

# 机器学习算法

感知机

我真的不懂忧郁



# 目录

<b>Nomenclature</b>	<b>iii</b>
<b>1 感知机 (Perceptron)</b>	<b>1</b>
1.1 超平面 . . . . .	1
1.2 模型 . . . . .	2
1.3 数据集的线性可分性 . . . . .	2
1.4 损失函数 . . . . .	2
1.5 感知机学习算法 . . . . .	3
1.6 感知机收敛定理 . . . . .	4
<b>2 感知机代码仿真</b>	<b>6</b>
2.1 算法描述 . . . . .	6
2.2 实战 . . . . .	7
<b>3 感知机的一些问题</b>	<b>10</b>
3.1 感知机不能实现异或 . . . . .	10
<b>A Source Code Example</b>	<b>11</b>
<b>B Latex Tikz Test Code</b>	<b>12</b>
<b>C Task Division Example</b>	<b>14</b>

# Nomenclature

*If a nomenclature is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.*

## Abbreviations

Abbreviation	Definition
ISA	International Standard Atmosphere
...	

## Symbols

Symbol	Definition	Unit
$V$	Velocity	[m/s]
...		
$\rho$	Density	[kg/m <sup>3</sup> ]
...		

# Chapter 1

## 感知机 (Perceptron)

### 1.1. 超平面

假设  $\mathbb{R}^n$  空间中的一个平面  $S$ ,  $S$  上的一个法向量为

$$\vec{n} = (\omega_1, \omega_2, \dots, \omega_n) \quad (1.1)$$

固定平面  $S$  的一个点  $P_0 = (x_0, x_1, \dots, x_n)$ , 对于  $S$  上任意一个点  $P = (x'_1, x'_2, \dots, x'_n)$ , 构成一个向量

$$\vec{x} = P\vec{P}_0 = (x'_1 - x_1, x'_2 - x_2, \dots, x'_n - x_n) \quad (1.2)$$

对于平面上任意一个向量, 都与法向量垂直, 因此平面方程表示为法向量和平面上任意一个向量的内积

$$\vec{n} \cdot \vec{x} = 0 \quad (1.3)$$

三维空间的平面称为平面, 如果是任意的维度, 广义的平面称为超平面。

#### 实例

对于平面上的点, 因为都与超平面垂直, 因此如果实例点落入超平面中,  $\vec{n} \cdot \vec{x} = 0$ 。对于位于超平面上方的点,  $PP_0$  和法向量  $n$  之间的夹角

$$\frac{PP_0 \cdot n}{\|PP_0\| \|n\|} \leq 90^\circ \quad (1.4)$$

因此在超平面上方的点都满足

$$PP_0 \cdot n \geq 0 \quad (1.5)$$

反之, 超平面下方的点满足

$$PP_0 \cdot n \leq 0 \quad (1.6)$$

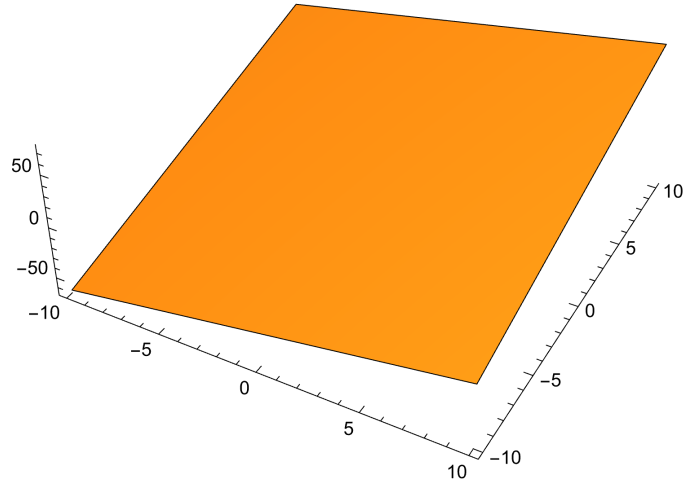


图 1.1: 超平面

## 1.2. 模型

**definition 1.2.1:** (感知机) 假设输入空间 (特征空间)  $\mathcal{X} \subset \mathbb{R}^n$ , 输出空间是  $\mathcal{Y} = \{-1, +1\}$ , 输入  $x \in \mathcal{X}$  表示实例的特征向量, 对应于输入空间 (特征空间) 的点; 输出  $y \in \mathcal{Y}$  表示实例的类别, 由输入空间到输出空间的如下符号函数

$$f(x) = \text{sign}(\omega \cdot x + b) \quad (1.7)$$

称为感知机。其中,  $\omega$  和  $b$  为感知机模型参数。

## 1.3. 数据集的线性可分性

假设一个数据集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (1.8)$$

如果存在一个超平面  $\omega \cdot x + b = 0$ , 使得数据集的正实例点和负实例点完全正确地划分到超平面的两侧处, 即对所有的  $y_i = +1$  的实例  $i$ , 有

$$\omega \cdot x_i + b > 0 \quad (1.9)$$

对所有的  $y_i = -1$  的实例  $i$ ,

$$\omega \cdot x_i + b < 0 \quad (1.10)$$

则  $T$  被称为线性可分的数据集。

## 1.4. 损失函数

### 实例点误分类的判别关系

根据输入空间到输出空间的映射方式, 误分类的实例点应该满足下面的关系。

$$-y_i(\omega \cdot x_i + b) > 0 \quad (1.11)$$

### 感知机的损失函数

损失函数一个自然选择是误分类点的总数，但是这样损失函数就不是参数  $\omega$  和  $b$ 。但是可以考虑：如果所有误分类点距离超平面的距离为 0，则这是一次完美的分割。

输入空间  $R^n$  中任意一点  $x$  到超平面  $S$  到距离为

$$\frac{1}{\|\omega\|} |\omega \cdot x + b| \quad (1.12)$$

对于误分类的数据来说，距离为

$$\frac{1}{\|\omega\|} y_i (\omega \cdot x + b) \quad (1.13)$$

1

所有误分类点到超平面  $S$  到总距离为

$$L = -\frac{1}{\|\omega\|} \sum_{x_i \in M}^n y_i (\omega \cdot x_i + b) \quad (1.14)$$

## 1.5. 感知机学习算法

给定一个训练数据集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \quad (1.15)$$

其中  $x_i \in \mathcal{X} = R^n$ ,  $y_i \in \mathcal{Y} = \{-1, 1\}$ ,  $i = 1, 2, \dots, N$ , 求参数  $\omega$ 、 $b$  使得损失函数最小化

$$\min_{\omega, b} L(\omega, b) = - \sum_{x_i \in M} y_i (\omega \cdot x_i + b) \quad (1.16)$$

其中  $M$  是误分类点集合。

### 原始形式

输入：训练数据集  $T$ ；设定学习率  $\eta$  ( $0 < \eta \leq 1$ )；

输出： $\omega, b; f(x) = \text{sign}(\omega \cdot x + b)$ ；

1. 选定初值  $\omega_0, b_0$ ；
2. 在训练集中选取数据  $(x_i, y_i)$ ；
3. 如果  $y_i(\omega \cdot x_i + b) \leq 0$ ，更新权重

$$\begin{aligned} \omega &\leftarrow \omega + \eta y_i x_i \\ b &\leftarrow b + \eta y_i \end{aligned} \quad (1.17)$$

4. 跳转至 2，直到没有误分类点。

<sup>1</sup> 因为  $y_i$  只能去 +1 和 -1，因此可以把绝对值符号去掉



## 对偶形式

输入：训练数据集  $T$ ；设定学习率  $\eta$  ( $0 < \eta \leq 1$ )；

输出： $\alpha, b; f(x) = \text{sign}(\sum_{j=1}^N \alpha_j y_j x_j \cdot x + b)$ ，其中  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$ ；

1.  $\alpha \leftarrow 0, b \leftarrow 0$ ；

2. 训练集上选取数据  $(x_i, y_i)$ ；

3. 如果  $y_i(\sum_{j=1}^N \alpha_j y_j x_j \cdot x_i + b) \leq 0$

$$\begin{aligned}\alpha_i &\leftarrow \alpha_i + \eta \\ b &\leftarrow b + \eta y_i\end{aligned}\tag{1.18}$$

4. 跳转到 2 直到没有误分类点数据。

对偶形式的基本想法是，将  $\omega$  和  $b$  表示为实例  $x_i$  和  $y_i$  的线性组合的形式，通过求解其系数而求得  $\omega$  和  $b$ 。不失一般性，在原始形式算法中假设初始值均为 0，对误分类点  $(x_i, y_i)$  通过

$$\begin{aligned}\omega &\leftarrow \omega + \eta y_i x_i \\ b &\leftarrow b + \eta y_i\end{aligned}\tag{1.19}$$

逐步修改  $\omega$  和  $b$ ，修改  $n$  次的  $\omega$  和  $b$  增量为  $\alpha_i y_i x_i$  和  $\alpha_i y_i$ 。这里  $\alpha_i = n_i \eta$ 。这样，从学习过程不难看出，最后学习到的  $\omega, b$  可以分别表示为

$$\omega = \sum_{i=1}^N \alpha_i y_i x_i\tag{1.20}$$

$$b = \sum_{i=1}^N \alpha_i y_i\tag{1.21}$$

这里  $\alpha_i \geq 1, i = 1, 2, \dots, N$ ，当  $\eta = 1$  时，表示第  $i$  个实例点由于误分类而进行更新的次数。实例点更新次数越多，意味着它距离超平面越近，也就越难正确分类。这样的实例对学习结果影响最大。

## 1.6. 感知机收敛定理

**theorem 1.6.1:** (Novikoff) 设训练数据集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$  是线性可分的数据集，其中  $x_i \in \mathcal{X} = R^n, y_i \in \mathcal{Y} = -1, +1, i = 1, 2, \dots, N$ ，则

1. 存在满足条件  $\|\hat{\omega}_{opt}\| = 1$  的超平面  $\hat{\omega}_{opt} \cdot \hat{x} = \omega_{opt} \cdot x + b_{opt} = 0$  将训练数据集完全正确分开；且存在  $\gamma > 0$ ，对于所有  $i = 1, 2, \dots, N$

$$y_i(\hat{\omega}_{opt} \cdot \hat{x}) = y_i(\omega_{opt} \cdot x_i + b_{opt}) \geq \gamma\tag{1.22}$$

2. 令  $R = \max_{1 \leq i \leq N} \|\hat{x}_i\|$ ，则感知机算法在训练数据集上的误分类次数  $k$  满足不等式

$$k \leq \left(\frac{R}{\gamma}\right)^2\tag{1.23}$$

**proof.** 首先证明第一点：

由于训练集数据是线性可分的，按照定义，存在超平面可以将数据集完全正确分开，取此超平面为

$$\hat{\omega}_{opt} \cdot \hat{x} = \omega_{opt} \cdot x + b_{opt} = 0 \quad (1.24)$$

由于等式右边等于 0，所以总可以取  $\|\hat{\omega}\| = 1$ 。对于有限的  $i = 1, 2, \dots, N$ ，正确分类下有

$$y_i(\hat{\omega}_{opt} \cdot \hat{x}_i) = y_i(\omega_{opt} \cdot x_i + b_{opt}) > 0 \quad (1.25)$$

所以 0 是序列  $\{y_i(\hat{\omega}_{opt} \cdot \hat{x}_i)\}$  的一个下界，假设序列的下确界

$$\gamma = \inf_i \{y_i(\omega_{opt} \cdot x_i + b_{opt})\} \quad (1.26)$$

使得

$$y_i(\hat{\omega}_{opt} \cdot \hat{x}_i) = y_i(\omega_{opt} \cdot x_i + b_{opt}) \geq \gamma \geq 0 \quad (1.27)$$

下面证明第二点：

训练点为  $x_i$ ，标签为  $y_i$ 。假设  $\hat{\omega}_{opt} = 0$  开始，如果实例被误分类，则更新权重。令  $\hat{\omega}_{k-1}$  是第  $k$  个误分类实例之前的扩充权重向量，即

$$\hat{\omega}_{k-1} = (\omega_{k-1}^T, b_{k-1})^T \quad (1.28)$$

如果第  $k$  个实例点是误分类点，则

$$y_i(\hat{\omega}_{k-1} \cdot \hat{x}_i) = y_i(\omega_{k-1} \cdot x_i + b_{k-1}) \leq 0 \quad (1.29)$$

在误分类的时候更新权重

$$\begin{aligned} \omega_k &\leftarrow \omega_{k-1} + \eta y_i x_i \\ b_k &\leftarrow b_{k-1} + \eta y_i \end{aligned} \quad (1.30)$$

即

$$\hat{\omega}_k = \hat{\omega}_{k-1} + \eta y_i \hat{x}_i \quad (1.31)$$

利用两个不等式，这两个不等式将在后续的引理中证明。

$$\hat{\omega}_k \cdot \hat{\omega}_{opt} \geq k\eta\gamma \quad (1.32)$$

$$\|\hat{\omega}_k\|^2 \leq k\eta^2 R^2 \quad (1.33)$$

代入方程 (1.31)，得到

$$\begin{aligned} k\eta\gamma &\leq \hat{\omega}_k \cdot \hat{\omega}_{opt} \leq \|\hat{\omega}_k\| \|\hat{\omega}_{opt}\| \leq \sqrt{k}\eta R \\ k^2\gamma^2 &\leq kR^2 \end{aligned} \quad (1.34)$$

于是

$$k \leq \left(\frac{R}{\gamma}\right)^2 \quad (1.35)$$



# Chapter 2

## 感知机代码仿真

### 2.1. 算法描述

代码实现思路为：初始化权重，如果实例点正确分类，则检查下一个实例点，指标指向下一个实例点，如果实例点被误分类，则更新权重，指标回到第一个检查的实例点重新开始检查，直到找到权重使得从第一个到最后一个实例点都被正确分类。

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 # 定义感知机模型类
5 class Perceptron:
6     def __init__(self):
7         self.count = 1
8         self.w = None
9         self.b = 0
10        self.l_late = 1
11
12    def fig(self):
13        # 模型方程
14        x_plot = np.linspace(2, 7, 10)
15        y_plot = -(perceptron.w[0]*x_plot+perceptron.b)/perceptron.w[1]
16        # 图示
17        plt.plot(x_plot, y_plot)
18        plt.scatter(x_trian[0:3,0], x_trian[0:3,1], marker='+', color='orange', label='postive')
19        plt.scatter(x_trian[3:6,0], x_trian[3:6,1], marker='+', color='green', label='negative')
20
21        # 设置图例
22        plt.legend(loc=1)
23
24        # 显示
25        plt.show()
26
27    # 算法主函数
28    def fit(self, x_trian, y_trian):
29        cnt=0
```

```

30     self.w=np.zeros(x_trian.shape[1])#shape 方法通常用于获取数组或矩阵的形状（维度），主要用于
        NumPy 库中的数组对象。
31     i = 0
32
33     while i < x_trian.shape[0]:
34         x = x_trian[i]
35         y = y_trian[i]
36         # 判断其所有点是否都没有误分类，如有更新w,b,重新判断
37         if y*(np.dot(self.w,x)+self.b)<=0:
38             self.w = self.w + self.l_rate*np.dot(x,y)
39             self.b = self.b+self.l_rate*y
40             i = 0
41         else:
42             i += 1
43
44         cnt=cnt+1
45
46     print("trainning_finish!\ntraining_cost\times:\t{}".format(cnt))

```

利用感知机模型，对下列的测试数据进行训练

```

1 # 训练集
2 x_trian = np.array([[3,3],[3,5],[4,7],[4,0],[2,1],[6,2]])
3 y_trian = np.array([1,1,1,-1,-1,-1])
4 print(x_trian.size,y_trian.size)

```

调用 Perceptron 类对数据进行训练，得到结果如下

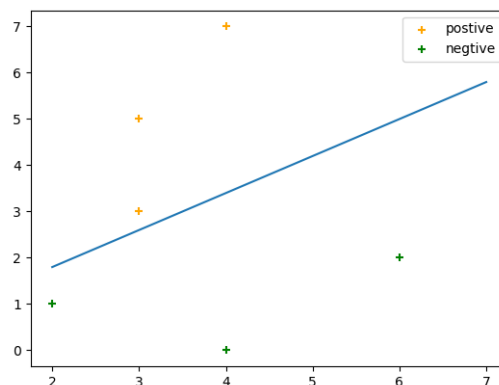


图 2.1: 训练成果

算法的训练次数为 21 次，权重向量为  $[-4.5.]$ ，偏置为  $-1$ 。

## 2.2. 实战

下面的例子是 *sklearn* 里面的鸢尾花的例子。首先导入需要的包和数据集。

```

1 import pandas as pd
2 import numpy as np
3 from sklearn.datasets import load_iris
4 from sklearn.linear_model import Perceptron
5 import matplotlib.pyplot as plt

```

加载数据集，并且把数据集转换为 Dataframe 类型。

```

1 # 处理数据
2 iris = load_iris()
3 data = pd.DataFrame(iris.data, columns=iris.feature_names)

```

DataFrame 是 pandas 中最重要的数据结构之一，类似于 Excel 表格或 SQL 表。DataFrame 可以包含多种类型的数据，包括整数、浮点数、字符串等，并且可以轻松地进行数据操作和分析。

鸢尾花的数据表格如下所示

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...	...	...	...	...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

图 2.2: 鸢尾花数据集

其中分为花萼的长宽和花瓣的长宽，接下来加载标签值。

```

1 data['label'] = iris.target # 标签值
2 data.columns = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'label']
3 print(data.head(4))

```

标签值如下

	sepal length	sepal width	petal length	petal width	label
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0

图 2.3: 鸢尾花数据集

我们用花萼的数据来训练模型，可视化数据集图像如下

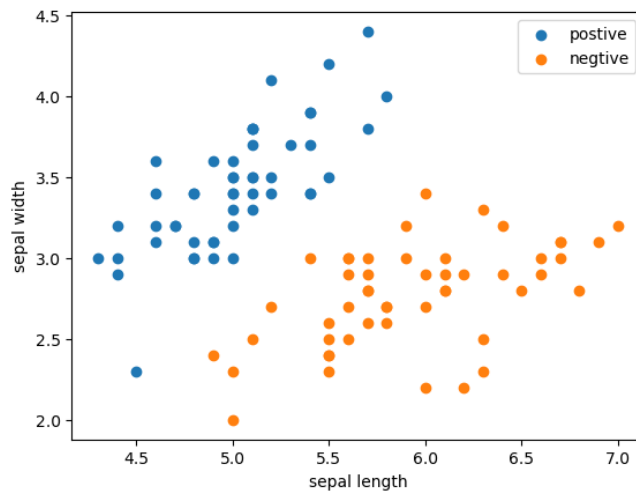


图 2.4: 鸢尾花数据集的二维分布

可以直观地看出，数据集明显是线性可分的。我们例化 `sklearn` 中的感知机模型，代码如下

```
1 x_train = np.array(data.iloc[:100,[0,1]])
2 y_train = np.array(data.iloc[:100,-1])
3
4 perceptron = Perceptron()
5 perceptron.fit(x_train,y_train)
6 print('w:',perceptron.coef_,'\n','b:',perceptron.intercept_,'\n','迭代次数:',perceptron.n_iter_)
7
8 # 评判模型
9 res = perceptron.score(x_train, y_train)
10 print('成功率:{:.0%}'.format(res))
```

拟合次数为 8 次，得到的拟合超平面如下

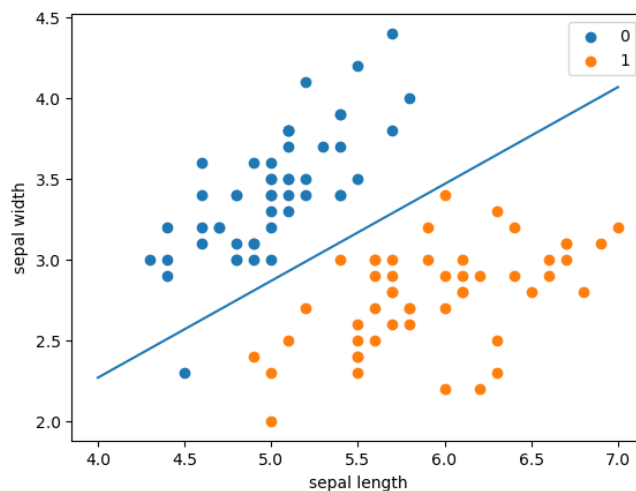


图 2.5: 鸢尾花数据集的二维分布

可见拟合结果中有一个被误分类，但是总体上是成功分类了。

# Chapter 3

## 感知机的一些问题

### 3.1. 感知机与逻辑运算

感知机是一个分类模型，可以把输出看作是为“真”或者“假”，因此可以类似逻辑门一样，可以猜想感知机可以用于实现逻辑运算。

### 3.2. 异或问题

# Chapter A

## Source Code Example

*Adding source code to your report/thesis is supported with the package listings. An example can be found below. Files can be added using `\lstinputlisting[language=<language>]{<filename>}`.*

```
1 """
2 ISA Calculator: import the function, specify the height and it will return a
3 list in the following format: [Temperature,Density,Pressure,Speed of Sound].
4 Note that there is no check to see if the maximum altitude is reached.
5 """
6
7 import math
8 g0 = 9.80665
9 R = 287.0
10 layer1 = [0, 288.15, 101325.0]
11 alt = [0,11000,20000,32000,47000,51000,71000,86000]
12 a = [-.0065,0,.0010,.0028,0,-.0028,-.0020]
13
14 def atmosphere(h):
15     for i in range(0,len(alt)-1):
16         if h >= alt[i]:
17             layer0 = layer1[:]
18             layer1[0] = min(h,alt[i+1])
19             if a[i] != 0:
20                 layer1[1] = layer0[1] + a[i]*(layer1[0]-layer0[0])
21                 layer1[2] = layer0[2] * (layer1[1]/layer0[1])**(-g0/(a[i]*R))
22             else:
23                 layer1[2] = layer0[2]*math.exp((-g0/(R*layer1[1]))*(layer1[0]-layer0[0]))
24     return [layer1[1],layer1[2]/(R*layer1[1]),layer1[2],math.sqrt(1.4*R*layer1[1])]
```

# Chapter B

## Latex Tikz Test Code

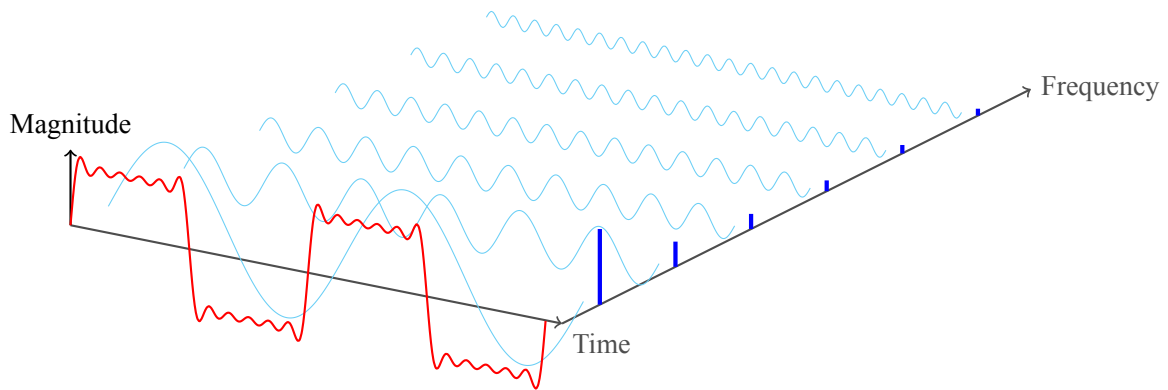


图 B.1: 测试图 1

```
1 \begin{figure}[H]
2   \begin{tikzpicture}[x={ (1cm,0.5cm)},z={ (0cm,0.5cm)},y={ (1cm,-0.2cm)}] %指定坐标系的方向与缩
   放
3     \draw[->,thick,black!70] (0,6.5,0) -- (6.2,6.5,0) node[right] {Frequency}; % 频率轴
4     \draw[->,thick,black!70] (0,0,0) -- (0,6.5,0) node[below right] {Time}; % 时间轴
5     \draw[->,thick] (0,0,0) -- (0,0,2) node[above] {Magnitude};
6
7     \foreach \n in {0.5,1.5,...,5.5}{
8       \draw [cyan!50, domain=0:2*pi,samples=200,smooth]
9         plot (\n,\x, {sin(4*\n*\x r)/\n });
10      \draw[blue, ultra thick] (\n,6.5,0) -- (\n,6.5,1/\n);
11    } % 频率逐渐增大振幅逐渐变小的正弦函数
12
13    \draw [red, thick, domain=0:2*pi,samples=200,smooth]
14      plot (0,\x, {sin(4*0.5*\x r)/0.5 + sin(4*1.5*\x r)/1.5 + sin(4*2.5*\x r)/2.5
15        + sin(4*3.5*\x r)/3.5 + sin(4*4.5*\x r)/4.5 + sin(4*5.5*\x r)/5.5} );
16    % 最后是手动加起来得到矩形波的逼近
17    \end{tikzpicture}
18    \caption{测试图1}
```



19 `\end{figure}`

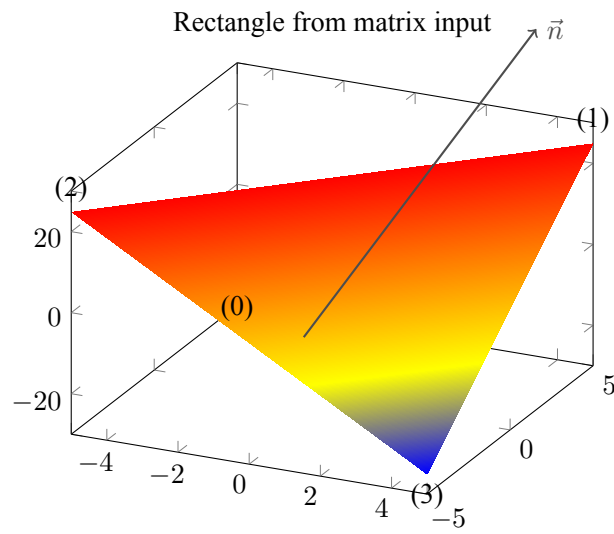


图 B.2: 三维绘图

```

1
2 \begin{tikzpicture}
3   \begin{axis}[nodes near coords={(\coordindex)},
4     title=Rectangle from matrix input]
5     % note that surf implies 'patch type=rectangle'
6     \addplot3[surf,shader=interp,samples=2,
7       patch type=rectangle]
8       {x*y};
9   \end{axis}
10 \end{tikzpicture}

```

# Chapter C

## Task Division Example

*If a task division is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.*

表 C.1: Distribution of the workload

Task	Student Name(s)
Summary	
Chapter 1 Introduction	
Chapter 2	
Chapter 3	
Chapter *	
Chapter * Conclusion	
Editors	
CAD and Figures	
Document Design and Layout	