函数表示为：

$$h(x) := sign(w^{*T} x + b^*)$$
$$= sign(\sum_{i \in SV} \alpha_i^* y_i x_i^T x + b^*)$$

3. 硬间隔核化 SVM

3.1 核技巧与硬间隔核化 SVM 的对偶型

当训练集样本线性不可分是，SVM 可以通过加入核方法（特征映射）解决非线性可分的问题。如果样本在原空间$\mathbb{R}^d$中线性不可分且维度d是有限的，可以证明，一定存在一个维度d'使得训练集在空间$\mathbb{R}^{d'}$中是线性可分的。

在通过特征映射将$x \in \mathbb{R}^d$映射为$\phi(x) \in \mathbb{R}^{d'}$之后，在$\mathbb{R}^{d'}$空间中，SVM 的参数$w$维度也要变为$d'$维，那么硬间隔 SVM 的基本型和对偶型可以分别表示为：
基本型：

$$\min_{w,b} \frac{1}{2} w^T w$$
$$s.t. y_i (w^T \phi(x_i) + b) \geq 1, i = 1, 2, ..., m$$

对偶型：

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \phi(x_i^T) \phi(x_j) - \sum_{i=1}^{m} \alpha_i$$
$$s.t. \alpha_i \geq 0, i = 1, 2, ..., m$$
$$\sum_{i=1}^{m} \alpha_i y_i = 0$$

如果将向量映射到$\mathbb{R}^{d'}$之后进行优化计算，计算复杂度将会大大提高（$O(d')$）。其中，对偶型可以使用核技巧进行优化，计算复杂度可以降为$O(d)$。在对偶型中，被映射到高维空间的向量在优化问题中总是以内积的形式存在，即$\phi(x_i^T)\phi(x_j)$，核技巧的本质就是把内积运算等价为在低维$\mathbb{R}^d$的一些运算，也就是说，我们希望构造一个核函数$\kappa(x_i, x_j) = \phi(x_i^T)\phi(x_j)$，使得$\kappa(x_i, x_j)$可以在$O(d)$的复杂度下进行计算。那么对偶型

可以写为以下形式，称为**硬间隔核化 SVM 的对偶型：**

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \kappa(x_i, x_j) - \sum_{i=1}^{m}\alpha_i$$

$$s.t. \alpha_i \geq 0, i = 1, 2, ..., m$$

$$\sum_{i=1}^{m}\alpha_i y_i = 0$$

3.2 核函数

常见的核函数有，线性核、多项式核和高斯核，他们的优点与缺点如图 4 所示：

名称	形式	优点	缺点
线性核	$\boldsymbol{x}_i^{\top}\boldsymbol{x}_j$	有高效实现，不易过拟合	无法解决非线性问题
多项式核	$(r\boldsymbol{x}_i^{\top}\boldsymbol{x}_j + c)^k$	比线性核更通用，k 直接描述了被映射空间的复杂度	超参数多，当 k 很大时会导致计算不稳定
高斯核	$\exp\left(-\gamma\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2\right)$	只有一个超参数 γ，没有多项式核计算不稳定的问题	计算慢，过拟合风险大

图 4：常见核函数及其优缺点

在选择完核函数之后，可以解得最优解

$$w^* = \sum_{i\in SV}\alpha_i^* y_i \phi(x_i) \qquad b^* = y_s - \sum_{i\in SV}\alpha_i^* y_i \kappa(x_i, x_s)$$

在求解对偶型得到 $\alpha_i^*(i = 1, 2, ..., m)$ 之后，硬间隔核化 SVM 的假设函数可以写为：

$$h(x) = sign(\sum_{i\in SV}\alpha_i^* y_i \kappa(x_i, x) + b^*)$$

4. 软间隔核化 SVM

在此之前我们已经了解并详细推导了硬间隔 SVM 基本型、对偶型与核化之后的版本，接下来推导一下软间隔 SVM。首先从硬间隔 SVM 的思想出发，硬间隔 SVM 的思想就是使得划分超平面尽可能与正负样本分开，即做到训练误差尽可能小。但是实际应用上，经常会有噪声样本存在，使用硬间隔的思想就可能会使得模型过拟合，如图 5 所示，如果使用硬间隔 SVM 模型进行拟合，很可能得到超平面 B，而事实上超平面 A 能够更具有泛化性。

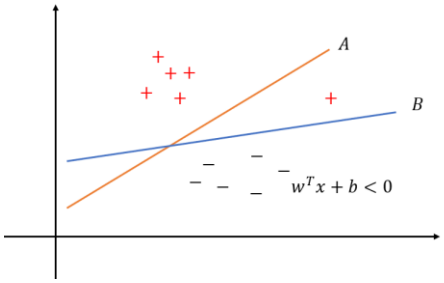

图 5：硬间隔 SVM 过拟合问题

4.1 软间隔核化 SVM 的基本型

软间隔 SVM 的思想就是将硬间隔 SVM 的严格约束放宽松一点，我们从硬间隔核化 SVM 讲起，过度到软间隔核化 SVM（因为核化不核化也就相差核函数而已）。可以证明，硬间隔核化 SVM 会严格使所有训练样本训练正确，回顾其基本型：

$$\min_{w,b} \frac{1}{2} w^T w$$
$$s.t. y_i(w^T \phi(x_i) + b) \geq 1, i = 1, 2, ..., m$$

约束中所有样本都需要满足 $y_i(w^T \phi(x_i) + b) \geq 1$，也就是让所有样本都分类正确。现在我们想要将该约束松弛一下，也就是允许被错误分类样本的出现，但这些样本尽可能少，优化问题变为如下问题：

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \mathrm{I}(y_i \neq sign(w^T \phi(x_i) + b))$$
$$s.t. y_i(w^T \phi(x_i) + b) \geq 1, if\ (y_i = sign(w^T \phi(x_i) + b))$$

其中，$\mathrm{I}(y_i \neq sign(w^T \phi(x_i) + b))$ 为指示函数，参数为一个条件，函数值为 1 当且仅当条件成立，否则函数值为 0。$\sum_{i=1}^{m} \mathrm{I}(y_i \neq sign(w^T \phi(x_i) + b))$ 的意思就是统计所有样本中预测错误的样本数量，并对预测错误的样本加上一个惩罚项，C 是一个超参数，表示惩罚的程度。

由于指示函数 $\mathrm{I}(y_i \neq sign(w^T \phi(x_i) + b))$ 不是连续函数，更不是凸函数，这使得问题的求解比较困难，我们需要简化这个优化问题。

我们引入一个松弛变量 $\xi_i \in \mathbb{R}$ 来描述第 i 个样本违背约束的程度，对于某个样本，违背约束的程度越大，ξ_i 的值就应该越大，那么有：

$$\xi_i := \begin{cases} 0, if\ (y_i(w^T \phi(x_i) + b) \geq 1) \\ 1 - y_i(w^T \phi(x_i) + b), otherwise \end{cases}$$

有了 ξ_i 的定义之后，我们对之前定义的软间隔优化问题进行改造，得到如下优化问题：

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i$$
$$s.t. y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, i = 1, 2, ..., m$$
$$\xi_i \geq 0, i = 1, 2, ..., m$$

该优化问题也称为**软间隔核化 SVM 的基本型。**

当样本 i 满足大间隔约束 $y_i(w^T \phi(x_i) + b) \geq 1$ 时， 有 $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i$ 对任意

$\xi_i \geq 0$ 成立，而优化目标的第二项是要最小化 ξ_i，故 $\xi_i = 0$

当样本 i 不满足大间隔约束 $y_i(w^T\phi(x_i)+b) \geq 1$ 时，有 $\xi_i \geq 1 - y_i(w^T\phi(x_i)+b) > 0$，由于要最小化 ξ_i，故 $\xi_i = 1 - y_i(w^T\phi(x_i)+b)$。

可见对于满足约束的样本与不满足约束的样本而言，基本型的定义与 ξ_i 的定义一致。

最后，可以像证明硬间隔 SVM 基本型一样类似地证明软间隔核化 SVM 的基本型所表示的优化问题属于凸二次规划问题，那么软间隔核化 SVM 同样具有全局最小值。

4.2 软间隔核化 SVM 的对偶型

类比 2.2 节推到硬间隔 SVM 对偶型的方法，我们分四个步骤来推导软间隔核化 SVM 的对偶型，首先回顾一下软间隔核化 SVM 的基本型。

$$\begin{aligned}
&\min_{w,b,\xi} \frac{1}{2} w^T w + C\sum_{i=1}^{m} \xi_i \\
&s.t. y_i(w^T\phi(x_i)+b) \geq 1 - \xi_i, i = 1,2,...,m \\
&\xi_i \geq 0, i = 1,2,...,m
\end{aligned}$$

- **把约束优化问题写成基本形式**

$$\begin{aligned}
&\min_{w,b,\xi} \frac{1}{2} w^T w + C\sum_{i=1}^{m} \xi_i \\
&s.t. 1 - \xi_i - y_i(w^T\phi(x_i)+b) \leq 0, i = 1,2,...,m \\
&-\xi_i \leq 0, i = 1,2,...,m
\end{aligned}$$

其中，优化变量 $u = \begin{bmatrix} w \\ b \\ \xi \end{bmatrix} \in \mathbb{R}^{m+d+1}$，包含 $2m$ 项不等式约束。

- **构建拉格朗日函数**

由于软间隔核化 SVM 优化问题存在两组不等式约束，故在拉格朗日函数中需要两组拉格朗日乘子 α, β，构建的拉格朗日函数如下：

$$L(w,b,\xi,\alpha,\beta) := \frac{1}{2} w^T w + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i(1 - \xi_i - y_i(w^T\phi(x_i)+b)) + \sum_{i=1}^{m} \beta_i(-\xi_i)$$

- **交换 min 和 max 的顺序得到对偶问题并求解**

得到对偶问题如下：

$$\begin{aligned}
&\max_{\alpha,\beta} \min_{w,b,\xi} L(w,b,\xi,\alpha,\beta) \\
&s.t. \alpha_i \geq 0, i = 1,2,...,m \\
&\beta_i \geq 0, i = 1,2,...,m
\end{aligned}$$

可以看到内层的优化问题为无约束的优化问题，可以直接通过令偏导等于 0 的方法求

解。首先求解 w 的最优解，令 $\frac{\partial L}{\partial w}=0$ 有：

$$\begin{aligned}\frac{\partial L}{\partial w}&=\frac{\partial}{\partial w}(\frac{1}{2}w^Tw+C\sum_{i=1}^{m}\xi_i+\sum_{i=1}^{m}\alpha_i(1-\xi_i-y_i(w^T\phi(x_i)+b))+\sum_{i=1}^{m}\beta_i(-\xi_i))\\&=w+\sum_{i=1}^{m}\alpha_i(-y_i\phi(x_i))\\&=w-\sum_{i=1}^{m}\alpha_iy_i\phi(x_i)=0\end{aligned}$$

故 w 的最优解为：

$$w^*=\sum_{i=1}^{m}\alpha_iy_i\phi(x_i)$$

然后计算 b 的最优解，令 $\frac{\partial L}{\partial b}=0$ 得：

$$\begin{aligned}\frac{\partial L}{\partial b}&=\frac{\partial}{\partial b}(\frac{1}{2}w^Tw+C\sum_{i=1}^{m}\xi_i+\sum_{i=1}^{m}\alpha_i(1-\xi_i-y_i(w^T\phi(x_i)+b))+\sum_{i=1}^{m}\beta_i(-\xi_i))\\&=-\sum_{i=1}^{m}\alpha_iy_i\\&=0\end{aligned}$$

故得：

$$\sum_{i=1}^{m}\alpha_iy_i=0$$

注意到这里与 2.2 节一样，并没有直接求解出 b 的最优解，而是得出一个等式约束，在下一节根据 KKT 条件可以得出 b 的最优解。

最后计算 ξ 的最优解，令 $\frac{\partial L}{\partial \xi}=0$ 得：

$$\begin{aligned}\frac{\partial L}{\partial \xi_i}&=\frac{\partial}{\partial \xi_i}(\frac{1}{2}w^Tw+C\sum_{i=1}^{m}\xi_i+\sum_{i=1}^{m}\alpha_i(1-\xi_i-y_i(w^T\phi(x_i)+b))+\sum_{i=1}^{m}\beta_i(-\xi_i))\\&=C-\alpha_i-\beta_i\\&=0\end{aligned}$$

故得：

$$\alpha_i+\beta_i=C$$

注意到这里同样没有直接得出 ξ 的最优解，而是得出一个等式约束。

把上述求解得到的结果代入拉格朗日函数中可以得到：

$$
\begin{aligned}
&L(w^*,b^*,\xi^*,\alpha,\beta)\\
&=\frac{1}{2}w^{*T}w^*+C\sum_{i=1}^{m}\xi_i^*+\sum_{i=1}^{m}\alpha_i(1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*))+\sum_{i=1}^{m}\beta_i(-\xi_i^*)\\
&=\frac{1}{2}w^{*T}w^*-\sum_{i=1}^{m}\alpha_i y_i w^{*T}\phi(x_i)-\sum_{i=1}^{m}\alpha_i y_i b^*+C\sum_{i=1}^{m}\xi_i^*-\sum_{i=1}^{m}\alpha_i\xi_i^*-\sum_{i=1}^{m}\beta_i\xi_i^*+\sum_{i=1}^{m}\alpha_i\\
&=\frac{1}{2}w^{*T}w^*-w^{*T}\sum_{i=1}^{m}\alpha_i y_i\phi(x_i)-b^*\sum_{i=1}^{m}\alpha_i y_i+\sum_{i=1}^{m}(C-\alpha_i-\beta_i)\xi_i^*+\sum_{i=1}^{m}\alpha_i\\
&=-\frac{1}{2}w^{*T}w^*+\sum_{i=1}^{m}\alpha_i\\
&=-\frac{1}{2}(\sum_{i=1}^{m}\alpha_i y_i\phi(x_i))^T(\sum_{j=1}^{m}\alpha_j y_j\phi(x_j))+\sum_{i=1}^{m}\alpha_i\\
&=-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\phi(x_i)^T\phi(x_j)+\sum_{i=1}^{m}\alpha_i\\
&=-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\kappa(x_i,x_j)+\sum_{i=1}^{m}\alpha_i
\end{aligned}
$$

故软间隔核化 SVM 的对偶问题最终变为：

$$
\begin{aligned}
&\max_{\alpha,\beta}-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\kappa(x_i,x_j)+\sum_{i=1}^{m}\alpha_i\\
&s.t.\alpha_i\geq 0,i=1,2,...,m\\
&\beta_i\geq 0,i=1,2,...,m\\
&\alpha_i+\beta_i=C,i=1,2,...,m\\
&\sum_{i=1}^{m}\alpha_i y_i=0
\end{aligned}
$$

其中的几个约束条件可以等价的变形为：

$$
\begin{aligned}
&\max_{\alpha}-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\kappa(x_i,x_j)+\sum_{i=1}^{m}\alpha_i\\
&s.t.0\leq\alpha_i\leq C,i=1,2,...,m\\
&\sum_{i=1}^{m}\alpha_i y_i=0
\end{aligned}
$$

可以证明，这个问题属于凸二次规划问题，包含 m 个优化变量，$2m+2$ 项不等式约束。因此，软间隔核化 SVM 的对偶型有全局最小值。

- **利用 KKT 条件得到主问题的最优解**

➢ 主问题可行： $1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*)\leq 0,-\xi_i^*\leq 0$

➢ 对偶问题可行：满足对偶问题的约束 $0\leq\alpha_i^*\leq C$

➢ 主问题最优： $w^*=\sum_{i=1}^{m}\alpha_i^* y_i\phi(x_i),\sum_{i=1}^{m}\alpha_i^* y_i=0,\alpha_i^*+\beta_i^*=C$

- 互补松弛：$\alpha_i^*(1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*))=0$，$\beta_i^*\xi_i^*=(C-\alpha_i^*)\xi_i^*=0$

由 KKT 条件中的对偶问题可行$0\le\alpha_i^*\le C$，我们可以根据α_i^*的取值去划分整个数据集，分为以下三类：

当$\alpha_i^*=0$时，由$(C-\alpha_i^*)\xi_i^*=0$可得，$\xi_i^*=0$，又由$\alpha_i^*(1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*))=0$得：$1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*)\le 0$，即$y_i(w^{*T}\phi(x_i)+b^*)\ge 1$，故此时样本位于最大间隔边界或之外，不是支持向量。

当$0<\alpha_i^*<C$时，由$(C-\alpha_i^*)\xi_i^*=0$可得，$\xi_i^*=0$，又由：

$\alpha_i^*(1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*))=0$可得：$1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*)=0$即：

$y_i(w^{*T}\phi(x_i)+b^*)=1$，故此时样本位于最大间隔边界之上，为支持向量。

当$\alpha_i^*=C$时，由$(C-\alpha_i^*)\xi_i^*=0$可得，$\xi_i^*\ge 0$，又由$\alpha_i^*(1-\xi_i^*-y_i(w^{*T}\phi(x_i)+b^*))=0$可得：$y_i(w^{*T}\phi(x_i)+b^*)=1-\xi_i^*$，故此时根据$\xi_i^*$的取值再细分为以下四类：

当$\xi_i^*=0$时，$y_i(w^{*T}\phi(x_i)+b^*)=1$，样本分类正确，且恰好位于最大间隔边界。

当$0<\xi_i^*<1$时，$y_i(w^{*T}\phi(x_i)+b^*)=1-\xi_i^*\in(0,1)$，样本分类正确，但不满足最大间隔约束，落在最大间隔内部。

当$\xi_i^*=1$时，$y_i(w^{*T}\phi(x_i)+b^*)=0$，此时样本恰好落在划分超平面上，样本分类错误，且不满足最大间隔约束。

当$\xi_i^*>1$时，$y_i(w^{*T}\phi(x_i)+b^*)<0$，此时样本分类错误，且不满足大间隔约束。

与 2.2 节类似，我们令

$$SV:=\{i\mid\alpha_i>0,i=1,2,...,m\}$$

表示所有支持向量编号的集合，则有：

$$w^*=\sum_{i=1}^{m}\alpha_i^*y_i\phi(x_i)=\sum_{i\notin SV}0y_i\phi(x_i)+\sum_{i\in SV}\alpha_i^*y_i\phi(x_i)=\sum_{i\in SV}\alpha_i^*y_i\phi(x_i)$$

也就是说，w^*仅由支持向量决定。

对于b^*的取值，对于某一支持向量x_s及其标签y_s。由于$y_s(w^{*T}x_s+b^*)=1$得：

$$b^*=\frac{1}{y_s}-w^{*T}\phi(x_s)=y_s-w^{*T}\phi(x_s)=y_s-\sum_{i\in SV}\alpha_i^*y_i\phi(x_i^T)\phi(x_s)=y_s-\sum_{i\in SV}\alpha_i^*y_i\kappa(x_i,x_s)$$

理论上只需要一个支持向量就可以表示出b^*，但实际上通常用所有支持向量求解之后取平均值。

因此，在求解参数w^*, b^*之后，就可以把软间隔核化 SVM 的假设函数表示为：

$$\begin{aligned} h(x) &:= sign(w^{*T}\phi(x) + b^*) \\ &= sign(\sum_{i \in SV} \alpha_i^* y_i \phi(x_i)^T \phi(x) + b^*) \\ &= sign(\sum_{i \in SV} \alpha_i^* y_i \kappa(x_i, x) + b^*) \end{aligned}$$

5. SVM 相关实验

5.1 SVM 优化问题的简单解决方法

前面提到的所有 SVM 的形式，最终都可以转化为一个凸二次规划问题，因此可以调用现成的二次规划软件库进行求解。以硬间隔 SVM 为例，回顾硬间隔 SVM 的对偶型：

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{m}\alpha_i \\ &s.t. \alpha_i \geq 0, i = 1, 2, ..., m \\ &\sum_{i=1}^{m}\alpha_i y_i = 0 \end{aligned}$$

可以转化为以下矩阵形式：

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\alpha^T Q \alpha - 1^T \alpha \\ &s.t. \alpha \geq 0, y^T \alpha = 0 \end{aligned}$$

其中，1是全 1 的列向量，$Q \in \mathbb{R}^{m \times m}$中的每一个元素为：

$$Q_{ij} := y_i y_j x_i^T x_j$$

在 matlab 当中，可以通过以下方式求解α：

$$\alpha = quadprog(Q, -\mathbf{1}, [], [], y^T, 0, \mathbf{0}, [], x_0)$$

5.2 SVM 在二维平面上的示例

首先使用 matlab 编程实现了硬间隔 SVM 与软间隔 SVM 优化模型，对如图 6 所示的二维线性可分数据进行分类，画出划分超平面并进行对比。可以看到，这个二维平面上的线性可分数据存在噪声点，如左上角的黑点，这个噪声点会对两个模型产生不一样的影响。

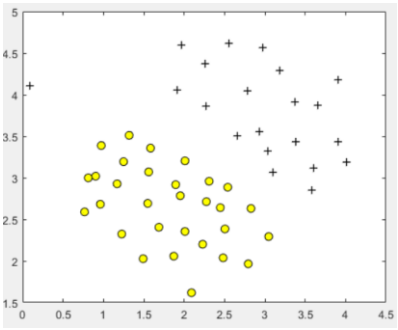

图 6：二维线性可分数据

使用硬间隔 SVM 模型训练出来的超平面如图 7 所示，可以看到，硬间隔 SVM 模型受噪声点的影响较大，由于其追求完美分类的效果，使得其训练出来的模型泛化性较差。

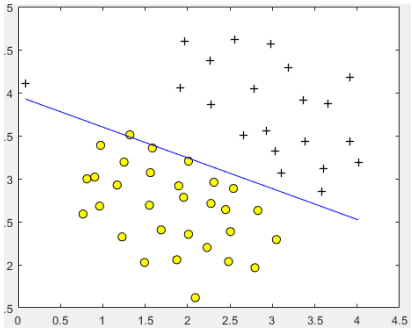

图 7：硬间隔 SVM 对二维线性可分数据的分类结果

使用软间隔 SVM 训练模型时，由于需要选择超参数C，这里我采用了留一交叉验证的方法进行超参数的选取，对于m个样本，每次以 1 个样本为测试集，以$m-1$个样本为训练集，进行m训练得到最终错误率，选择错误率最小的参数C。最终选出来的超参数$C=0.03$，训练出的划分超平面如图 8 所示，可以看出，软间隔 SVM 训练出来的模型淡化了噪声点的影响，使得模型泛化性更强。

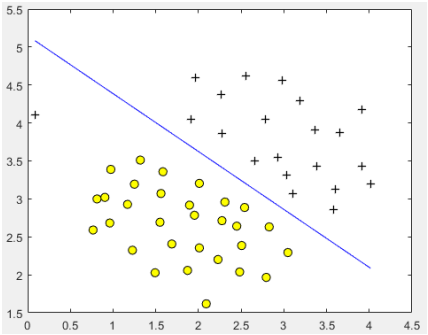

图 8：软间隔 SVM 对二维线性可分数据的分类结果

在这之后，我探究了在非线性数据上使用硬间隔核化 SVM 与软间隔核化 SVM 进行分类的效果，并进行比对，使用了高斯核函数进行训练。使用的数据如图 9 所示，为一个非线性可分数据。

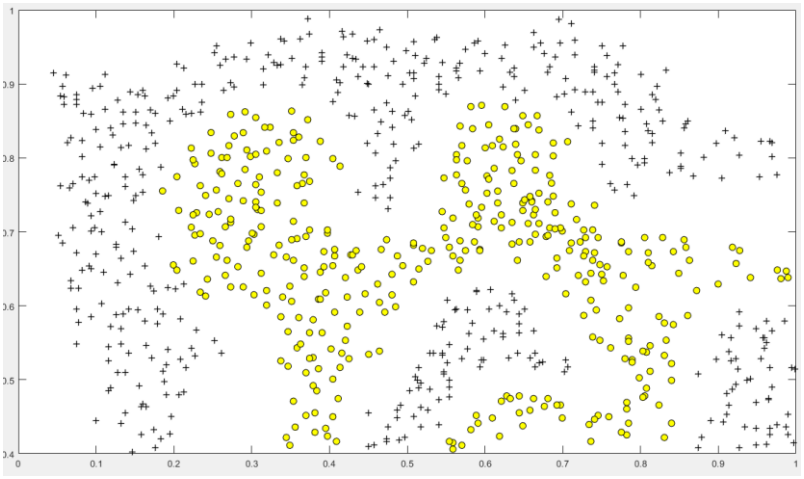

图 9：二维线性不可分数据

使用硬间隔核化 SVM 进行模型训练时，需要设置超参数σ，与之前选择超参数的方法一样，我使用了留一法进行模型性能的测试并选择错误率最低的模型对应的参数作为σ的取值，最终取值$\sigma=0.03$，最终训练得出决策边界如图 10 所示，可以看出，硬间隔核化 SVM 成功的把所有样本都分开了。很明显，该模型对训练样本过拟合了。

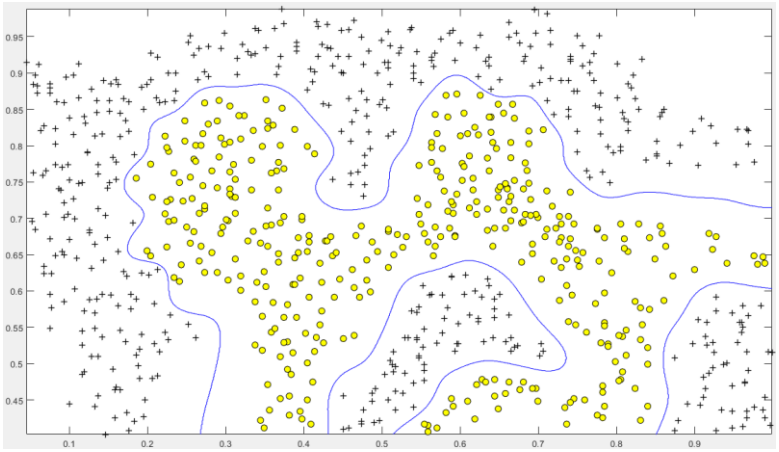

图 10：硬间隔核化 SVM 对二维线性不可分数据的分类结果

同样，使用软间隔核化 SVM 训练模型并画出决策边界，可以发现，该模型泛化性较强，没有过拟合训练样本。对于超参数的选择，仍用留一法进行选取，最终选取$C=\sigma=0.03$。

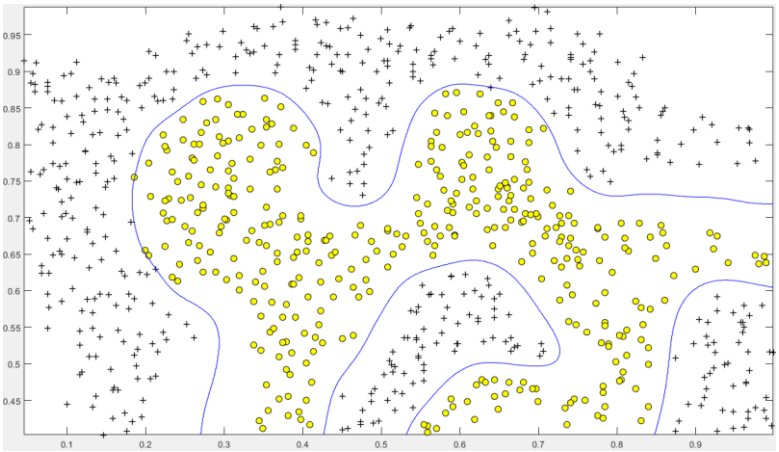

图 11：软间隔核化 SVM 对二维线性不可分数据的分类结果

5.3 使用 SVM 进行多分类

SVM 是一个二分类器，要使用 SVM 进行多分类有许多思路，这里使用其中的一种：OvO 拆解（One vs One，一对一）。假设现在类别数为 K，那么在训练时，OvO 拆解将这 K 个类别两两配对，对于每个配对，将其中一类作为正类、另一类作为负类，训练一个二分类模型。总共会产生 $C_K^2 = \frac{K(K-1)}{2}$ 个二分类模型。在预测阶段时，使用 $\frac{K(K-1)}{2}$ 个二分类模型同时对待预测样本进行预测，得到 $\frac{K(K-1)}{2}$ 个二分类预测结果，将被预测最多的类别作为最终的分类结果。

5.4 SVM 在人脸数据上的表现效果

使用了 PCA 与 LDA 进行降维(10~100)，比较使用 SVM 进行多分类（采用 OvO 拆解策略）与使用 KNN 进行多分类的识别率。对于数据集的选取，ORL 与 AR 数据集选取了 40 类人脸，Yale 数据集选取了 15 类人脸，训练集与测试集的划分比例为 7：3。实验结果如表 1、表 2、表 3、表 4 所示。由表 1 和表 2 可以看出，使用 PCA 降维之后的数据进行 SVM 分类的方法在人脸数据集上的表现非常出色，在数据维数较低时就已经可以达到较高的识别率。与 KNN 分类器相比，在不同维度下识别率都有较显著的提升。其中，以 AR 数据集上的提升最为显著。

表 1：PCA+KNN 在不同维度下对不同数据集下的识别率

	10	20	30	40	50	60	70	80	90	100
ORL	0.842	0.933	**0.942**	**0.942**	0.933	0.925	0.917	0.925	0.933	0.933
AR	0.300	0.419	0.459	0.503	0.522	0.522	0.538	0.544	0.538	**0.547**
Yale	0.950	0.983	0.967	0.967	0.967	**0.983**	0.967	0.967	**0.983**	**0.983**
FERET	0.413	0.400	0.438	0.438	**0.450**	0.438	0.438	**0.450**	**0.450**	**0.450**

表 2：PCA+SVM 在不同维度下对不同数据集下的识别率

	10	20	30	40	50	60	70	80	90	100
ORL	0.925	0.942	**0.958**	**0.958**	**0.958**	**0.958**	**0.958**	**0.958**	**0.958**	**0.958**
AR	0.560	0.800	0.875	0.878	0.894	0.906	**0.913**	0.906	0.906	**0.913**
Yale	0.950	0.983	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
FERET	0.450	0.500	0.513	0.538	0.538	**0.550**	**0.550**	**0.550**	**0.550**	**0.550**

由表 3 与表 4 可以看出 LDA 与 PCA 一样，使用了 SVM 进行分类之后，识别率也有所提升。

表 3：LDA+KNN 在不同维度下对不同数据集的识别率

	1	5	9	13	17	21	25	29	33	37
ORL	0.033	0.175	0.300	0.358	0.441	0.508	0.550	0.575	0.616	**0.641**
AR	0.022	0.097	0.166	0.231	0.347	0.397	0.431	0.500	0.522	**0.534**

表 4：LDA+SVM 在不同维度下对不同数据集的识别率

	1	5	9	13	17	21	25	29	33	37
ORL	0.153	0.185	0.324	0.438	0.464	0.548	0.572	0.645	0.756	**0.841**
AR	0.132	0.167	0.175	0.254	0.384	0.451	0.524	0.614	0.624	**0.735**

四、实验结论或体会

这次实验中，我详细推导了 SVM 的各种形式，并对相关数学原理进行证明解释，使得我对 SVM 的整个原理基本理解透彻。做了许多相关的实验，实验结果与理论一致，使得我对于 SVM 的理解更加深刻。

指导教师批阅意见： 成绩评定： 指导教师签字： 年　　月　　日
备注：

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。