| 序号 | 页码 | 原文 | 更正 |
|---|---|---|---|
| 1 | 12 | 对 $w_j$ 求偏导数并令其为0，可得 $$w_j=\frac{\sum_{i=1}^{N}x_i y_i}{\sum_{i=1}^{N}x_i^{j+1}},$$ $j=0,1,2,\cdots,M$ 于是求得拟合多项式系数 $w_0^*,w_1^*,\cdots,w_M^*$. | 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 $w_0^*,w_1^*,\cdots,w_M^*$. 求解过程这里不予叙述，读者可参阅有关材料. |
| 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 | $=\log P(Z\|Y,\theta^{(i+1)})=0$ (9.23) | $=\log\left[\sum_Z P(Z\|Y,\theta^{(i+1)})\right]=0$ (9.23) |
| 4 | 198 | $W_i(y_{i-1},y_i\|x)=\sum_{i=1}^{K}w_k f_k(y_{i-1},y_i,x,i)$ (11.23) | $W_i(y_{i-1},y_i\|x)=\sum_{k=1}^{K}w_k f_k(y_{i-1},y_i,x,i)$ (11.23) |
| 5 | 14 | 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. | 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. |
| 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 | 式(7.73) $f*g\bullet\sum_{i=1}^{m}\sum_{j=1}^{l}\alpha_i\beta_j K(x_i,z_j)$ | 式(7.73) $f*g=\sum_{i=1}^{m}\sum_{j=1}^{l}\alpha_i\beta_j K(x_i,z_j)$ |
| 9 | 163 | 式(9.28) | *该编号移到164页第一公式后* |

| | | | |
|---|---|---|---|
| 10 | 222 | 倒数第 9 行<br>式(B.23) | 倒数第 9 行<br>式(B.24) |
| 11 | 35 | 表 2.2 第 6 列<br>4<br>$x_3$<br>2<br>0<br>2<br>0 | 表 2.2 第 6 列<br>4<br>$x_3$<br>1<br>0<br>3<br>-2 |
| 12 | 115 | 第 4 行<br>损失函数$[y_i(wx_i+b)]_+$ | 第 4 行<br>损失函数$[-y_i(wx_i+b)]_+$ |
| 13 | 63 | 5.2.3 信息增益比<br>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．<br>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 $D$ 的经验熵 $H(D)$ 之比<br>$$g_R(D,A)=\frac{g(D,A)}{H(D)} \quad (5.10)$$ | 5.2.3 信息增益比<br>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．<br>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<br>$$g_R(D,A)=\frac{g(D,A)}{H_A(D)} \quad (5.10)$$<br>其中<br>$$H_A(D)=-\sum_{i=1}^{n}\frac{\vert D_i\vert}{\vert D\vert}\log_2\frac{\vert D_i\vert}{\vert D\vert}$$<br>$n$ 是特征 $A$ 取值的个数． |

<table>
<tr><td>14</td><td>114</td><td>证明　可将最优化问题(7.63)写成问题（7.60）～（7.62）．令<br>$$1-y_i(w\cdot x_i+b)=\xi_i，$$<br>$\xi_i\ge 0$<br>（7.64）<br>则 $y_i(w\cdot x_i+b)\ge 1$．于是 $w,b,\xi_i$ 满足约束条件(7.61)～(7.62)．由(7.64)有 $[1-y_i(w\cdot x_i+b)]_+=[\xi_i]_+=\xi_i$，所以最优化问题(7.63)可写成</td><td>证明　可将最优化问题(7.63)写成问题（7.60）～（7.62）．令<br>$$[1-y_i(w\cdot x_i+b)]_+=\xi_i$$<br>（7.64）<br>则 $\xi_i\ge 0$，式（7.62）成立．由式（7.64），当 $1-y_i(w\cdot x_i+b)>0$ 时，有 $y_i(w\cdot x_i+b)=1-\xi_i$；当 $1-y_i(w\cdot x_i+b)\le 0$ 时，$\xi_i=0$，有 $y_i(w\cdot x_i+b)\ge 1-\xi_i$．故式（7.61）成立．于是 $w,b,\xi_i$ 满足约束条件(7.61)～(7.62)．所以最优化问题(7.63)可写成</td></tr>
<tr><td>15</td><td>第159页</td><td>利用 Jensen 不等式(Jensen inequality)</td><td>利用 Jensen 不等式(Jensen inequality) ①<br>脚注①　这里用到的是<br>$\log\sum_j \lambda_j y_j \ge \sum_j \lambda_j \log y_j$，其中 $\lambda_j\ge 0$，$\sum_j \lambda_j=1$．</td></tr>
<tr><td rowspan="3">16</td><td rowspan="2">第163页</td><td colspan="2">(原稿)<br>那么，完全数据的对数似然函数为<br>$$\log P(y,\gamma\mid\theta)=\sum_{k=1}^{K} n_k\log\alpha_k+\sum_{j=1}^{N}\gamma_{jk}[\log(\frac{1}{\sqrt{2\pi}})-\log\sigma_k-\frac{1}{2\sigma_k^2}(y_j-\mu_k)^2]$$</td></tr>
<tr><td colspan="2">(修改稿)　加大括号<br>那么，完全数据的对数似然函数为<br>$$\log P(y,\gamma\mid\theta)=\sum_{k=1}^{K}\{n_k\log\alpha_k+\sum_{j=1}^{N}\gamma_{jk}[\log(\frac{1}{\sqrt{2\pi}})-\log\sigma_k-\frac{1}{2\sigma_k^2}(y_j-\mu_k)^2]\}$$</td></tr>
<tr><td>第163页</td><td colspan="2">(原稿)<br>2．EM 算法的 E 步：确定 $Q$ 函数．<br>$$Q(\theta,\theta^{(i)})=E[\log P(y,\gamma\mid\theta)\mid y,\theta^{(i)}]$$<br>$$=E\{\sum_{k=1}^{K} n_k\log\alpha_k+\sum_{j=1}^{N}\gamma_{jk}[\log(\frac{1}{\sqrt{2\pi}})-\log\sigma_k-\frac{1}{2\sigma_k^2}(y_j-\mu_k)^2]\}$$</td></tr>
</table>

| | | |
|---|---|---|
| | | (修改稿) 加大括号<br>2．EM 算法的 E 步：确定 $Q$ 函数.<br>$Q(\theta,\theta^{(i)}) = E[\log P(y,\gamma \mid \theta) \mid y,\theta^{(i)}]$<br>$= E\{\sum_{k=1}^{K}\{n_k \log\alpha_k + \sum_{j=1}^{N}\gamma_{jk}[\log(\frac{1}{\sqrt{2\pi}}) - \log\sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2]\}\}$ |
| | 第164页 | (原稿)<br>$Q(\theta,\theta^{(i)}) = \sum_{k=1}^{K} n_k \log\alpha_k + \sum_{k=1}^{N}\hat{\gamma}_{jk}[\log(\frac{1}{\sqrt{2\pi}}) - \log\sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2]$ (9.29) |
| | | (修改稿) 加上大括号，第二个和号的求和指标由 $k$ 改为 $j$<br>$Q(\theta,\theta^{(i)}) = \sum_{k=1}^{K}\{n_k \log\alpha_k + \sum_{j=1}^{N}\hat{\gamma}_{jk}[\log(\frac{1}{\sqrt{2\pi}}) - \log\sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2]\}$<br>(9.29) |
| 17 | 第 73 页 算 法 5.7 | (原稿)<br>（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 $T$ |
| | | (修改稿)<br>(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 $T$ |
| 18 | 第 73 页 算 法 5.7 | (原稿)<br>(6) 如果 $T$ 不是由根结点单独构成的树，则回到步骤(4). |
| | | (修改稿)<br>(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 $T_k = T_n$ |
| 19 | 第 205 页 算 法 11.2 | (原稿)<br>（6）计算 $g_{k+1} = g(w^{(k+1)})$，若 $g_k = 0$，则停止计算； |
| | | (修改稿)<br>（6）计算 $g_{k+1} = g(w^{(k+1)})$，若 $g_{k+1} = 0$，则停止计算； |
| 20 | 第 200 页 | (原稿)<br>$\beta_i(y_i \mid x) = M_i(y_i, y_{i+1} \mid x)\beta_{i-1}(y_{i+1} \mid x)$ (11.30) |
| | | (修改稿)<br>$\beta_i(y_i \mid x) = M_{i+1}(y_i, y_{i+1} \mid x)\beta_{i+1}(y_{i+1} \mid x)$ (11.30) |

| | | |
|---|---|---|
| 21 | 第196页 | (原稿)<br>倒数第 11 行 $\lambda_2 = 0.5$ |
| | | (修改稿)<br>$\lambda_2 = 0.6$ |
| 22 | 第208页 | (原稿)<br>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)<br>$\delta_2(1) = \max\{1 + \lambda_2 t_2 + \mu_3 s_3, 0.5 + \lambda_4 t_4 + \mu_3 s_3\} = 2.4 \quad \Psi_2(1) = 1$ |
| | | (原稿)<br>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)<br>$\delta_3(1) = \max\{2.4 + \mu_5 s_5, 2.5 + \lambda_3 t_3 + \mu_3 s_3\} = 4.3 \quad \Psi_3(1) = 2$ |
| | | (原稿)<br>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)<br>$\delta_3(2) = \max\{2.4 + \lambda_1 t_1 + \mu_4 s_4, 2.5 + \lambda_5 t_5 + \mu_4 s_4\} = 3.9 \quad \Psi_3(2) = 1$ |
| 23 | 第156页式 | (原稿)<br>式(9.5)左端 $\mu^{(i+1)}$ |
| | | (修改稿)<br>$\mu_j^{(i+1)}$ |
| 24 | 第198页 | (原稿)<br>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 $\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i \mid x)$ 表示， |
| | | (修改稿)<br>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 $\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i \mid x)$ 表示. |

<table>
<tr><td rowspan="2">25</td><td rowspan="2">第 200 页</td><td>(原稿)<br>$$\alpha_i^T(y_i \mid x) = \alpha_{i-1}^T(y_{i-1} \mid x) M_i(y_{i-1}, y_i \mid x), \quad i = 1, 2, \cdots, n+1 \qquad (11.27)$$<br>(原稿)<br>$$\beta_i(y_i \mid x) = M_i(y_i, y_{i+1} \mid x)\beta_{i+1}(y_{i+1} \mid x) \qquad (11.30)$$</td></tr>
<tr><td>(修改稿)<br>$$\alpha_i^T(y_i \mid x) = \alpha_{i-1}^T(y_{i-1} \mid x)[M_i(y_{i-1}, y_i \mid x)], \quad i = 1, 2, \cdots, n+1 \qquad (11.27)$$<br>(修改稿)<br>$$\beta_i(y_i \mid x) = [M_i(y_i, y_{i+1} \mid x)]\beta_{i+1}(y_{i+1} \mid x) \qquad (11.30)$$</td></tr>
<tr><td rowspan="2">26</td><td rowspan="2">第 29 页，倒数第 2 行公式</td><td>（原稿）（公式右边的 $x$ 少了下标 $i$ ）<br>$$\min_{w,b} L(w,b) = -\sum_{x_i \in M} y_i(w \bullet x + b)$$</td></tr>
<tr><td>（修改稿）<br>$$\min_{w,b} L(w,b) = -\sum_{x_i \in M} y_i(w \bullet x_i + b)$$</td></tr>
</table>