

序号	页码	原文	更正
1	12	<p>对 w_j 求偏导数并令其为 0, 可得</p> $w_j = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^{j+1}},$ <p>$j = 0, 1, 2, \dots, M$ 于是求得拟合多项式系数 $w_0^*, w_1^*, \dots, w_M^*$.</p>	<p>这一问题可用最小二乘法求得拟合多项式系数的唯一解, 记作 $w_0^*, w_1^*, \dots, w_M^*$. 求解过程这里不予叙述, 读者可参阅有关材料.</p>
2	77	$F(-x + \mu) - \frac{1}{2} = -F(x - \mu) + \frac{1}{2}$	$F(-x + \mu) - \frac{1}{2} = -F(x + \mu) + \frac{1}{2}$
3	161	$= \log P(Z Y, \theta^{(i+1)}) = 0 \quad (9.23)$	$= \log \left[\sum_Z P(Z Y, \theta^{(i+1)}) \right] = 0 \quad (9.23)$
4	198	$W_i(y_{i-1}, y_i x) = \sum_{k=1}^K w_k f_k(y_{i-1}, y_i, x, i)$ <p>(11.23)</p>	$W_i(y_{i-1}, y_i x) = \sum_{k=1}^K w_k f_k(y_{i-1}, y_i, x, i)$ <p>(11.23)</p>
5	14	<p>第 13,14 行 可以假设复杂的模型有较大的先验概率, 简单的模型有较小的先验概率.</p>	<p>第 13,14 行 可以假设复杂的模型有较小的先验概率, 简单的模型有较大的先验概率.</p>
6	141	(0.0715,0.0715,0.0715,0.0715,0.0715,0.0715,0.1666,0.1666,0.1666,0.0715)	(0.07143,0.07143,0.07143,0.07143,0.07143,0.07143,0.16667,0.16667,0.16667,0.07143)
7	43	第 8 行 (参阅图 3.8)	第 8 行 (参阅图 3.5)
8	119	<p>式(7.73)</p> $f * g \bullet \sum_{i=1}^m \sum_{j=1}^l \alpha_i \beta_j K(x_i, z_j)$	<p>式(7.73)</p> $f * g = \sum_{i=1}^m \sum_{j=1}^l \alpha_i \beta_j K(x_i, z_j)$
9	163	式(9.28)	该编号移到 164 页第一公式后

10	222	倒数第 9 行 式(B.23)	倒数第 9 行 式(B.24)
11	35	表 2.2 第 6 列 4 x_3 2 0 2 0	表 2.2 第 6 列 4 x_3 1 0 3 -2
12	115	第 4 行 损失函数 $[y_i(wx_i + b)]_+$	第 4 行 损失函数 $[-y_i(wx_i + b)]_+$
13	63	<p>5.2.3 信息增益比</p> <p>信息增益值的大小是相对训练数据集而言的，并没有绝对意义。在分类问题困难时，也就是训练数据集的经验熵大的时候，信息增益值会偏大。反之，信息增益值会偏小。使用信息增益比(information gain ratio) 可以对这一问题进行校正。这是特征选择的另一准则。</p> <p>定义 5.3 (信息增益比) 特征 A 对训练数据集 D 的信息增益比</p> <p>$g_R(D, A)$ 定义为其信息增益 $g(D, A)$ 与训练数据集 D 的经验熵 $H(D)$ 之比</p> $g_R(D, A) = \frac{g(D, A)}{H(D)} \quad (5.10)$	<p>5.2.3 信息增益比</p> <p>以信息增益作为划分训练数据集的特征，存在偏向于选择取值较多的特征的问题。使用信息增益比(information gain ratio)可以对这一问题进行校正。这是特征选择的另一准则。</p> <p>定义 5.3 (信息增益比) 特征 A 对训练数据集 D 的信息增益比 $g_R(D, A)$ 定义为其信息增益 $g(D, A)$ 与训练数据集 D 关于特征 A 的值的熵 $H_A(D)$ 之比，即</p> $g_R(D, A) = \frac{g(D, A)}{H_A(D)} \quad (5.10)$ <p>其中</p> $H_A(D) = - \sum_{i=1}^n \frac{ D_i }{ D } \log_2 \frac{ D_i }{ D }$ <p>n 是特征 A 取值的个数。</p>

14	114	<p>证明 可将最优化问题(7.63)写成问题 (7.60) ~ (7.62). 令</p> $1 - y_i(w \square x_i + b) = \xi_i,$ $\xi_i \geq 0$ <p>(7.64)</p> <p>则 $y_i(w \square x_i + b) \geq 1$. 于是 w, b, ξ_i 满足约束条件(7.61)~(7.62). 由(7.64)有 $[1 - y_i(w \square x_i + b)]_+ = [\xi_i]_+ = \xi_i$, 所以最优化问题(7.63)可写成</p>	<p>证明 可将最优化问题(7.63)写成问题 (7.60) ~ (7.62). 令</p> $[1 - y_i(w \square x_i + b)]_+ = \xi_i$ <p>(7.64)</p> <p>则 $\xi_i \geq 0$, 式 (7.62) 成立. 由式 (7.64),</p> <p>当 $1 - y_i(w \square x_i + b) > 0$ 时, 有</p> $y_i(w \square x_i + b) = 1 - \xi_i; \quad \text{当}$ <p>$1 - y_i(w \square x_i + b) \leq 0$ 时, $\xi_i = 0$, 有</p> $y_i(w \square x_i + b) \geq 1 - \xi_i. \quad \text{故式 (7.61) 成立.}$ <p>于是 w, b, ξ_i 满足约束条件(7.61)~(7.62). 所以最优化问题(7.63)可写成</p>
15	第 159 页	利用 Jensen 不等式(Jensen inequality)	<p>利用 Jensen 不等式(Jensen inequality) ①</p> <p>脚注① 这里用到的是</p> $\log \sum_j \lambda_j y_j \geq \sum_j \lambda_j \log y_j, \quad \text{其中}$ $\lambda_j \geq 0, \quad \sum_j \lambda_j = 1.$
16	第 163 页	<p>(原稿)</p> <p>那么, 完全数据的对数似然函数为</p> $\log P(y, \gamma \theta) = \sum_{k=1}^K n_k \log \alpha_k + \sum_{j=1}^N \gamma_{jk} [\log(\frac{1}{\sqrt{2\pi}}) - \log \sigma_k - \frac{1}{2\sigma_k^2} (y_j - \mu_k)^2]$	
		<p>(修改稿) 加大括号</p> <p>那么, 完全数据的对数似然函数为</p> $\log P(y, \gamma \theta) = \sum_{k=1}^K \{n_k \log \alpha_k + \sum_{j=1}^N \gamma_{jk} [\log(\frac{1}{\sqrt{2\pi}}) - \log \sigma_k - \frac{1}{2\sigma_k^2} (y_j - \mu_k)^2]\}$	
	第 163 页	<p>(原稿)</p> <p>2. EM 算法的 E 步: 确定 Q 函数.</p> $Q(\theta, \theta^{(i)}) = E[\log P(y, \gamma \theta) y, \theta^{(i)}]$ $= E\{\sum_{k=1}^K n_k \log \alpha_k + \sum_{j=1}^N \gamma_{jk} [\log(\frac{1}{\sqrt{2\pi}}) - \log \sigma_k - \frac{1}{2\sigma_k^2} (y_j - \mu_k)^2]\}$	

		<p>(修改稿) 加大括号</p> <p>2. EM 算法的 E 步: 确定 Q 函数.</p> $Q(\theta, \theta^{(i)}) = E[\log P(y, \gamma \theta) y, \theta^{(i)}]$ $= E\left\{\sum_{k=1}^K \{n_k \log \alpha_k + \sum_{j=1}^N \gamma_{jk} [\log(\frac{1}{\sqrt{2\pi}}) - \log \sigma_k - \frac{1}{2\sigma_k^2} (y_j - \mu_k)^2]\}\right\}$
	第 164 页	<p>(原稿)</p> $Q(\theta, \theta^{(i)}) = \sum_{k=1}^K n_k \log \alpha_k + \sum_{k=1}^N \hat{\gamma}_{jk} [\log(\frac{1}{\sqrt{2\pi}}) - \log \sigma_k - \frac{1}{2\sigma_k^2} (y_j - \mu_k)^2] \quad (9.29)$ <p>(修改稿) 加上大括号, 第二个和号的求和指标由 k 改为 j</p> $Q(\theta, \theta^{(i)}) = \sum_{k=1}^K \{n_k \log \alpha_k + \sum_{j=1}^N \hat{\gamma}_{jk} [\log(\frac{1}{\sqrt{2\pi}}) - \log \sigma_k - \frac{1}{2\sigma_k^2} (y_j - \mu_k)^2]\}$ <p>(9.29)</p>
17	第 73 页 算 法 5.7	<p>(原稿)</p> <p>(4) 自上而下地访问内部结点 t, 如果有 $g(t) = \alpha$, 进行剪枝, 并对叶结点 t 以多数表决法决定其类, 得到树 T</p> <p>(修改稿)</p> <p>(4) 对 $g(t) = \alpha$ 的内部结点 t 进行剪枝, 并对叶结点 t 以多数表决法决定其类, 得到树 T</p>
18	第 73 页 算 法 5.7	<p>(原稿)</p> <p>(6) 如果 T 不是由根结点单独构成的树, 则回到步骤(4).</p> <p>(修改稿)</p> <p>(6) 如果 T_k 不是由根结点及两个叶结点构成的树, 则回到步骤(2); 否则令 $T_k = T_n$</p>
19	第 205 页 算 法 11.2	<p>(原稿)</p> <p>(6) 计算 $g_{k+1} = g(w^{(k+1)})$, 若 $g_k = 0$, 则停止计算;</p> <p>(修改稿)</p> <p>(6) 计算 $g_{k+1} = g(w^{(k+1)})$, 若 $g_{k+1} = 0$, 则停止计算;</p>
20	第 200 页	<p>(原稿)</p> $\beta_i(y_i x) = M_i(y_i, y_{i+1} x) \beta_{i-1}(y_{i+1} x) \quad (11.30)$ <p>(修改稿) $\beta_i(y_i x) = [M_i(y_i, y_{i+1} x)] \beta_{i+1}(y_{i+1} x) \quad (11.30)$</p>

21	第 196 页	(原稿) 倒数第 11 行 $\lambda_2 = 0.5$
		(修改稿) $\lambda_2 = 0.6$
22	第 208 页	(原稿) 第 9 行 $\delta_2(1) = \max\{1 + \lambda_2 t_2, 0.5 + \lambda_4 t_4\} = 1.6 \quad \Psi_2(1) = 1$
		(修改稿) $\delta_2(1) = \max\{1 + \lambda_2 t_2 + \mu_3 s_3, 0.5 + \lambda_4 t_4 + \mu_3 s_3\} = 2.4 \quad \Psi_2(1) = 1$
		(原稿) 第 12 行 $\delta_3(1) = \max\{1.6 + \mu_5 s_5, 2.5 + \lambda_3 t_3 + \mu_3 s_3\} = 4.3 \quad \Psi_3(1) = 2$
		(修改稿) $\delta_3(1) = \max\{2.4 + \mu_5 s_5, 2.5 + \lambda_3 t_3 + \mu_3 s_3\} = 4.3 \quad \Psi_3(1) = 2$
		(原稿) 第 13 行 $\delta_3(2) = \max\{1.6 + \lambda_1 t_1 + \mu_4 s_4, 2.5 + \lambda_5 t_5 + \mu_4 s_4\} = 3.2 \quad \Psi_3(2) = 1$
		(修改稿) $\delta_3(2) = \max\{2.4 + \lambda_1 t_1 + \mu_4 s_4, 2.5 + \lambda_5 t_5 + \mu_4 s_4\} = 3.9 \quad \Psi_3(2) = 1$
23	第 156 页式	(原稿) 式(9.5)左端 $\mu^{(i+1)}$
		(修改稿) $\mu_j^{(i+1)}$
24	第 198 页	(原稿) 这样, 给定观测序列 x , 标记序列 y 的非规范化概率可以通过 $n+1$ 个矩阵的乘积 $\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i x)$ 表示,
		(修改稿) 这样, 给定观测序列 x , 相应标记序列 y 的非规范化概率可以通过该序列 $n+1$ 个矩阵适当元素的乘积 $\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i x)$ 表示.

25	第 200 页	(原稿) $\alpha_i^T(y_i x) = \alpha_{i-1}^T(y_{i-1} x) M_i(y_{i-1}, y_i x), \quad i = 1, 2, \dots, n+1 \quad (11.27)$ (原稿) $\beta_i(y_i x) = M_i(y_i, y_{i+1} x) \beta_{i+1}(y_{i+1} x) \quad (11.30)$
		(修改稿) $\alpha_i^T(y_i x) = \alpha_{i-1}^T(y_{i-1} x) [M_i(y_{i-1}, y_i x)], \quad i = 1, 2, \dots, n+1 \quad (11.27)$ (修改稿) $\beta_i(y_i x) = [M_i(y_i, y_{i+1} x)] \beta_{i+1}(y_{i+1} x) \quad (11.30)$
26	第 29 页, 倒数第 2 行公式	(原稿) $\min_{w,b} L(w,b) = - \sum_{x_i \in M} y_i (w \bullet x + b) \quad (\text{公式右边的 } x \text{ 少了下标 } i)$
		(修改稿) $\min_{w,b} L(w,b) = - \sum_{x_i \in M} y_i (w \bullet x_i + b)$
27	第 104 页第 8 行	(原稿) $\nabla_b L(w,b,\alpha) = \sum_{i=1}^N \alpha_i y_i = 0 \quad (\text{公式右边少个负号})$
		(修改稿) $\nabla_b L(w,b,\alpha) = - \sum_{i=1}^N \alpha_i y_i = 0$
28	第 74 页中间	(原稿) (2) 样本集合 D 对特征 A 的信息增益比 (C4.5) $g_R(D,A) = \frac{g(D,A)}{H(D)}$ 其中, $g(D,A)$ 是信息增益, $H(D)$ 是 D 的熵.
		(修改稿) (2) 样本集合 D 对特征 A 的信息增益比 (C4.5) $g_R(D,A) = \frac{g(D,A)}{H_A(D)}$ 其中, $g(D,A)$ 是信息增益, $H_A(D)$ 是 D 关于特征 A 的值的熵.
29	第 221	(原稿)

	页 第 3 行	这是因为搜索方向是 $p_k = -\lambda g_k$ ，由式(B.8)有
		(修改稿)
		这是因为搜索方向是 $p_k = -H_k^{-1} g_k$ ，由式(B.8)有 (修改搜索方向表达式)
30	第 109 页 式 (7.34) 下第 2 行	(原稿) 可以证明 w 的解是唯一的，但 b 的解不唯一， b 的解存在于一个区间[11].
		(修改稿) 可以证明 w 的解是唯一的，但 b 的解可能不唯一，而是存在于一个区间[11].
	第 112 最后 2 行 至 113 页 第 1 行	(原稿) 步骤 (2) 中，对任一适合条件 $0 < \alpha_j^* < C$ 的 α_j^* ，按式 (7.51) 都可求出 b^* ， 但是由于原始问题 (7.32) ~ (7.34) 对 b 的解并不唯一[11]，所以实际计算时 可以取在所有符合条件的样本点上的平均值。
		(修改稿) 步骤 (2) 中，对任一适合条件 $0 < \alpha_j^* < C$ 的 α_j^* ，按式 (7.51) 都可求出 b^* 。从 理论上，原始问题 (7.32) ~ (7.34) 对 b 的解可能不唯一[11]，然而在实际应 用中，往往只会出现算法叙述的情况。
	第 132 页中间	(原稿) 线性可分支持向量机的解 w^* 唯一但 b^* 不唯一。
		(修改稿) 线性支持向量机的解 w^* 唯一但 b^* 不一定唯一。
31	第 179 页 第 6 行	(原稿) 此式当 $t=1$ 和 $t=T-1$ 时分别为式 (10.17) 和式 (10.21)。
		(修改稿) 此式当 $t=1$ 和 $t=T-1$ 时分别为式 (10.21) 和 (10.17)。
32	第 119 页 式 (7.72)	(原稿) $g(\bullet) = \sum_{i=1}^l \beta_j K(\bullet, z_j) \quad (7.72)$
		(修改稿) $g(\bullet) = \sum_{j=1}^l \beta_j K(\bullet, z_j) \quad (7.72)$
33	第 227 页 至 228 页	(原稿) $\nabla_{\alpha} L(x^*, \alpha^*, \beta^*) = 0 \quad (C.22)$ $\nabla_{\beta} L(x^*, \alpha^*, \beta^*) = 0 \quad (C.23)$

		$\alpha_i^* c_i(x^*) = 0, \quad i = 1, 2, \dots, k \quad (\text{C.24})$ $c_i(x^*) \leq 0, \quad i = 1, 2, \dots, k \quad (\text{C.25})$ $\alpha_i^* \geq 0, \quad i = 1, 2, \dots, k \quad (\text{C.26})$ $h_j(x^*) = 0, \quad j = 1, 2, \dots, l \quad (\text{C.27})$ <p>特别指出，式（C.24）称为 KKT 的对偶互补条件.</p>
		<p>（修改稿）</p> $\alpha_i^* c_i(x^*) = 0, \quad i = 1, 2, \dots, k \quad (\text{C.22})$ $c_i(x^*) \leq 0, \quad i = 1, 2, \dots, k \quad (\text{C.23})$ $\alpha_i^* \geq 0, \quad i = 1, 2, \dots, k \quad (\text{C.24})$ $h_j(x^*) = 0, \quad j = 1, 2, \dots, l \quad (\text{C.25})$ <p>特别指出，式（C.22）称为 KKT 的对偶互补条件. （删除原式（C.22）（C.23），后面 4 式编号改为（C.22）至（C.25））</p>
34	44 页图 3.5	<p>（原稿）</p> <p>B 点和 C 点各在一直线附近</p>
		<p>（修改稿）</p> <p>B 点和 C 点都在其最近直线上</p>
35	第 140 页 第 4 行	<p>（原稿）</p> <p>两相比较，误分类样本的权值被放大 $e^{2\alpha_m} = \frac{e_m}{1-e_m}$ 倍.</p>
		<p>（修改稿）</p> <p>两相比较，由式(8.2)知误分类样本的权值被放大 $e^{2\alpha_m} = \frac{1-e_m}{e_m}$ 倍. （分子分母颠倒）.</p>
36	第 138 页 式（8.1）， 第 139 页 式（8.8）	<p>（原稿）</p> $e_m = P(G_m(x_i) \neq y_i) = \sum_{i=1}^N w_{mi} I(G_m(x_i) \neq y_i) \quad (8.1)$ $e_m = P(G_m(x_i) \neq y_i) = \sum_{G_m(x_i) \neq y_i} w_{mi} \quad (8.8)$
		<p>（修改稿）</p>

		$e_m = \sum_{i=1}^N P(G_m(x_i) \neq y_i) = \sum_{i=1}^N w_{mi} I(G_m(x_i) \neq y_i) \quad (8.1)$ $e_m = \sum_{i=1}^N P(G_m(x_i) \neq y_i) = \sum_{G_m(x_i) \neq y_i} w_{mi} \quad (8.8)$ <p>(添加一个求和号)</p>
--	--	---