

# Mathematical Model of Mobile Phone Sales Forecast

Mengmeng Tong

Liaoning Normal University, Dalian Liaoning  
Email: 1078862668@qq.com

Received: Jun. 28<sup>th</sup>, 2020; accepted: Jul. 9<sup>th</sup>, 2020; published: Jul. 16<sup>th</sup>, 2020

---

## Abstract

Forecasting the sales volume and making a good sales volume forecast are important bases for manufacturers to make a good production plan. In this paper, the sales volume and the order volume prediction of the latest generation of mobile phone products are studied. On the basis of the data given in the topic, the author selects the appropriate scientific principles and forecasting methods, to process and mine the data, and establish a forecasting model, and then processes and analyzes the model, studies and makes a forecasting model of mobile phone sales, so as to provide convenience for the sales of merchants. This paper focuses on the relevant knowledge of statistical learning theory for statistical estimation and prediction. Because the practical part of statistical learning theory is support vector machine, which can effectively realize the design method of minimizing structural risk, this paper adopts the method of least square support vector machine. Based on machine learning theory and statistical theory, a complete modeling work is carried out by using support vector machine. The prediction results show that the model has high prediction accuracy, and based on various conditions such as influencing factors of mobile phone sales, it realizes the systematic prediction of mobile phone sales, which has good development and can provide convenient conditions for electronic enterprise manufacturers.

## Keywords

Machine Learning Theory, Least Squares Support Vector Machine, Data Preprocessing

---

# 手机销售量预测数学模型

佟萌萌

辽宁师范大学, 辽宁 大连  
Email: 1078862668@qq.com

收稿日期: 2020年6月28日; 录用日期: 2020年7月9日; 发布日期: 2020年7月16日

## 摘要

预测销售量,做好销量预测是生产商制定好的生产方案的重要依据。本文针对最新一代手机产品的销售量和订单量预测的问题进行研究。以题目所给的数据为基础,选择适合的科学原理和预测方法,对数据进行处理、挖掘,建立预测模型,然后再对模型处理和分析,研究并做出手机销量的预测模型,为商家的销售提供便利条件。本文重点选用统计学习理论的相关知识进行统计估计和预测。由于统计学习理论中较实用的部分为支持向量机,是一种有效实现结构风险最小化的设计方法,所以本文采用最小二乘支持向量机的方法。以机器学习理论与统计理论为基础,运用支持向量机的方法进行了完整的建模工作。预测结果表明,该模型具有较高的预测精度,同时基于手机销售的影响因素等多方面条件,实现了系统地对手机销售预测,具有较好的发展性,可以为电子企业生产商提供方便条件。

## 关键词

机器学习理论,最小二乘支持向量机,数据预处理

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着社会科技、经济不断的发展,人们的生活水平也提高了,手机的使用已经变为“生活必需品”,手机销售是近几年电子产品中占销售量比重较高的种类之一,随着消费者需求的提高,如何把握不同的变化,合理地预测不同手机的功能、使用性能、价位及市场不同进货渠道和销售方式已经成为手机销售行业的一个热点问题,预测手机销售量问题是企业管理人员根据以往的销售量情况结合市场需求的分析对其手机产品销售量进行精确的预算,用来计划企业的经营方案和产销计划,通过手机销售量的预测可以精确地制定销售计划、选择进货渠道,从而使企业取得较好的效益。

本文针对最新一代手机产品的销售量和订单量预测的问题进行研究。以题目所给的数据为基础,选择适合的科学原理和预测方法,对数据进行处理、挖掘,建立预测模型,对模型进行系统的处理和分析,建立精准的手机销量的预测模型,为商家和企业管理者销售提供指导。

## 2. 背景介绍

电子产品一直存在着产品的生产量与销售量不能及时匹配的问题,产量过剩会导致产品积压;产量不足会影响收入。预测产品的销量并依此生产,是电子企业关心的问题。一般生产商的生产量以订单和销售量的预测为依据。本文需要解决的问题有:

问题 1: 建立销量预测模型,依据附件中给出的数据,预测销售商在未来 1 至 20 周出售给顾客最新的一代产品的销售量,即 a 类产品有三代,但只需预测 a-3 的销量,即第三代; b 类产品有两代,但只需要预测 b-2 的销售量; c 类产品只有一代,即 c-1。

问题 2: 给出订单预测模型,并根据附件给出的数据,对最新一代产品预测不同销售区域未来 1 至 20 周的订单量。订单分三类: 中间商与销售商之间的订单(A 类),生产商与销售商之间的订单(B 类),生产商与中间商之间的订单(C 类)。本题要求需要预测销售商的订单量(A + B 类订单量)和生产商的订单量

(B + C 类订单)。

### 3. 模型假设与符号说明

#### 3.1. 模型假设

- 1) 假设附件中所给的数据完全真实可靠。
- 2) 假设在已知数据时间段和预测时间段内，市场没有发生大的政策改变、销售模式改变。
- 3) 假设所给的电子产品没有出现较大的质量问题导致大规模产品召回。
- 4) 忽略电子产品退回后对订单的影响。

#### 3.2. 符号说明

在整篇文章中，我们主要采用了以下符号见表 1：

**Table 1.** Symbol description

**表 1.** 符号说明

符号	说明
$\ w\ ^2$	VC 维数
$x_i$	销售量影响因素
$y_i$	手机销售量
$a_i$	Lagrange 乘子
$L$	Lagrange 函数
$M$	训练样本集
$N$	测试样本集
$\Sigma$	求和符号

### 4. 模型建立与求解

#### 4.1. 问题一分析

本问题要求根据已给的数据，预测未来 20 周新一代产品的销售量，即 a-3、b-2、c 产品销量进行预测。一般从长期看来手机销量是有一定规律的；短期看手机销售量是受到消费者需求影响的；对销售量进行预测方法有很多种，本文运用最小二乘支持向量机，利用已知数据和最小二乘支持向量机算法证明，回归问题可以通过两类分类方法来解决，并通过数值试验，验证了最小二乘支持向量机分类算法解决回归问题的精准性。

#### 4.2. 模型建立与输入输出量选取

##### 4.2.1. 建立网络模型

考虑到 SVM 函数有比较高端的精度逼近能力，因此可以借助文献[1]中的 SVM 网络拓扑结构利用样本数据来训练，利用此方法探究内部的时间序列变化的规律，及逼近时间序列样本中隐含的函数关系，于是就可以完成对时间序列的映射关系，因此达到建立模型的目的。

##### 4.2.2. 选择输入量和输出量

依据变量和变量之间已有的关系，借助回归方法进行处理和分析，对自变量和因变量之间存在的关

系进行分析，最后得到了回归方程。实际上因变量和自变量实际上是非随机变量和随机变量。如果自变量的数据是已经知道的，我们就可以利用回归方程求出因变量的值。也就是，通过分析已知数据的变化规律，得到回归方程的关系式，然后确定模型参数，最后我们就可以进行预测了。

根据已知数据的分析，销售量的主要影响因素为生厂商收到的订单量、销售商下单量、顾客退给销售商的数量，将这几个影响因素作为输入因子进行销售量模型的建模。

在进行销售量建模时，将 2015 年 8 月 3 日至 2017 年 2 月 6 日的的数据作为支持向量机模型的训练样本，并将 2016 年 2 月的以后 20 周的销售量数据作为对基于最小二乘的支持向量机训练的检验样本。销售商下单量、厂商收到的订单量和顾客退给销售商的数据作为输入因子，销售量作为输出因子。对已给数据分析和处理后我们可以发现，输入因子和输出因子的量纲和数据的相对取值范围也是不同。要是直接作为输入因子进行 SVM 训练和样本检验，这样就会降低模型的精度。所以，我们需要处理对输入的数据进行出来，即对三个输入因子和一个输出因子进行归一化处理，处理后的数据的范围相对比较集中，有利于模型的准确预测。

### 4.2.3. 确定销售量预测结构模型

根据数据，我们发现手机销售量有非线性和不确定性，所以，我们要想建立一个确定的关系来描述手机最终销售量和影响销售量得各种因素是很难的。但利用手机销售量时间序列来综合表达手机销售量与影响销售量因素的结果，此外可分析得出手机销售系统发展情况，因此，通过分析手机销售量的时间序列特征，我们仔细分析和探索手机销售量系统变化的规律，对手机销售量进行预测。

本文在统计学习理论上，建立 SVM 的量预测模型，结合以上内容及数据分析，得到图 1 所示的预测模型，其中  $x_i, K(x_i, x), y_i$  为表 1 中符号[1]。

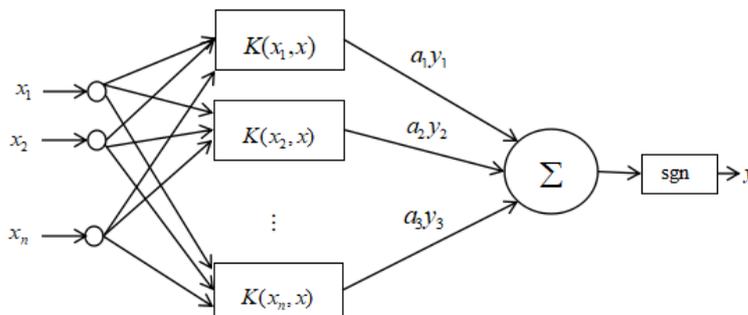


Figure 1. Support vector machine model  
图 1. 支持向量机模型

### 4.2.4. 时间序列模型的定阶

如图 1 所示是支持向量机预测模型的拓扑结构。因为输入维数是  $n$ ，所以模型的阶数为  $n$ 。因为支持向量机的任意逼近的非线性映射的能力，并且可以自动最优化生成对于上述映射确定的预测模型网络结构。所以，我们可以采用合适的模型并选择准则优化选取支持向量机的输入节点数，最后得到合适的预测模型结构。

### 4.2.5. 建立预测模型

最小二乘支持向量机(LSSVM)是支持向量机其中的一种算法，将标准支持向量机算法中的不等式约束转化为等式约束。对以上的非线性回归问题，设训练样本为

$$(x_1, y_1), \dots, (x_l, y_l) \in R^n \times R$$

非线性回归函数为

$$f(x) = w^T \varphi(x) + b$$

对于最小二乘支持向量机，优化问题变为

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + r \frac{1}{2} \left( \sum_{i=1}^l \xi_i^2 \right) \\ \text{st} & \begin{cases} y_i = w^T \varphi(x_i) + b + \xi_i \\ \xi_i \geq 0 \\ i = 1, \dots, l \end{cases} \end{aligned} \quad (1)$$

求解式(1)的优化问题，可以引入 Lagrange 函数

$$L = \frac{1}{2} \|w\|^2 + r \frac{1}{2} \left( \sum_{i=1}^l \xi_i^2 \right) - \sum_{i=1}^l a_i (w^T \varphi(x_i) + b + \xi_i - y_i)$$

其中， $a_i$  为 Lagrange 乘子；常数  $r > 0$ ，它对超出误差的样本的惩罚程度进行控制，最优的  $a_i$  和  $b$  可以根据 KKT (Karush-Kuhn-Tucker) 条件得到

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^l a_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^l a_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow a_i = r \xi_i \\ \frac{\partial L}{\partial \xi_i^*} = 0 \rightarrow w^T \varphi(x_i) + b + \xi_i = 0 \end{cases}$$

根据上式，上述问题即可转化为下面的求解如下的线性方程

$$\begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & K(X_1, X_2) + 1/r & K(X_1, X_2) & \dots & K(X_1, X_n) \\ 1 & K(X_2, X_1) & K(X_2, X_2) + 1/r & \dots & K(X_2, X_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K(X_n, X_1) & K(X_n, X_2) & \dots & K(X_n, X_n) + 1/r \end{bmatrix} \begin{bmatrix} b \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

上式中， $K(x_i, x_j)$  为核函数。

从而得到非线性回归函数的解为

$$f(x) = w^T \varphi(x) + b = \sum_{i=1}^l a_i K(x_i, x) + b \quad (2)$$

销售量及其影响因素的样本资料序列为  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ,  $x_i \in R^d, y_i \in R$ ，其中  $y_i$  为销售量， $x_i$  为销售量的影响因素。因此销售量的支持向量机模型表示如下：

$$f(x) = \sum_{i=1}^l a_i k(x_i, x) + b \quad (3)$$

本文主要介绍运用最小二乘支持向量机的方法：建立预测模型，根据手机销售量特点，找出影响手机销售量主要原因，运用文献[2] [3]中最小二乘支持向量机算法，建立手机销售量预测模型。

本文中基于 SVM 预测方法的手机销售量预测主要步骤为:

- 1) 对已知数据进行分析 and 预处理处理(特征提取、归一化);
- 2) 形成训练样本集和测试样本集;
- 3) 针对训练样本建立目标函数;
- 4) 解出目标函数;
- 5) 将求解式所得的参数代入回归函数  $f(x)$ , 将样本对未来 20 周手机销售量进行预测;
- 6) 依据 SVM 手机销售模型建立手机销售量预测的系统。

基于上述步骤的基础上建立手机销售量预测模型逻辑框图, 如图 2 所示。

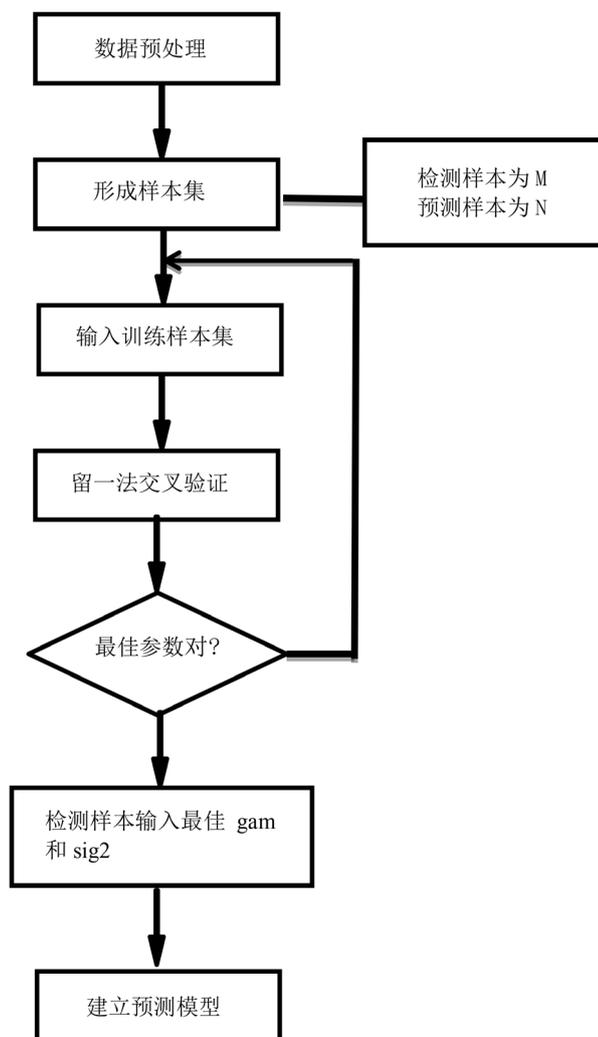


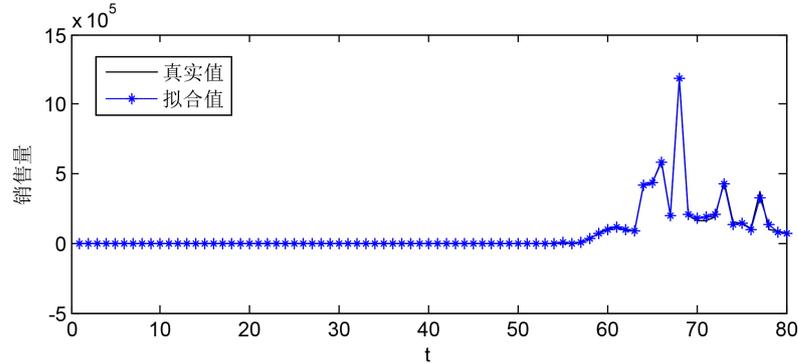
Figure 2. Modeling process of predicting mobile phone sales  
图 2. 预测手机销售量的建模过程

#### 4.2.6. 数据处理

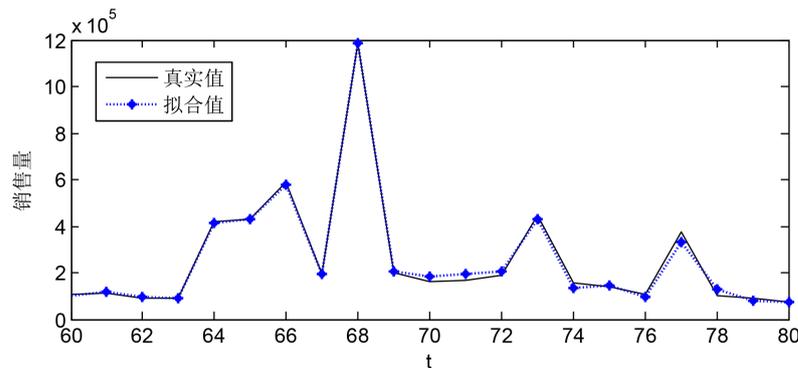
对已给当中的数据进行分析 and 处理, 即平稳化处理一些存在异常的数据, 并及时补充部分缺失的数据。一般情况, 经过不断的分析、研究时间数据, 合理的补充缺失的。对于问题 2 中数据, 我们分别将相同日期相同地区的 A 和 B 类、B 和 C 订单数量进行整合。

### 4.3. 问题一求解

a-3 销售量预测，以产品 a-3 的销售量为因变量，以生产商收到的订单、销售商的下单量、顾客退货量为自变量，建立预测模型，模型拟合值与预测值的拟合程度如图 3 和图 4 所示：

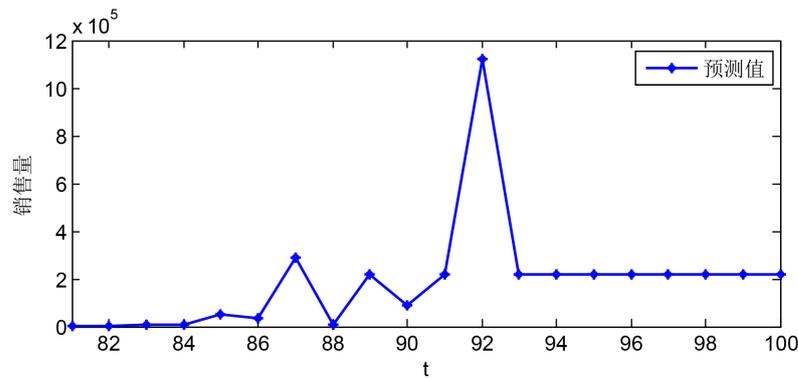


**Figure 3.** a-3 model fitting diagram of predicted data and actual data (1~80 weeks)  
**图 3.** a-3 预测数据与实际数据模型拟合图(1~80 周)



**Figure 4.** a-3 model fitting diagram of predicted data and actual data (60~80 weeks)  
**图 4.** a-3 预测数据与实际数据模型拟合图(60~80 周)

以此模型为基础，通过逐步预测与检验，得到如上两个模型拟合图 5，通过模型的拟合程度，得出模型建立的适用性与有效性。



**Figure 5.** a-3 predicted data predicted value in the next 20 weeks  
**图 5.** a-3 预测数据未来 20 周预测值

上图所示为 a-3 在未来 20 周内销售量的变化情况。如图可知，在未来 20 周内，该产品除了个别周次(第 12 周)有较大的销售量，其余几周的销售量较小，在第 7 周和第 9 周销售量增加，并逐渐趋于稳定。

b-2 销售量预测，以产品 b-2 的销售量为因变量，以生产商收到的订单、销售商的下单量、顾客退货量为自变量，建立预测模型，模型拟合值与预测值的拟合程度如图 6 和图 7 所示。图 8 为 b-2 预测数据未来 20 周销售量的预测值。

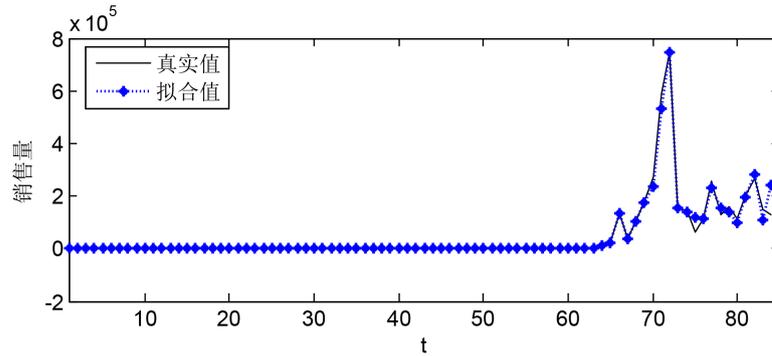


Figure 6. b-2 model fitting diagram of predicted data and actual data (0~80 weeks)  
图 6. b-2 预测数据与实际数据模型拟合图(0~80 周)

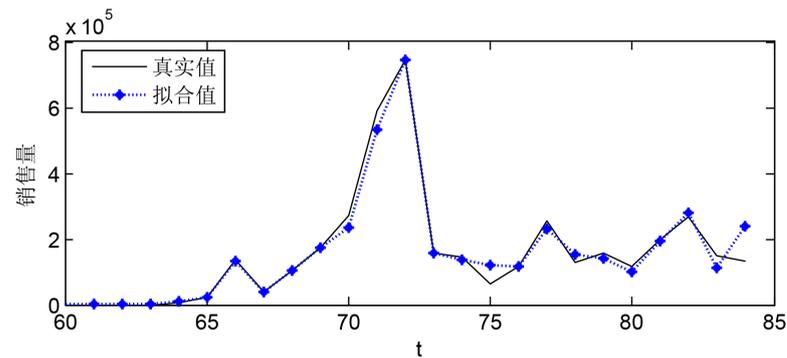


Figure 7. b-2 model fitting diagram of predicted data and actual data (60~85 weeks)  
图 7. b-2 预测数据与实际数据模型拟合图(60~85 周)

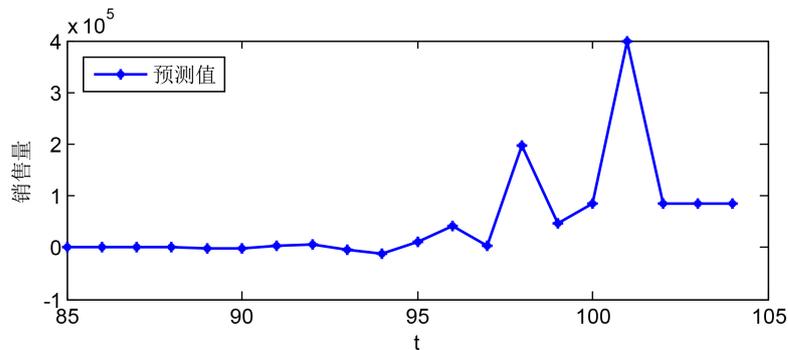


Figure 8. b-2 sales forecast for the next 20 weeks  
图 8. b-2 未来 20 周销售量预测值

上图所示为 b-2 在未来 20 周内销售量的变化情况。如图可知，该产品在前 12 周的销售量比较稳定，在第 12 周销售量开始增加，在第 14 周和 17 周的销售量较大，其余各周逐渐增加，并趋于稳定。

c-1 销售量预测, 以产品 c-1 的销售量为因变量, 以生产商收到的订单、销售商的下单量、顾客退货量为自变量, 建立预测模型, 预测数据与实际数据模型拟合图(0~50 周)如图 9 所示, 图 10 为 b-2 预测数据未来 20 周销售量的预测值。

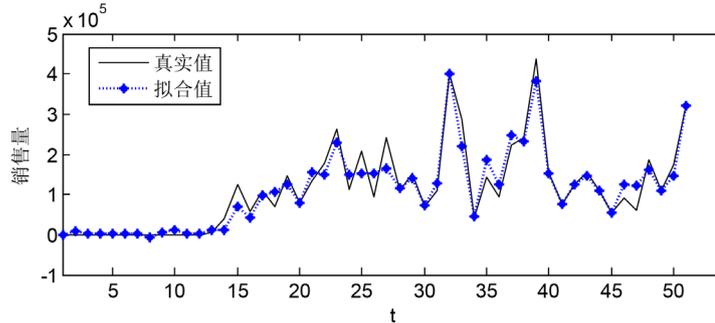


Figure 9. Model fitting diagram of c-1 predicted data and actual data (0~50 weeks)  
图 9. c-1 预测数据与实际数据模型拟合图(0~50 周)

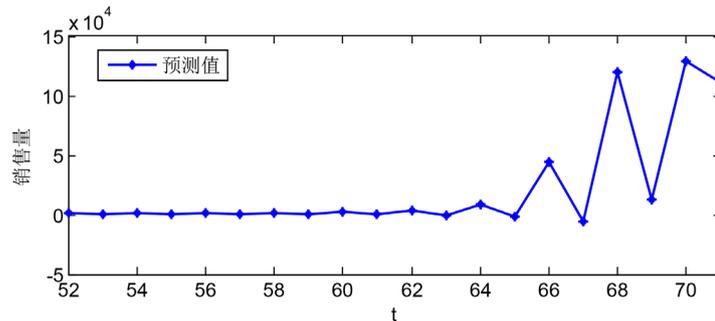


Figure 10. c-1 sales forecast for the next 20 weeks  
图 10. c-1 未来 20 周销售量预测值

上图所示为 c-1 在未来 20 周内销售量的变化情况。如图可知, 在未来 20 周, 该产品前 12 周的销售量基本稳定, 后 8 周的销售量波动较大。

#### 4.4. 问题二求解

产品 e-2 在 a 区 A + B 类订单预测, 按照时间序列以产品 e-2 的 A + B 类订单为输出变量, 建立预测模型, 模型拟合值与预测值的拟合程度如图 11 所示。图 12 为 e-2 未来 20 周订单量的预测值。

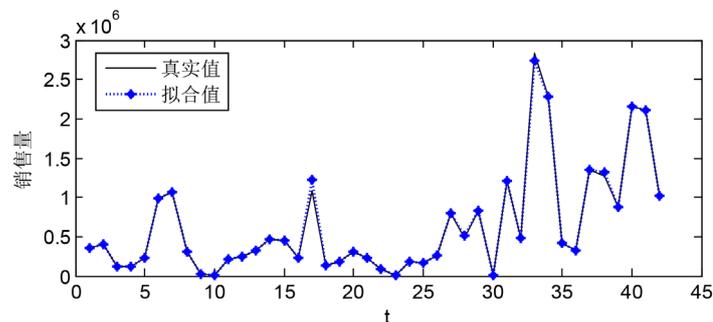
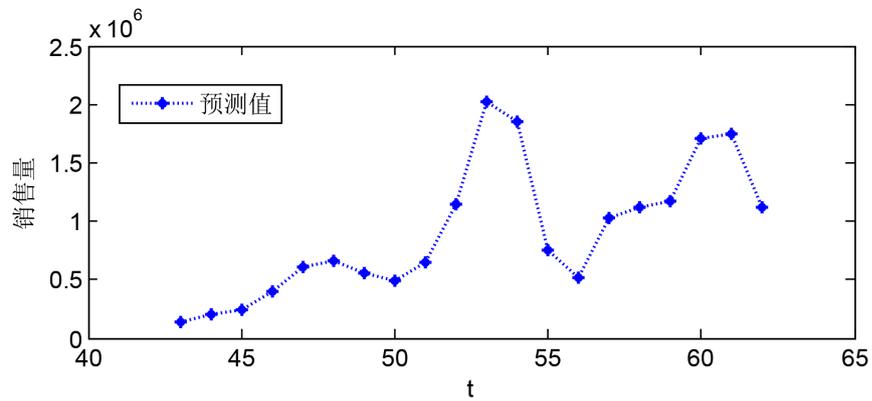
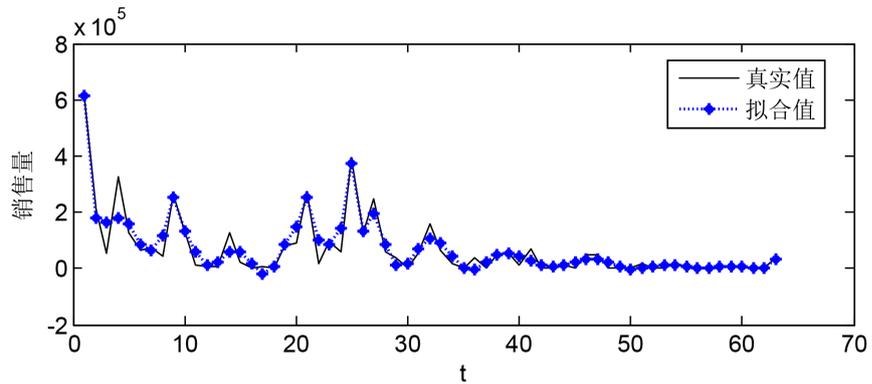


Figure 11. Model fitting diagram of predicted data and actual data (0~45 weeks)  
图 11. 预测数据与实际数据模型拟合图(0~45 周)

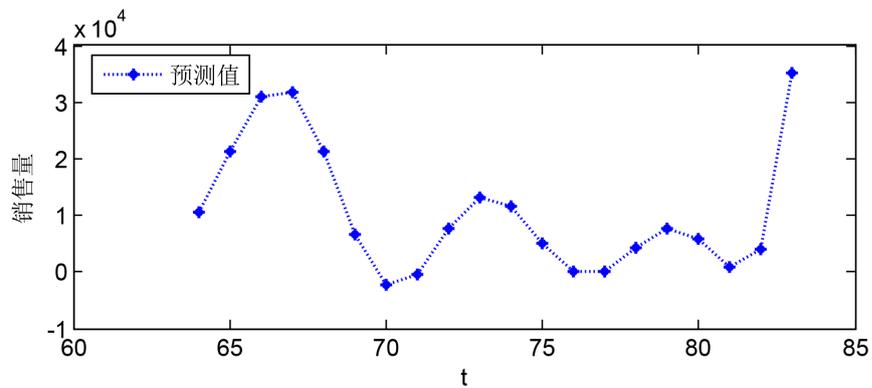


**Figure 12.** e-2 forecast value of Class A + B orders in Area A in the next 20 weeks  
**图 12.** e-2 未来 20 周在 a 区 A + B 类订单量预测值

2) 产品 d-3 在 a 区 B + C 类订单预测, 按照时间序列以产品 d-3 的 B + C 类订单为输出变量, 建立预测模型, 模型拟合值与预测值的拟合程度如图 13。图 14 为 d-3 未来 20 周销售量的预测值。



**Figure 13.** Model fitting diagram of predicted data and actual data (0~70 weeks)  
**图 13.** 预测数据与实际数据模型拟合图(0~70 周)



**Figure 14.** d-3 forecast value of B + C orders in area a in the next 20 weeks  
**图 14.** d-3 未来 20 周在 a 区 B + C 类订单量预测值

3) 产品在 a 区 A + B 类订单预测, 按照时间序列以产品 d-3 的 A + B 类订单为输出变量, 建立预测模型, 模型拟合值与预测值的拟合程度如图 15。图 16 为 d-3 预测数据未来 20 周销售量的预测值。

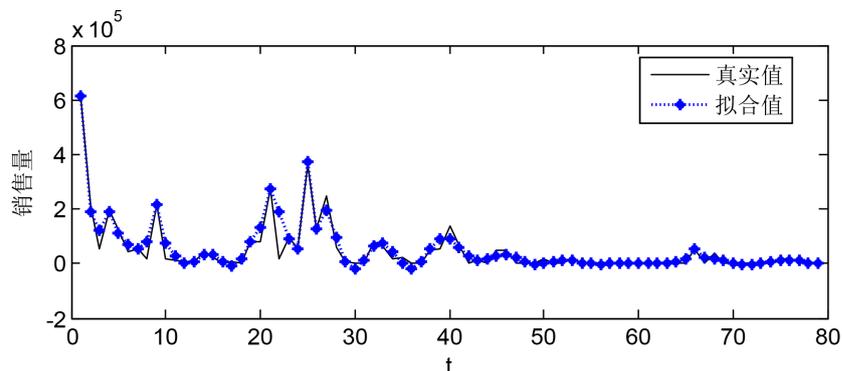


Figure 15. Model fitting diagram of predicted data and actual data (0~80 weeks)

图 15. 预测数据与实际数据模型拟合图(0~80 周)

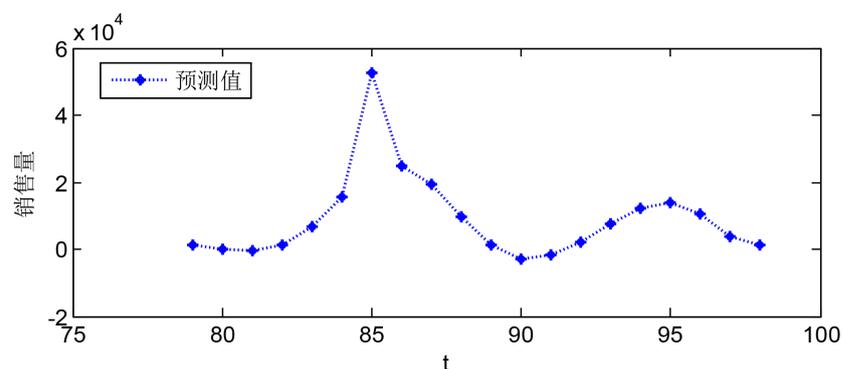


Figure 16. d-3 forecast value of Class A + B orders in Area A in the next 20 weeks

图 16. d-3 未来 20 周在 a 区 A + B 类订单量预测值

## 5. 模型评价

### 5.1. 模型优点

本文是在统计学习理论基础上进行统计估计和预测。由于统计学习理论是支持向量机中最实用的部分，能够较好达到结构风险最小化的设计思想，因此运用最小二乘支持向量机的方法。以机器学习理论与统计理论为基础，运用支持向量机的方法进行了完整的模型建立。预测结果表明，此模型有较强的准确性，同时基于手机销售的影响因素等多方面条件，实现了对手机销售的系统预测，具有比较好的发展前景，可以给企业管理、生产者提供参考，为销售量预测提供了一种新的办法。

### 5.2. 模型缺点

根据对附件中的数据处理分析得到，由于产品在销售区域的订单量数据不足，进行预测有效性较低。模型设计较简单，需要进一步加强。

## 参考文献

- [1] 胡俊. 基于最小二乘法支持向量机的小麦产量预测方法研究[D]: [硕士学位论文]. 西安: 西安科技大学, 2014.
- [2] 曹旭东. 数学建模原理与方法[M]. 北京: 高等教育出版社, 2014.
- [3] 刘科峰, 张韧, 于鹏, 王彦磊, 余丹丹. 基于小波分解和最小二乘支持向量机的西太平洋副高预测[J]. 热大气象学报, 2007, 12(5): 32-41.