国家科学技术学术著作出版基金
NFAPST

支持向量机

——理论、算法与拓展

邓乃扬 田英杰 著

科学出版社
www.sciencep.com

国家科学技术学术著作出版基金资助出版

支持向量机

——理论、算法与拓展

邓乃扬　田英杰　著

科学出版社

北京

内 容 简 介

本书以分类问题(模式识别、判别分析)和回归问题为背景，介绍支持向量机的基本理论、方法和应用. 特别强调对所讨论的问题和处理方法的实质进行直观的解释和说明，因此具有很强的可读性. 为使具有一般高等数学知识的读者能够顺利阅读，书中首先介绍了最优化的基础知识.

本书可作为理工类、管理学等专业的高年级本科生、研究生和教师的教材或教学参考书，也可供相关领域的科研人员和实际工作者阅读参考.

图书在版编目(CIP)数据

支持向量机：理论、算法与拓展/邓乃扬，田英杰著. —北京：科学出版社，2009

ISBN 978-7-03-025031-5

I. 支… II. ①邓… ②田… III. 向量计算机-算法理论 IV. TP301.6

中国版本图书馆 CIP 数据核字 (2009) 第 121196 号

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2009 年 8 月第 一 版 开本：B5(720×1000)
2009 年 8 月第一次印刷 印张：16
印数：1—2 500 字数：307 000

定价：48.00 元

(如有印装质量问题，我社负责调换)

序　言

支持向量机 (support vector machines, SVMs) 是借助于最优化方法解决数据挖掘中若干问题的有力工具, 它在一定程度上克服了“维数灾难”和“过学习”等传统困难, 并在文本分类、生物信息、语音识别、遥感图像分析、故障识别和预测、时间序列预测、信息安全等诸多领域有了成功的应用.

支持向量机不仅有着统计学习理论 (statistical learning theory, SLT) 的坚实理论基础, 而且具有直观的几何解释和完美的数学形式. 虽然自 20 世纪 90 年代由 Vapnik 提出以来一直处于飞速发展的阶段, 但是支持向量机的理论基础和各种算法实现的基本框架已经形成. 自 2000 年开始, 国外已陆续有专著出版.

2004 年, 作者在科学出版社出版了学术专著《数据挖掘中的新方法 —— 支持向量机》, 该书是国内第一本专门对支持向量机进行全面完整介绍和论述的著作. 出版后读者反映良好, 并被中国科学院研究生院、清华大学、中国农业大学等用作研究生教材, 被评为 2006 年北京市精品教材. 经过五年的继续研究与教学实践, 我们决定对该书进行全面改写, 增加新的研究成果, 并更名为“支持向量机 —— 理论、算法与拓展”.

本书特别强调可读性, 强调直观对理解问题实质的重要作用. 在给出系统严谨的论述之前, 一般先用图像等直观手段引进各种概念、方法和结论, 并特别注意对它们的本质给予形象的解释和说明. 对于原始文献中没有提及直观解释的内容, 我们也尽可能予以补充, 例如, 对求解多类分类问题的 Crammer-Singer 支持向量机, 我们给出了清晰的几何解释. 与第一部著作相比, 本书的逻辑系统更加清晰, 对问题的阐述也更加简明.

本书主要讨论分类问题和回归问题. 基于把回归问题转化为分类问题的研究工作, 全书以分类问题为主线, 形成了统一的格局. 主要内容如下: 第 1 章介绍最优化基础. 该章着重于凸规划的介绍, 添加了引领最优化方法应用研究的锥规划 (包括二阶锥规划和半定规划), 以及 Hilbert 空间中的凸规划理论. 这是以后章节以及进一步研究新的支持向量机的最优化基础. 第 2 章和第 3 章则分别对线性分类问题和线性回归问题直观地导出最基本的线性支持向量机. 第 4 章介绍核的基本概念, 并在此基础上介绍求解一般的分类问题和回归问题的支持向量机. 第 5 章则从最大间隔法的统计学解释入手, 讨论支持向量机的统计学习理论基础. 第 6 章介绍支持向量机实际应用中的模型选择问题. 第 7 章介绍实现支持向量机的几个主流算法. 第 8 章是前面讨论过的基本的支持向量机的变形与拓广, 包括求解多类分类

问题、半监督问题、带有扰动的问题和多示例问题的支持向量机.

本书包含了我们自己的研究工作. 例如第 5.6 节, 就是我们完善支持向量机的统计学习理论基础的研究成果. 我们相信, 这是迄今为止对 C- 支持向量分类机的一种最确切、最直接、最简明的统计学习理论解释. 此外, 这里还给出了 C- 支持向量分类机中的参数 C 以全新的意义. 又如第 8 章中给出的利用顺序回归的思想求解多类分类问题的支持向量机, 通过构建二阶锥规划或者半定规划模型求解半监督问题和带有扰动的问题的支持向量机, 以及处理多示例问题的支持向量机等内容.

本书所设定的读者范围较广, 既包括初涉支持向量机的人员, 也包括希望利用支持向量机解决实际问题的人士, 还可作为对支持向量机进行深入研究的参考书.

本书得以出版, 我们要感谢国家自然科学基金连续多年对我们研究工作的资助和国家科学技术学术著作出版基金的资助; 感谢北京航空航天大学的王日爽教授、曲阜师范大学的王长钰和王宜举教授、大连理工大学的夏尊铨和张立卫教授、北京交通大学的修乃华教授、北京理工大学的刘宝光教授、空军指挥学院的李意起教授、上海大学的白延琴教授、中国农业大学的经玲教授和甄苓副教授, 感谢中国科学院马志明院士、章祥荪研究员、石勇教授, 中国微软亚洲研究院李航研究员, 香港中文大学张树中教授对本书的关心和支持, 感谢桂林电子科技大学的朱志斌教授和中国人民大学的张春华博士给予的宝贵建议, 同时感谢我们讨论班的成员: 杨志霞博士、赵琨博士和王永翠、邵小健、秦如新、赵艳梅、高婷婷、邵元海、徐岩、李玉欣等同学, 他们都对本书提供了帮助.

由于作者水平所限, 书中难免有不妥之处, 欢迎读者批评指正. 来函请发至 dengnaiyang@vip.163.com 或 tianyingjie1213@163.com.

作　者

2009 年 1 月

符 号 表

R	实数集合
$x \in R^n$	输入和 n 维欧氏空间
$y \in \mathcal{Y}$	输出和输出集合
(x_i, y_i)	第 i 个训练点
$T = \{(x_1, y_1), \cdots, (x_l, y_l)\}$	训练集
l	训练点个数
$[x]_i$, $[x_i]_j$	向量 x 的第 i 个分量, 向量 x_i 的第 j 个分量
$\mathrm{x} = \Phi(x)$	Hilbert 空间中的向量和输入空间到 Hilbert 空间的映射
$[\mathrm{x}]_i$, $[\mathrm{x}_i]_j$	向量 x 的第 i 个分量, 向量 x_i 的第 j 个分量
$(x \cdot x'), (\mathrm{x} \cdot \mathrm{x}')$	x 与 x' 的内积, x 与 x' 的内积
$\mathcal{H}$	Hilbert 空间
w	R^n空间中的权向量
w_i	权向量 w 的第 i 个分量
w	Hilbert 空间中的权向量
w_i	权向量 w 的第 i 个分量
b	阈值
$K(x, x')$	核函数 $(\Phi(x) \cdot \Phi(x'))$
K	核矩阵 (Gram 矩阵)
$\\|\cdot\\|_p$	p-范数
$\\|\cdot\\|$	2-范数
$\\|\cdot\\|_1$	1-范数
h	VC 维
C	惩罚参数
ξ	松弛变量
ξ_i	松弛变量 ξ 的第 i 个分量
α	对偶变量, Lagrange 乘子
α_i	对偶变量 α 的第 i 个分量
β	对偶变量, Lagrange 乘子
β_i	对偶变量 β 的第 i 个分量
$P(\cdot)$	通常表示概率分布或概率

目　录

第 1 章　最优化基础

1.1　欧式空间上的最优化问题

本节讨论的欧式空间上的最优化问题就是通常所说的最优化问题.

1.1.1　最优化问题实例

我们通过一个实例说明这里要讲的最优化问题.

例 1.1.1　设平面上有两个线段 u_1u_2 和 v_1v_2. 试求这两个线段上相距最近的两个点 u^* 和 v^*.

这个问题可以写成一个最优化问题. 事实上, 线段 u_1u_2 和 v_1v_2 上的点分别可表为

$$u = \alpha u_1 + (1-\alpha)u_2, \quad \alpha \in [0,1] \tag{1.1.1}$$

和

$$v = \beta v_1 + (1-\beta)v_2, \quad \beta \in [0,1]. \tag{1.1.2}$$

显然, 点 u 和 v 的距离是变量 α 和 β 的函数

$$f(\alpha,\beta) = \|u-v\|^2 = a_{11}\alpha^2 - 2a_{12}\alpha\beta + a_{22}\beta^2 + b_1\alpha + b_2\beta + c, \tag{1.1.3}$$

其中

$$\begin{aligned} &a_{11} = \|u_1-u_2\|^2, && a_{12} = (u_1-u_2)^{\mathrm{T}}(v_1-v_2), && a_{22} = \|v_1-v_2\|^2, \\ &b_1 = 2(u_1-u_2)^{\mathrm{T}}(u_2-v_2), && b_2 = 2(v_1-v_2)^{\mathrm{T}}(v_2-u_2), && c = \|u_2-v_2\|^2. \end{aligned} \tag{1.1.4}$$

因此, 我们需要在按上式确定系数 $a_{ij},\ i,j=1,2,\ b_i,\ i=1,2$ 和 c 之后, 选择 α 和 β, 使由式 (1.1.3) 给出的函数 $f(\alpha,\beta)$ 取最小值. 注意, 现在并不允许 α 和 β 任意取值, 而是限定在 $\alpha \in [0,1]$ 和 $\beta \in [0,1]$ 的范围内, 选择使 $f(\alpha,\beta)$ 取最小值的 α,β. 这个问题可表述成以下形式

$$\begin{aligned} \min \quad & f(\alpha,\beta) = a_{11}\alpha^2 - 2a_{12}\alpha\beta + a_{22}\beta^2 + b_1\alpha + b_2\beta + c, && (1.1.5) \\ \text{s.t.} \quad & 0 \leqslant \alpha \leqslant 1, && (1.1.6) \\ & 0 \leqslant \beta \leqslant 1, && (1.1.7) \end{aligned}$$

其中系数 $a_{ij},\ i,j=1,2,\ b_i,\ i=1,2$ 和 c 由式 (1.1.4) 给出. 这里的“min”表示“最小化 (minimize)”, 而“s.t.”表示“受限于 (subject to)”. 所以问题 (1.1.5)~(1.1.7)

的含义是：考虑两个变量 α 和 β, 在式 (1.1.6) 和式 (1.1.7) 所示的范围内, 寻求使式 (1.1.5) 所示的函数 f 取最小值的 α^* 和 β^*. 把 α^* 和 β^* 代入式 (1.1.1) 和式 (1.1.2), 便可得到两线段上相距最近的点 u^* 和 v^*：

$$u^* = \alpha^* u_1 + (1-\alpha^*)u_2, \quad v^* = \beta^* v_1 + (1-\beta^*)v_2. \tag{1.1.8}$$

这里我们关心的是问题 (1.1.5)~(1.1.7). 如果引进一个 2 维向量 $x = ([x]_1, [x]_2)^{\mathrm{T}} = (\alpha,\beta)^{\mathrm{T}}$, 则上述问题可表述为

$$
\begin{aligned}
\min \quad & f_0(x) = a_{11}[x]_1^2 - 2a_{12}[x]_1[x]_2 + a_{22}[x]_2^2 + b_1[x]_1 + b_2[x]_2 + c, && (1.1.9)\\
\text{s.t.} \quad & -[x]_1 \leqslant 0, && (1.1.10)\\
& [x]_1 - 1 \leqslant 0, && (1.1.11)\\
& -[x]_2 \leqslant 0, && (1.1.12)\\
& [x]_2 - 1 \leqslant 0, && (1.1.13)
\end{aligned}
$$

其中 $a_{ij},\ i,j=1,2,\ b_i,\ i=1,2$ 和 c 是已知的系数.

1.1.2　最优化问题及其解

我们可以拓广问题 (1.1.9)~(1.1.13) 为更一般的形式：把 2 个变量构成的 2 维向量 x 拓广为由 n 个变量构成的 n 维向量, 把其中的函数拓广为一般的函数, 把 4 个不等式形式的限定条件拓广为 m 个, 再增加 p 个等式形式的限定条件. 这样便得到了一般形式的最优化问题

$$
\begin{aligned}
\min \quad & f_0(x), && (1.1.14)\\
\text{s.t.} \quad & f_i(x) \leqslant 0,\ i=1,\cdots,m, && (1.1.15)\\
& h_i(x) = 0,\ i=1,\cdots,p. && (1.1.16)
\end{aligned}
$$

上述最优化问题是目标函数的最小化 ($\min f_0(x)$). 从本质上说, 它也涵盖了目标函数的最大化 ($\max \bar{f}(x)$) 问题, 因为只需取 $f_0(x) = -\bar{f}(x)$ 就可以把后者转化为前者. 因此可以认为这类最大化问题也是最优化问题.

定义 1.1.2(无约束问题和约束问题)　当问题 (1.1.14)~(1.1.16) 中的 $m=p=0$ 时, 即它不含任何限定条件时, 称该问题为无约束问题. 当 $m+p>0$, 即包含限定条件时，称为约束问题.

定义 1.1.3(目标函数、约束条件和约束函数)　称问题 (1.1.14)~(1.1.16) 中的 $f_0(x)$ 为目标函数. 称其中的 $f_i(x) \leqslant 0, i=1,\cdots,m$ 和 $h_i(x)=0, i=1,\cdots,p$ 为约束条件, 并分别称它们为不等式约束和等式约束. 称 $f_i(x), i=1,\cdots,m$ 和 $h_i(x)$, $i=1,\cdots,p$ 为约束函数.

定义 1.1.4(可行点和可行域) 称满足所有约束条件的点为可行点, 并称全体可行点组成的集合 D 为可行域

$$D=\{x|f_i(x)\leqslant 0, i=1,\cdots,m\,;h_i(x)=0\,,i=1,\cdots,p\,;x\in R^n\}\,. \qquad (1.1.17)$$

显然, 无约束问题的可行域是整个 R^n 空间, 而约束问题的可行域往往是 R^n 空间中的一部分.

定义 1.1.5(最优值) 问题 (1.1.14)~(1.1.16) 的最优值是指目标函数 f_0 在可行域 D 上的下确界 p^*:

$$p^*=\inf\{f_0(x)|x\in D\}, \qquad (1.1.18)$$

其中 D 是由定义 1.1.4 给出的问题的可行域. 如果可行域是空集 (无可行点), 则定义 $p^*=\infty$.

定义 1.1.6(整体解和局部解) 称 x^* 是问题 (1.1.14)~(1.1.16) 的整体解, 如果 x^* 是可行点, 而且目标函数在 x^* 处达到问题的最优值, 即 $f_0(x^*)=p^*$, 其中 p^* 是问题的最优值. 称 x^* 是问题 (1.1.14)~(1.1.16)的局部解, 如果 x^* 是可行点, 而且存在着 $\varepsilon>0$, 使得

$$f_0(x^*)=\inf\{f_0(x)|x\in D\,;\|x-x^*\|\leqslant\varepsilon\}, \qquad (1.1.19)$$

其中 D 是由定义 1.1.4 给出的问题的可行域. 换句话说, 如果 x^* 是下列问题的整体解

$$\begin{aligned}
\min\quad & f_0(x), && (1.1.20)\\
\text{s.t.}\quad & f_i(x)\leqslant 0,\ i=1,\cdots,m, && (1.1.21)\\
& h_i(x)=0,\ i=1,\cdots,p, && (1.1.22)\\
& \|x-x^*\|\leqslant\varepsilon. && (1.1.23)
\end{aligned}$$

显然, 我们提出最优化问题的初衷是寻求它的整体解, 局部解可以看作是放松了一些条件的整体解. 此外, 问题的整体解和局部解都可能不止一个, 我们分别称所有整体解和所有局部解组成的集合为相对应的解集. 请注意, 以后所说的“问题的解”常常指局部解.

1.1.3 最优化问题的几何解释

通过两个变量的最优化问题, 可以十分清晰地理解其几何意义. 现在用一个例子予以说明.

例 1.1.7 设例 1.1.1 中的线段 u_1u_2 和 v_1v_2 由下列方式给定:

$$u_1=(0,0)^{\mathrm T},\quad u_2=(1,0)^{\mathrm T},\quad v_1=(1,1)^{\mathrm T},\quad v_2=(2,2)^{\mathrm T}, \qquad (1.1.24)$$

参看图 1.1.1. 于是最优化问题 (1.1.9)~(1.1.13) 具体化为

$$
\begin{aligned}
\min \quad & f_0(x) = [x]_1^2 - 2[x]_1[x]_2 + 2[x]_2^2 + 2[x]_1 - 6[x]_2 + 5, & (1.1.25)\\
\text{s.t.} \quad & f_1(x) = -[x]_1 \leqslant 0, & (1.1.26)\\
& f_2(x) = [x]_1 - 1 \leqslant 0, & (1.1.27)\\
& f_3(x) = -[x]_2 \leqslant 0, & (1.1.28)\\
& f_4(x) = [x]_2 - 1 \leqslant 0. & (1.1.29)
\end{aligned}
$$

试通过考察目标函数和可行域的几何图像求解该问题.

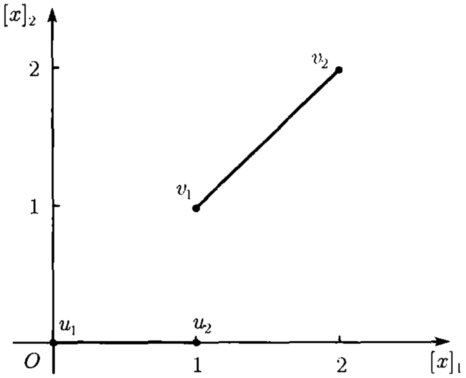

图 1.1.1

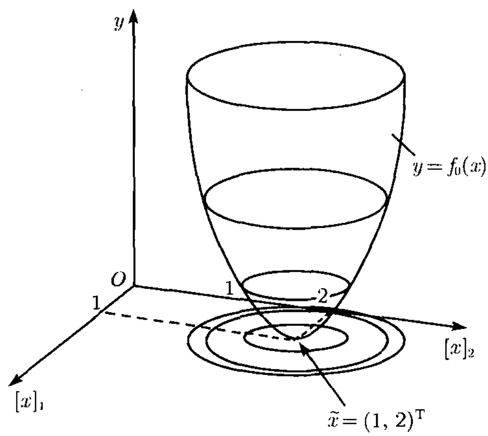

图 1.1.2

先考虑目标函数 $f_0(x)$, 经过变形, 可以写为

$$f_0(x) = ([x]_1 - [x]_2 + 1)^2 + ([x]_2 - 2)^2. \tag{1.1.30}$$

图 1.1.2 给出了曲面 $y = f_0(x)$ 的图像. 该曲面的最低点恰好在平面 $[x]_1 O [x]_2$ 上, 它在该平面的坐标为 $\tilde{x} = (1,2)^{\mathrm{T}}$, 所以 $\tilde{x}$ 是无约束问题 $\min f_0(x)$ 的整体解. 然而

它不是约束问题的解, 因为约束条件 (1.1.26)~(1.1.29) 表明可行域由正方形 $ABOC$ 的内部及边界组成, $\tilde{x}$ 不在可行域内, 参看图 1.1.3. 为得到约束问题的解, 我们把 $\tilde{x}$ 看作水平面 $y=0$ 和曲面 $y=f_0(x)$ 的交点在平面 $[x]_1O[x]_2$ 上的投影. 同时把水平面 $y=0$ 上移至 $y=k$. 考虑当 k 增加时水平面 $y=k$ 和曲面 $y=f_0(x)$ 的交线在平面 $[x]_1O[x]_2$ 上的投影. 它们应该是以 $\tilde{x}$ 为中心的不断膨胀的一族椭圆, 参看图 1.1.2 和图 1.1.3. 不难看出, 当 k 达到 1 时, 椭圆开始与可行域相交, 其交点为四边形的顶点$x^*=B=(0,1)^{\mathrm{T}}$. 它就是约束问题的整体解, 因为该点处的目标函数值为 $f_0(x^*)=1$, 而对应 $f_0(x)=1$ 的等高线 (椭圆) 在 x^* 处的切线恰好是水平的, 正方形 $ABOC$ 中除 x^* 的任意点都不会位于等高线 $f_0(x)=1$ 上, 也不会进入它的内部.

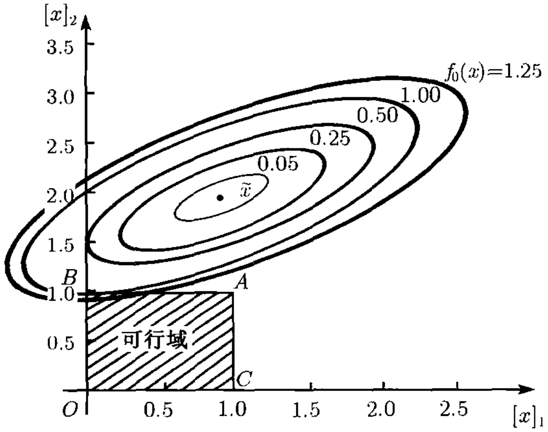

图 1.1.3

本节介绍的最优化问题 (1.1.14)~(1.1.16) 中的目标函数和约束函数几乎可以是任意的函数, 其涵盖范围相当广泛. 对于这类问题, 本书不再做进一步介绍, 有兴趣的读者可参看文献 [3, 4, 7, 8, 10, 13, 14, 29, 55, 73, 82, 83]. 下面我们讨论一些特殊的最优化问题.

1.2 欧式空间上的凸规划

本节介绍上节所述的最优化问题中一类重要且极具应用价值的问题, 即欧式空间上的凸规划问题. 对这类问题的研究相对比较成熟, 可参看文献 [32, 33, 39] 等.

1.2.1 凸集和凸函数

1. 凸集

首先引入 n 维欧式空间 R^n 上的凸集的概念.

定义 1.2.1(凸集) 设集合 $S \subset R^n$. 称 S 是凸集, 如果对任意 $x_1, x_2 \in S$ 和任意的 $\lambda \in [0,1]$, 都有

$$\lambda x_1 + (1-\lambda)x_2 \in S\ . \tag{1.2.1}$$

平面 R^2 上的凸集有着明显的几何意义. 直观地说, 图 1.2.1(a) 所示曲线内部形成的集合 S_1 就是一个凸集, 因为连接其中任意两点的线段完全属于这个集合, 而图 1.2.1(b) 所示曲线内部形成的集合 S_2 就不是一个凸集, 因为它不具有上述性质.

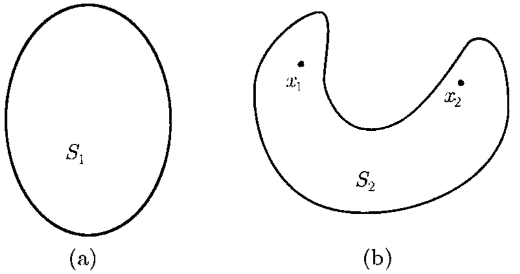

图 1.2.1

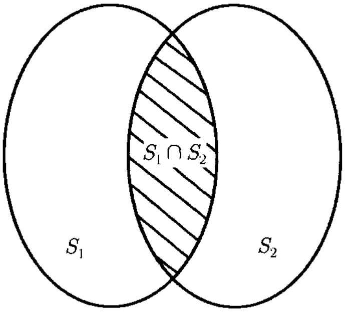

图 1.2.2

定理 1.2.2 若 S_1 和 S_2 都是凸集, 则交集 $S_1 \cap S_2$ 也是凸集.

从图 1.2.2 可以体会定理 1.2.2 的正确性.

2. 凸函数

下面引进定义在 R^n 空间上的凸函数的概念.

定义 1.2.3(凸函数) 设 $S \subset R^n$ 是非空凸集, f 是定义在 S 上的函数. 称函数 f 是 S 上的凸函数, 如果对任意 $x_1, x_2 \in S$ 和任意的 $\lambda \in (0,1)$, 都有

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2)\ . \tag{1.2.2}$$

称 f 为 S 上的严格凸函数, 如果当 $x_1 \neq x_2$ 时, 式(1.2.2) 中的严格不等号成立.

对于 1 维空间 R 上的函数, 即单变量的函数 $y = f(x)$ 来说, 式(1.2.2)的几何意义是很清楚的: 曲线上任意两点 $(x_1, f(x_1))$ 和 $(x_2, f(x_2))$ 的连线都位于曲线 $y = f(x)$ 的上方. 因此凸函数的图像是向下凸的 (参看图 1.2.3(a)), 而图 1.2.3(b) 所示的函数就不是凸函数. 类似地, 不难想象相应的 2 维空间上的凸函数的图像大体如图 1.2.4(a) 所示, 而图 1.2.4(b) 和图 1.2.4(c) 所示的函数都不是凸函数.

下面研究连续可微凸函数的特征. 先考虑连续可微的 1 维空间 R 上的凸函数 $f(x)$. 回忆微积分的知识我们知道, 它的向下凸的特性可用“它的二阶导数 f'' 非负”来描述. 因此, 对任意的点 $\bar{x}$, 应有

$$f(x) = f(\bar{x}) + f'(\bar{x})(x-\bar{x}) + \frac{1}{2}f''(\bar{x}+\theta(x-\bar{x}))(x-\bar{x})^2 \geqslant f(\bar{x}) + f'(\bar{x})(x-\bar{x}), \tag{1.2.3}$$

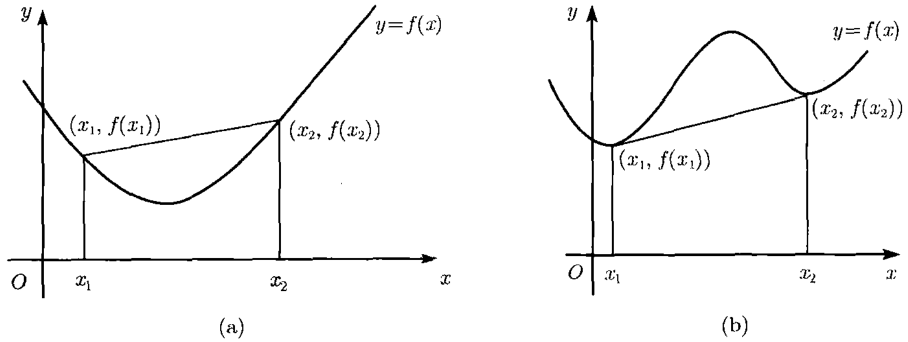

图 1.2.3

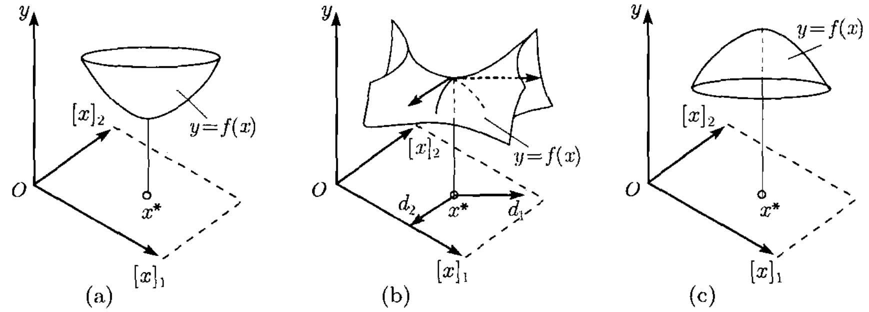

图 1.2.4

其中 $\theta \in (0,1)$. 此式有着明显的几何意义: 对任意的 $\bar{x}$, 曲线 $y = f(x)$ 位于曲线在 $\bar{x}$ 处的切线

$$y = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) \tag{1.2.4}$$

的上方, 如图 1.2.5 所示. 可以设想 2 维空间 R^2 中的连续可微的凸函数 $f(x)$ 应该有类似的特征: 考虑图 1.2.4(a) 中函数 $y = f(x)$ 的曲面, 则对任意的 $\bar{x}$, 该曲面位

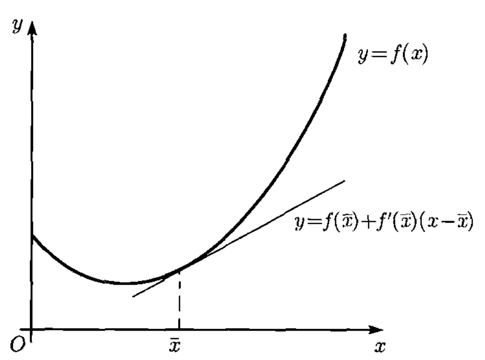

图 1.2.5