

勘误修订

[本书因颇受欢迎，出版社提出重印，于是作者借机要求在每次重印时加入新的修订，省却让读者等待第二版的麻烦。为方便读者，所有修订内容都列举在此。其中部分修订是为了更便于读者理解，并非原文有误]

(第一版第 30 次印刷, 2018 年 12 月):

p.38, 式 (2.27): " \max " \rightarrow " \min "

p.39, 倒数第 1 行: "若平均错误率……临界值范围" \rightarrow "若 τ_t 位于临界值范围"

p.152, 第 10 行: " $6/8 = 0.750$ " \rightarrow " $5/8 = 0.625$ "

p.153, 第 3 行: " 0.063 " \rightarrow " 0.052 "

p.153, 第 6 行: " 0.063 " \rightarrow " 0.052 "

p.172, 式 (8.2): " H " \rightarrow " F "

p.173, 式 (8.3): " H " \rightarrow " F "

p.174, 图 8.3 最后一行: " H " \rightarrow " F "

p.175, 式 (8.12) 前一行: "最小化" \rightarrow "最小化 $\ell_{\text{exp}}(H_{t-1} + \alpha_t h_t \mid \mathcal{D})$, 可简化为最小化"

p.185, 式 (8.28) 前一行: "是" \rightarrow "定义为"

p.284, 倒数第三行: " $y_i =$ " \rightarrow " $y_i \in$ "

(第一版第 29 次印刷, 2018 年 11 月)

(第一版第 28 次印刷, 2018 年 8 月)

(第一版第 27 次印刷, 2018 年 6 月):

p.42, 表 2.5 后第四行: " $(k^2 - 1)/12$ " \rightarrow " $(k^2 - 1)/12N$ "

p.159, 倒数第 9 行、第 10 行中两处: "字节长度" \rightarrow "编码位数"

p.160, 第 1 行、第 4 行中两处: "字节数" \rightarrow "编码位数"

p.160, 第 7 行、第 10 行中两处: "字节" \rightarrow "编码位"

p.174, 式 (8.7) 中两处、(8.8) 中 7 处、边注中两处: " $f(x)$ " \rightarrow " $f(\mathbf{x})$ "

(第一版第 26 次印刷, 2018 年 5 月):

p.58, 倒数第二行: "对率函数" \rightarrow "下面我们会看到, 对率回归求解的目标函数"

p.230, 倒数第三行: "方差" \rightarrow "协方差矩阵"

(第一版第 25 次印刷, 2018 年 3 月):

p.39, 最后一行: " $[-\infty, \cdot \rightarrow (-\infty, \cdot, \cdot, \infty] \rightarrow \cdot, \infty$)"

p.199, 式 (9.12): 分母的 " (μ_i, μ_j) " \rightarrow " (C_i, C_j) "

(第一版第 24 次印刷, 2018 年 1 月):

p.112, 图 5.14a: 修订文件

p.303, 倒数第二行: 去掉 "[Zhou et al., 2004]"

p.304, 第一行: "当" \rightarrow "考虑到有标记样本通常很少而未标记样本很多, 为缓解过拟合, 可在式 (13.21) 中引入针对未标记样本的 L_2 范数项 $\mu \sum_{i=l+1}^{l+u} \|\mathbf{F}_i\|^2$, 在"; 同时插入边注: "参见 11.4 节"

(第一版第 23 次印刷, 2017 年 10 月):

p.27, 式 (2.1): 第一个 " \mapsto " \rightarrow " \rightarrow ", 第二个 " \mapsto " \rightarrow " $=$ "

p.80, 倒数第 2 行: "算法 4.2" \rightarrow "图 4.2 算法"

p.131, 图 6.5: 修订文件

(第一版第 22 次印刷, 2017 年 9 月):

p.156, 倒数第 7 行: "(7.23)" \rightarrow "(7.21)"

p.320, 第 8 行: "其余 $n - 2$ " \rightarrow "此前 $t - 2$ "

(第一版第 21 次印刷, 2017 年 8 月)

(第一版第 20 次印刷, 2017 年 7 月):

p.60, 图 3.3 中: " $y = \mathbf{w}^T \mathbf{x}$, y " \rightarrow "投影方向 \mathbf{w} "

p.133, 式 (6.42) 加边注: "传统意义上的"结构风险"是指引入模型结构因素后的总体风险 (或许更宜译为"带结构风险"), 本书则是指总体风险中直接对应于模型结构因素的部分, 这样从字面上更直观, 或有助于理解其与机器学习其他内容间的联系. 参见 p.160."

(第一版第 19 次印刷, 2017 年 6 月):

p.159, 第一行加边注: "一般需先对图剪枝, 仅保留有向图中 x, y, \mathbf{z} 及它们的祖先结点"

p.230, 式 (10.15) 上面一行加边注: ” 严格来说, 协方差矩阵是 $\frac{1}{m-1} \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T$, 但前面的常数项在此不发生影响”

(第一版第 18 次印刷, 2017 年 5 月):

p.187, 式 (8.39) 下面一行: ” \leq ” \rightarrow ” \geq ”

(第一版第 17 次印刷, 2017 年 4 月):

p.384, 图 16.10, 步骤 9: ” $\pi(x, a)$ ” \rightarrow ” $\pi(x)$ ”

p.388, 图 16.13, 步骤 4: ” $\pi^\epsilon(x)$ ” \rightarrow ” $a = \pi^\epsilon(x)$ ”

p.388, 图 16.13, 步骤 8: 去掉”, $a = a'$ ”

(第一版第 16 次印刷, 2017 年 3 月):

p.417, 第 3 段第 1 行: ” 通往人工智能的途径” \rightarrow ” 一种人工智能途径”

(第一版第 15 次印刷, 2017 年 2 月):

p.206, 9.4.3 节前倒数第 5 行: ” c_2 ” \rightarrow ” c_1 ”

(第一版第 14 次印刷, 2016 年 12 月):

p.34, 图 2.4(b): 修订文件

p.206, 9.4.3 节前倒数第 2 行: ”(0.722; 0.442)” \rightarrow ”(0.722; 0.447)”

p.209, 式 (9.38) 上面一行: ” 样本” \rightarrow ” 混合成分”

p.215, 图 9.11 第 5 步: ” $j = 1, 2, \dots, m$ ” \rightarrow ” $j = i + 1, \dots, m$ ”

p.230, 式 (10.14) 结尾: ”” \rightarrow ”,”

p.230, 式 (10.14) 下面一行开头顶格插入: ” 其中 $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d)$.”

p.231, 式 (10.17): 两处” \mathbf{W} ” \rightarrow ” \mathbf{w}_i ”, ” λ ” \rightarrow ” λ_i ”

p.231, 式 (10.17) 下面第二行: ” \mathbf{W} ” \rightarrow ” \mathbf{W}^* ”

p.231, 图 10.5 最后一行: ” \mathbf{W} ” \rightarrow ” \mathbf{W}^* ”

p.232, 第一行: ” \mathbf{W} ” \rightarrow ” \mathbf{W}^* ”

p.232, 式 (10.19) 前第二行: ” \mathbf{W} ” \rightarrow ” $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d)$ ”

p.232, 式 (10.19) 前第二行: ” 即 PCA 欲求解” \rightarrow ” 则对于 \mathbf{w}_j , 由式 (10.17) 有”

p.232, 式 (10.19): 两处” \mathbf{W} ” \rightarrow ” \mathbf{w}_j ”; ” λ ” \rightarrow ” λ_j ”

p.233, 式 (10.20): 三处” \mathbf{W} ” \rightarrow ” \mathbf{w}_j ”; 两处” λ ” \rightarrow ” λ_j ”; ” α_i ” \rightarrow ” α_i^j ”

p.233, 式 (10.20) 下一行: ” α_i ” \rightarrow ” α_i^j ”; ” λ ” \rightarrow ” λ_j ”; ” \mathbf{W} ” \rightarrow ” \mathbf{w}_j ”

p.233, 式 (10.20) 下一行: ". 假定" \rightarrow "是 α_i 的第 j 个分量. 假定"

p.233, 式 (10.21): 两处 " \mathbf{W} " \rightarrow " \mathbf{w}_j "; " λ " \rightarrow " λ_j "

p.233, 式 (10.22): " \mathbf{W} " \rightarrow " \mathbf{w}_j "; " α_i " \rightarrow " α_i^j "

p.233, 式 (10.24): 两处 " \mathbf{A} " \rightarrow " α^j "; " λ " \rightarrow " λ_j "

p.233, 式 (10.24) 下面一行: " $\mathbf{A} = (\alpha_1; \alpha_2; \dots; \alpha_m)$ " \rightarrow " $\alpha^j = (\alpha_1^j; \alpha_2^j; \dots; \alpha_m^j)$ "

p.233, 式 (10.25) 下面一行: 去掉", α_i^j 是 α_i 的第 j 个分量"

(第一版第 13 次印刷, 2016 年 11 月):

p.36, 倒数第 5 行: "(TPR, FPR)" \rightarrow "(FPR, TPR)"

p.120, 第 7 行: "(1927 -)" \rightarrow "(1927 - 2016)"

p.203, 图 9.2 下面一行: " \mathbf{x}_{27} " \rightarrow " \mathbf{x}_{24} "

p.203, 图 9.2 下面第 3 行: "(0.532; 0.472)" \rightarrow "(0.478; 0.437)"

p.203, 图 9.2 下面第 5 行: "0.166" \rightarrow "0.220"

p.203, 图 9.2 下面第 7 行: 大括号中增加" \mathbf{x}_3 ", 去掉" \mathbf{x}_{15} "

p.203, 倒数第 5 行: 大括号中去掉" \mathbf{x}_3 ", 增加" \mathbf{x}_{15} "

p.203, 倒数第 3 行: "(0.473; 0.214)" \rightarrow "(0.493; 0.207)"

p.203, 倒数第 3 行: "(0.623; 0.388)" \rightarrow "(0.602; 0.396)"

p.204, 图 9.3: 修订文件

(第一版第 12 次印刷, 2016 年 11 月)

(第一版第 11 次印刷, 2016 年 10 月)

(第一版第 10 次印刷, 2016 年 9 月):

p.156, 式 (7.24) 分母: " N_i " \rightarrow " $N \times N_i$ "

p.156, 式 (7.25) 下面一行: "其中 N_i " \rightarrow "其中 N 是 D 中可能的类别数, N_i "

p.156, 式 (7.25) 下面第 4 行, 分母: " $17 + 3$ " \rightarrow " $17 + 3 \times 2$ "

p.156, 式 (7.25) 下面第 4 行: "0.350" \rightarrow "0.304"

(第一版第 9 次印刷, 2016 年 8 月)

(第一版第 8 次印刷, 2016 年 5 月):

p.5, 第 2 段倒数第 3 行: "3、2、2" \rightarrow "3、3、3"

p.5, 第 2 段倒数第 2 行: " $4 \times 3 \times 3 + 1 = 37$ " \rightarrow " $4 \times 4 \times 4 + 1 = 65$ "

p.26, 边注第 2 行: "2.6 节" \rightarrow "2.5 节"

p.41, 式 (2.33) 上面一行: "正态分布, 且均值……因此变量" \rightarrow "正态分布. McNemar 检验考虑变量"

p.41, 式 (2.33) 旁加边注: " $e_{01} + e_{10}$ 通常很小, 需考虑连续性校正, 因此分子中有 -1 项"

p.45, 第一个边注: "由式 (2.37)" \rightarrow "考虑到噪声不依赖于 f , 由式 (2.37)"

p.63, 式 (3.45) 下面一行: " $N - 1$ 个最大" \rightarrow " d' 个最大非零"

p.63, 式 (3.45) 下面第 2 行: "矩阵." \rightarrow "矩阵, $d' \leq N - 1$."; 加边注: "最多有 $N - 1$ 个非零特征值"

p.63, 式 (3.45) 下面第 3 行: " $N - 1$ 维" \rightarrow " d' 维"

p.63, 式 (3.45) 下面第 4 行: " $N - 1$ 通常远小于数据原有的属性数" \rightarrow " d' 通常远小于数据原有的属性数 d "

p.100, 图 5.5, 左图最上面的"阈值 0.5" \rightarrow "阈值 1.5"

p.100, 图 5.5, 左图最右边的"阈值 0.5" \rightarrow "阈值 -1.5 "

p.100, 图 5.5, 左图中间的" $1 -1 -1 1$ " \rightarrow " $1 1 -1 -1$ "

p.125, 式 (6.18): " y_s " \rightarrow " $1/y_s$ "

p.136, 式 (6.54): 右边最后一项中的四处" i " \rightarrow " j "

p.136, 式 (6.54): 右边最后一项中最后的" \mathbf{x} " \rightarrow " \mathbf{x}_i "

p.152, 第三个式子等号右端: " 0.375 " \rightarrow " 0.625 "

p.153, 第 3 行: " 0.038 " \rightarrow " 0.063 "

p.153, 第 6 行: " 0.038 " \rightarrow " 0.063 "

p.160, 式 (7.29) 下面第 2 行: "需多少字节来描述 D " \rightarrow "对 D 描述得有多好"; 加边注: "可以从统计学习角度理解, 将两项分别视为结构风险和经验风险"

p.239, 式 (10.39) 第二行式子: 去掉上标" 2 "

p.244, 第 13 行: "Locally" \rightarrow "Nonlinear dimensionality reduction by locally"

p.244, 第 14 行: " 2316 " \rightarrow " 2326 "

p.249, 式 (11.2): " $i = 1$ " \rightarrow " $k = 1$ "

p.253, 倒数第 5 行: "[Boyd and Vandenberghe, 2004]" \rightarrow "[Combettes and Wajs, 2005]"

p.263, 倒数第 4 行, 插入: "Combettes, P. L. and V. R. Wajs. (2005). "Signal recovery by proximal forward-backward splitting." *Mutiscale Modeling & Simulation*, 4(4):1168–1200."

p.277, 式 (12.29): " $E(h) - \hat{E}(h)$ " \rightarrow " $|E(h) - \hat{E}(h)|$ "
p.299, 式 (13.9) 后第三段第 2 行: "关于 D_u " \rightarrow "涉及 C_u "

(第一版第 7 次印刷, 2016 年 4 月):

p.42, 表 2.5 下面一段的第三行: "服从正态分布, 其均值" \rightarrow "的均值"
p.42, 倒数第二行加边注: "原始检验要求 k 较大 (例如 > 30), 若 k 较小则倾向于认为无显著区别"

(第一版第 6 次印刷, 2016 年 4 月):

p.56, 图 3.1 中, 红色第一和第二个点的坐标互换
p.114, 图 5.15 中, 卷积层 $16@10 \times 10$ 和采样层 $16@5 \times 5$ 各去掉 8 个方块
p.301, 式 (13.12) 的下一行: " $(\mathbf{f}_l^T \mathbf{f}_u^T)^T$ " \rightarrow " $(\mathbf{f}_l^T; \mathbf{f}_u^T)$ "
p.372, 图 16.2: 从"s= 健康"到"s= 溢水"的" $r=1$ " \rightarrow " $r=-1$ "
p.376, 图 16.5 的边注: "第 4 行中式 (16.4) 的参数" \rightarrow "该参数在第 4 行使用"
p.385, 第二行: "在使用策略时并不需要 ϵ - 贪心" \rightarrow "而不是为了最终使用"
p.387, 倒数第二行: " ϵ - 贪心策略, 而执行 (第 5 行) 的是原始策略" \rightarrow "原始策略, 而执行 (第 4 行) 的是 ϵ - 贪心策略"
p.393, 第四段第一行: 去掉"[Kuleshov and Precup, 2000] 和"
p.395, 去掉最后一行
p.396, 去掉第一行
p.402, 式 (A.32) 加边注: "机器学习中 \mathbf{W} 通常是对称矩阵"

(第一版第 5 次印刷, 2016 年 3 月):

p.62, 第 1 行加边注: " $(\mu_0 - \mu_1)^T \mathbf{w}$ 是标量"
p.78, 图 4.4, 从右往左数: 第二个叶结点改为"好瓜", 第三个叶结点改为"坏瓜"
p.85, 图 4.8, 从右往左数: 第二个叶结点改为"好瓜", 第三个叶结点改为"坏瓜"
p.85, 图 4.8, 中间分支底层: "硬挺" \rightarrow "硬滑"
p.89, 图 4.9, 中间分支底层: "硬挺" \rightarrow "硬滑"
p.103, 最后一行的式子: 求和的" q " \rightarrow " l "

p.399, 式 (A.9): " $A_{1\sigma n}$ " \rightarrow " $A_{n\sigma n}$ "
p.400, 第 1 行: "(1,4,3,2)" \rightarrow "(3,1,2)"
p.402, 式 (A.32) 最后一行的式子中: " $2\mathbf{A}$ " \rightarrow " $2\mathbf{A}^T$ "

(第一版第 4 次印刷, 2016 年 3 月):

p.59, 式 (3.27) 加边注: "考虑 $y_i \in \{0, 1\}$ "

(第一版第 3 次印刷, 2016 年 3 月):

p.15, 第 5 行: "居功" \rightarrow "厥功"
p.55, 最后一行: 式子括号中的逗号改为分号
p.125, 第 3 行: "减小" \rightarrow "增大"
p.125, 第 4 行, 第 6 行: "减幅" \rightarrow "增幅"
p.125, 第 5 行: "减小" \rightarrow "增长"

(第一版第 2 次印刷, 2016 年 2 月):

p.38, 第 6 行: " $\epsilon^{m'}$ " \rightarrow " $\binom{m}{m'}\epsilon^{m'}$ "
p.119, 第 14 行: "318-362" \rightarrow "533-536"
p.404, 式 (B.3) 最后一行的式子 \rightarrow " $\lambda g(\mathbf{x}) = 0$ "

(第一版第 1 次印刷, 2016 年 1 月):

p.6, 图 1.2: 图中两处"清脆" \rightarrow "浊响"
p.28, 第 3 段倒数第 2 行: "大量" \rightarrow "不少"
p.28, 边注: "例如……上百亿个参数" \rightarrow "机器学习常涉及两类参数: 一类是算法的参数, 亦称"超参数", 数目常在 10 以内; 另一类是模型的参数, 数目可能很多, 例如……上百亿个参数. 两者调参方式相似, 均是产生多个模型之后基于某种评估方法来进行选择; 不同之处在于前者通常是由人工设定多个参数候选值后产生模型, 后者则是通过学习来产生多个候选模型 (例如神经网络在不同轮数停止训练)."
p.31, 倒数第 3 行: "Event" \rightarrow "Even"
p.256, 第 4 段: "固定住 α_i " \rightarrow "以 α_i 为初值"
p.256, 最后一段第 1 行: " $\mathbf{E}_i =$ " \rightarrow " $\mathbf{E}_i = \mathbf{X} -$ "
p.385, 式 (16.25) 和 (16.26): 两处" r_i " \rightarrow " R_i "
p.385, 式 (16.25) 下一行: "若改用……" \rightarrow "其中 R_i 表示第 i 条轨迹上

自状态 x 至结束的累积奖赏. 若改用……”

p.386, 式 (16.28) 下一行: ”始终为 1” \rightarrow ”对于 $a_i = \pi(x_i)$ 始终为 1”

p.386, 图 16.11, 第 4 步: 两处” $\pi(x)$ ” \rightarrow ” $\pi(x_i)$ ”

p.386, 图 16.11, 第 6 步的式子 \rightarrow ” $R = \frac{1}{T-t} \left(\sum_{i=t+1}^T r_i \right) \prod_{i=t+1}^{T-1} \frac{\mathbb{I}(a_i = \pi(x_i))}{p_i}$ ”

p.386, 图 16.11, 边注”计算修正的累积奖赏.” \rightarrow ”计算修正的累积奖赏. 连乘内下标大于上标的项取值为 1.”; 去掉边注”重要性采样系数”