**勘误修订 (Latex格式)**

**[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27次印刷, 2018年6月):**

* **p.5, 第2段倒数第3行: "3、2、2" --> "3、3、3"**
* **p.5, 第2段倒数第2行: "$4 \times 3 \times 3 + 1 = 37$" --> "$4 \times 4 \times 4 + 1 = 65$"**
* **p.6, 图1.2: 图中两处"清脆" --> "浊响"**
* **p.15, 第5行: "居功" --> "厥功"**
* **p.28, 第3段倒数第2行: "大量" --> "不少"**
* **p.28, 边注: "例如 ……上百亿个参数" --> "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.27, 式(2.1)：第一个"$\mapsto$" --> "$\to$", 第二个"$\mapsto$" --> "$=$"**
* **p.34, 图 2.4(b): 修订[文件](http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/mlbookfig24b.zip)**
* **p.36, 倒数第5行: "(TPR, FPR)" --> "(FPR, TPR)"**
* **p.39, 最后一行："$[-\infty,$" --> "$(-\infty,$"，"$, \infty]$" --> "$, \infty)$"**
* **p.42, 表2.5后第四行："$(k^2-1)/12$" --> "$(k^2-1)/12N$"**
* **p.26, 边注第2行: "2.6 节" --> "2.5 节"**
* **p.41, 式(2.33)上面一行: "正态分布, 且均值 …… 因此变量" --> "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)" --> "考虑到噪声不依赖于$f$, 由式(2.37)"**
* **p.31, 倒数第3行: "Event" --> "Even"**
* **p.38, 第6行: "$\epsilon^{m'}$" --> "${m \choose m'} \epsilon^{m'}$"**
* **p.42, 表2.5下面一段的第三行: "服从正态分布，其均值" --> "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"**
* **p.58, 倒数第二行："对率函数" --> "下面我们会看到, 对率回归求解的目标函数"**
* **p.60, 图3.3中："$y = \bm{w}^{\rm T}\bm{x}$, $y$" --> "投影方向~$\bm{w}$"**
* **p.80, 倒数第2行："算法4.2" --> "图 4.2 算法"**
* **p.63, 式(3.45)下面一行: "$N-1$个最大" --> "$d'$个最大非零"**
* **p.63, 式(3.45)下面第2行: "矩阵." --> "矩阵, $d'\le N-1$."; 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" --> "$d'$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" --> "$d'$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$0.5$" --> "阈值$1.5$"**
* **p.100, 图5.5, 左图最右边的 "阈值$0.5$" --> "阈值$-1.5$"**
* **p.100, 图5.5, 左图中间的"1  -1  -1  1" --> "1  1  -1  -1"**
* **p.56, 图3.1中，红色第一和第二个点的坐标互换**
* **p.55, 最后一行: 式子括号中的逗号改为分号**
* **p.59, 式(3.27)加边注: "考虑 $y\_i \in \{0, 1\}$"**
* **p.62, 第1行加边注: "$(\bm{\mu}\_0 - \bm{\mu}\_1)^{\rm T} \bm{w}$ 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”--> “硬滑”**
* **p.89, 图4.9, 中间分支底层: “硬挺”--> “硬滑”**
* **p.112, 图 5.14a: 修订[文件](http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/mlbookfig514a.zip)**
* **p.120, 第7行: "(1927 -- )" --> "(1927 -- 2016)"**
* **p.131, 图 6.5: 修订[文件](http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/mlbookfig65.zip)**
* **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中其他内容间的联系. 参见p.160."**
* **p.156, 倒数第7行："(7.23)" --> "(7.21)"**
* **p.156, 式(7.24)分母: "$N\_i$" --> "N \times N\_i"**
* **p.156, 式(7.25)下面一行: "其中 $N\_i$" --> "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: "$17+3$" --> "$17 + 3 \times 2$"**
* **p.156, 式(7.25)下面第4行: "0.350" --> "0.304"**
* **p.159, 倒数第9行、第10行中两处: "字节长度" --> "编码位数"**
* **p.159, 第一行加边注："一般需先对图剪枝, 仅保留有向图中~$x$, $y$, $\bf{z}$~及它们的祖先结点"**
* **p.160, 第1行、第4行中两处: "字节数" --> "编码位数"**
* **p.160, 第7行、第10行中两处: "字节" --> "编码位"**
* **p.174, 式(8.7)中两处、(8.8)中 7 处、边注中两处: "$f(x)$" --> "$f(\bm{x})$"**
* **p.187, 式(8.39)下面一行: "$\le$" --> "$\ge$"**
* **p.199, 式(9.12)：分母的 "$(\bm{\miu}\_i, \bm{\miu}\_j)$" --> "$(C\_i, C\_j)$"**
* **p.103, 最后一行的式子: 求和的"$q$" --> "$l$"**
* **p.114, 图5.15中, 卷积层 16@10x10 和 采样层 16@5x5 各去掉 8 个方块**
* **p.125, 式(6.18): "$y\_s$" --> "$1/y\_s$"**
* **p.136, 式(6.54): 右边最后一项中的四处 "$i$" --> "$j$"**
* **p.136, 式(6.54): 右边最后一项中最后的 "${\bm x}$" --> "${\bm x}\_i$"**
* **p.119, 第14行: "318--362" --> "533--536"**
* **p.125, 第3行: "减小" --> "增大"**
* **p.125, 第4行，第6行: "减幅" --> "增幅"**
* **p.125, 第5行: "减小" --> "增长"**
* **p.152, 第三个式子等号右端: "$0.375$" --> "$0.625$"**
* **p.153, 第3行: "$0.038$" --> "$0.063$"**
* **p.153, 第6行: "$0.038$" --> "$0.063$"**
* **p.160, 式(7.29)下面第2行: "需多少字节来描述$D$" --> "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.203, 图9.2下面一行: "$\bm{x}\_{27}$" --> "$\bm{x}\_{24}$"**
* **p.203, 图9.2下面第3行: "(0.532; 0.472)" --> "(0.478; 0.437)"**
* **p.203, 图9.2下面第5行: "0.166" --> "0.220"**
* **p.203, 图9.2下面第7行: 大括号中增加 "\bm{x}\_3", 去掉"\bm{x}\_{15}"**
* **p.203, 倒数第5行: 大括号中去掉 "\bm{x}\_3", 增加"\bm{x}\_{15}"**
* **p.203, 倒数第3行: "(0.473; 0.214)" --> "(0.493; 0.207)"**
* **p.203, 倒数第3行: "(0.623; 0.388)" --> "(0.602; 0.396)"**
* **p.204, 图9.3: 修订[文件](http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/mlbookfig93.zip)**
* **p.206, 9.4.3节前倒数第5行: "$c\_2$" --> "$c\_1$"**
* **p.206, 9.4.3节前倒数第2行: "(0.722; 0.442)" --> "(0.722; 0.447)"**
* **p.209, 式(9.38)上面一行: "样本" --> "混合成分"**
* **p.215, 图9.11第5步: "$j = 1, 2, \ldots, m$" --> "$j = i+1, \ldots, m$"**
* **p.230, 式(10.14)结尾: "." --> ","**
* **p.230, 式(10.14)下面一行开头顶格插入: "其中${\bf W} = (\bm{w}\_1, \bm{w}\_2, \ldots, \bm{w}\_d)$."**
* **p.230, 倒数第三行："方差" --> "协方差矩阵"**
* **p.230, 式(10.15)上面一行加边注: "严格来说, 协方差矩阵是~${1 \over {m-1}}\sum\nolimits\_{i=1}^m\bm{x}\_i\bm{x}\_i^{\rm T}$, 但前面的常数项在此不发生影响"**
* **p.303, 倒数第二行：去掉 "[Zhou et al., 2004]"**
* **p.231, 式(10.17): 两处"${\bf W}$"-->"${\bm w}\_i$", "$\lambda$" --> "$\lambda\_i$"**
* **p.231, 式(10.17)下面第二行: "${\bf W}$" --> "${\bf W}^\*$"**
* **p.231, 图10.5最后一行: "${\bf W}$" --> "${\bf W}^\*$"**
* **p.232, 第一行: "${\bf W}$" --> "${\bf W}^\*$"**
* **p.232, 式(10.19)前第二行: "${\bf W}$" --> "${\bf W} = (\bm{w}\_1, \bm{w}\_2, \ldots, \bm{w}\_d)$"**
* **p.232, 式(10.19)前第二行: "即PCA欲求解" --> "则对于 $\bm{w}\_j$, 由式(10.17)有"**
* **p.232, 式(10.19): 两处"${\bf W}$"-->"${\bm w}\_j$"; "$\lambda$" --> "$\lambda\_j$"**
* **p.233, 式(10.20): 三处"${\bf W}$"-->"${\bm w}\_j$"; 两处"$\lambda$"-->"$\lambda\_j$"; "${\bm \alpha}\_i$"-->"$\alpha\_i^j$"**
* **p.233, 式(10.20)下一行: "${\bm \alpha}\_i$"-->"$\alpha\_i^j$"; "$\lambda$"-->"$\lambda\_j$"; "${\bf W}$"-->"${\bm w}\_j$"**
* **p.233, 式(10.20)下一行: ". 假定" --> "是${\bm \alpha}\_i$的第$j$个分量. 假定"**
* **p.233, 式(10.21): 两处"${\bf W}$"-->"${\bm w}\_j$"; "$\lambda$"-->"$\lambda\_j$"**
* **p.233, 式(10.22): "${\bf W}$"-->"${\bm w}\_j$"; "${\bm \alpha}\_i$"-->"$\alpha\_i^j$"**
* **p.233, 式(10.24): 两处"${\bf A}$"-->"${\bm \alpha}^j$"; "$\lambda$"-->"$\lambda\_j$"**
* **p.233, 式(10.24)下面一行: "${\bf A} = ({\bm \alpha}\_1; {\bm \alpha}\_2; \ldots; {\bm \alpha}\_m)$" --> "${\bm \alpha}^j = (\alpha\_1^j; \alpha\_2^j; \ldots; \alpha\_m^j)$"**
* **p.233, 式(10.25)下面一行: 去掉 ", $\alpha\_i^j$ 是 ${\bm \alpha}\_i$ 的第 $j$ 个分量"**
* **p.239, 式(10.39)第二行式子: 去掉上标 "$2$"**
* **p.244, 第13行: "Locally" --> "Nonlinear dimensionality reduction by locally"**
* **p.244, 第14行: "2316" --> "2326"**
* **p.249, 式(11.2): "$i=1$" --> "$k=1$"**
* **p.253, 倒数第5行: "[Boyd and Vandenberghe, 2004]" --> "[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.'' \textit{Mutiscale Modeling & Simulation}, 4(4):1168--1200."**
* **p.277, 式(12.29): "$E(h) - \hat{E}(h)$" --> "$\left| E(h) - \hat{E}(h) \right|$"**
* **p.299, 式(13.9)后第三段第2行: "关于 $D\_u$" --> "涉及 $C\_u$"**
* **p.301, 式(13.12)的下一行: "$({\bm f}\_l^{\rm T}\,{\bm f}\_u^{\rm T})^{\rm T}$" --> "$({\bm f}\_l^{\rm T}; {\bm f}\_u^{\rm T})$"**
* **p.256, 第4段: "固定住${\bf \alpha}\_i$" --> "以${\bf \alpha}\_i$为初值"**
* **p.256, 最后一段第1行: "${\bf E}\_i =$" --> "${\bf E}\_i = {\bf X} - $"**
* **p.304, 第一行："当" --> "考虑到有标记样本通常很少而未标记样本很多, 为缓解过拟合, 可在式(13.21)中引入针对未标记样本的~L$\_2$~范数项~$\mu \sum\_{i=l+1}^{l+u}\|\mathbf{F}\_{i}\|^{2}$, 在"; 同时插入边注: "参见~11.4~节"**
* **p.320, 第8行："其余~$n-2$" --> "此前~$t-2$"**
* **p.384, 图16.10, 步骤9: "$\pi(x, a)$" --> "$\pi(x)$"**
* **p.385, 式(16.25)和(16.26): 两处"$r\_i$" --> "$R\_i$"**
* **p.385, 式(16.25)下一行: "若改用……" --> "其中$R\_i$表示第$i$条轨迹上自状态$x$至结束的累积奖赏. 若改用……"**
* **p.386, 式(16.28)下一行: "始终为1" --> "对于$a\_i=\pi(x\_i)$始终为1"**
* **p.386, 图16.11, 第4步: 两处 "$\pi(x)$" --> "$\pi(x\_i)$"**
* **p.386, 图16.11, 第6步的式子 --> "$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}$"**
* **p.386, 图16.11, 边注"计算修正的累积奖赏." --> "计算修正的累积奖赏. 连乘内下标大于上标的项取值为1."; 去掉边注"重要性采样系数."**
* **p.388, 图16.13, 步骤4: "$\pi^{\epsilon}(x)$" --> "$a = \pi^{\epsilon}(x)$"**
* **p.388, 图16.13, 步骤8: 去掉", $a = a'$"**
* **p.417, 第3段第1行: "通往人工智能的途径" --> "一种人工智能途径"**
* **p.372, 图16.2: 从"s=健康"到"s=溢水"的 "r=1" --> "r=-1"**
* **p.376, 图16.5的边注: "第 4 行中式(16.4)的参数" --> "该参数在第4行使用"**
* **p.385, 第二行: "在使用策略时并不需要$\epsilon-$贪心" --> "而不是为了最终使用"**
* **p.387, 倒数第二行: "$\epsilon-$贪心策略, 而执行(第5行)的是原始策略" --> "原始策略, 而执行(第4行)的是$\epsilon-$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.399, 式(A.9): "$A\_{1 \sigma n}$" --> "$A\_{n \sigma n}$"**
* **p.400, 第1行: "(1,4,3,2)" --> "(3,1,2)"**
* **p.402, 式(A.32)最后一行的式子中: "$2{\mathbf A}$" --> "$2{\mathbf A}^{\rm T}$"**
* **p.402, 式(A.32)加边注: "机器学习中 $\bf W$ 通常是对称矩阵"**
* **p.404, 式(B.3)最后一行的式子 --> "$\lambda g({\bm x})=0$"**