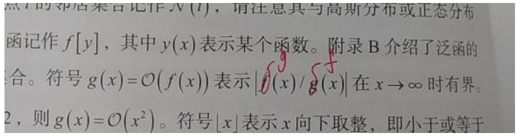
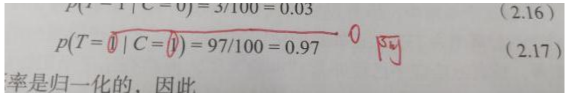
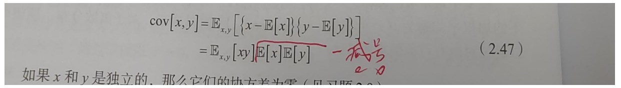
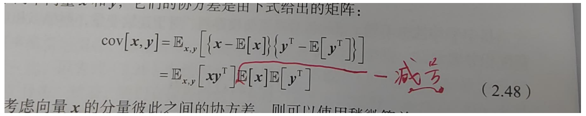
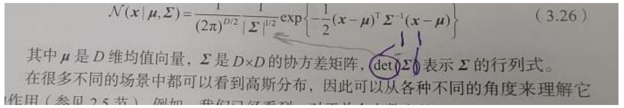
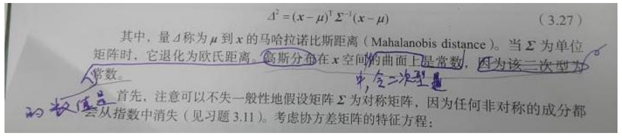
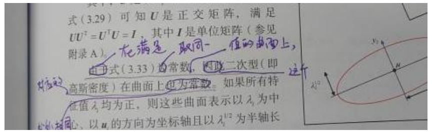
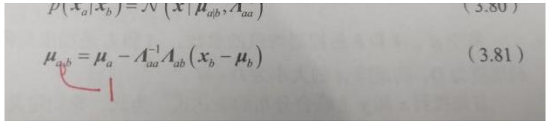
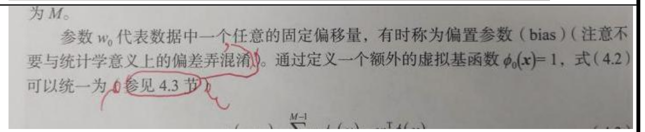
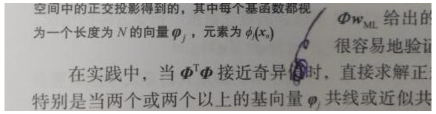
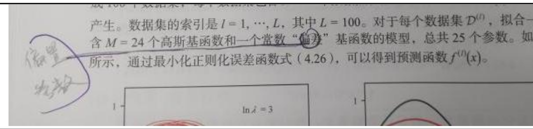
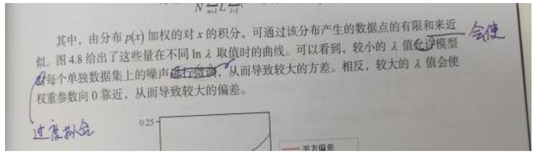
《深度学习：基础与概念》勘误表

2025 年 6 月第 2 次印刷勘误

序号	章	页	行	误	正	备注
1	前言	4	7	$f(x)/g(x)$	$g(x)/f(x)$	 函记作 $f[y]$，其中 $y(x)$ 表示某个函数。附录 B 介绍了泛函的 合。符号 $g(x)=\mathcal{O}(f(x))$ 表示 $\lvert f(x)/g(x)\rvert$ 在 $x\to\infty$ 时有界。 2，则 $g(x)=\mathcal{O}(x^2)$。符号 $\lfloor x\rfloor$ 表示 x 向下取整，即小于或等于
2	2	27				 $p(T=0\mid C=0)=97/100=0.97$ (2.17) 率是归一化的，因此
3	2	32	式 2.47	$\mathbb{E}_{x,y}[xy]\mathbb{E}[x]\mathbb{E}[y]$	$\mathbb{E}_{x,y}[xy]-\mathbb{E}[x]\mathbb{E}[y]$	 $\mathrm{cov}[x,y]=\mathbb{E}_{x,y}[\{x-\mathbb{E}[x]\}\{y-\mathbb{E}[y]\}]$ $=\mathbb{E}_{x,y}[xy]\mathbb{E}[x]\mathbb{E}[y]$ (2.47) — 减号 如果 x 和 y 是独立的，那么它们的协方差为零
4	2	32	式 2.48	$\mathbb{E}_{x,y}[xy^{\mathrm{T}}]\mathbb{E}[x]\mathbb{E}[y^{\mathrm{T}}]$	$\mathbb{E}_{x,y}[xy^{\mathrm{T}}]-\mathbb{E}[x]\mathbb{E}[y^{\mathrm{T}}]$	 $\mathrm{cov}[\mathbf{x},\mathbf{y}]=\mathbb{E}_{\mathbf{x},\mathbf{y}}[\{\mathbf{x}-\mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}}-\mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}]$ $=\mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}]\mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]$ (2.48) — 减号 考虑向量 $\mathbf{x}$ 的分量彼此之间的协方差
5	3	60	2	$\det(\boldsymbol{\Sigma})$	$\lvert\boldsymbol{\Sigma}\rvert$	 $\mathcal{N}(\mathbf{x}\mid\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{D/2}}\frac{1}{\lvert\boldsymbol{\Sigma}\rvert^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$ (3.26) 其中 $\boldsymbol{\mu}$ 是 D 维均值向量，$\boldsymbol{\Sigma}$ 是 $D\times D$ 的协方差矩阵，$\det(\boldsymbol{\Sigma})$ 表示 $\boldsymbol{\Sigma}$ 的行列式。 在很多不同的场景中都可以看到高斯分布，因此可以从各种不同的角度来理解它

6	3	60	↑4	高斯分布在 $\boldsymbol{x}$ 空间的曲面上是常数，因为该二次型为常数。	在 $\boldsymbol{x}$ 空间中，令二次型是常数的曲面上，高斯分布的数值是常数。	 $\Delta^2=(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ (3.27) 其中，量Δ称为 $\boldsymbol{\mu}$ 到 $\boldsymbol{x}$ 的马哈拉诺比斯距离（Mahalanobis distance）。当 $\boldsymbol{\Sigma}$ 为单位矩阵时，它退化为欧氏距离。高斯分布在 $\boldsymbol{x}$ 空间的曲面上是常数，因为该二次型为常数。首先，注意可以不失一般性地假设矩阵 $\boldsymbol{\Sigma}$ 为对称矩阵，因为任何非对称的成分都会从指数中消失（见习题 3.11）。考虑协方差矩阵的特征方程： 中，令二次型是 的数值是
序号	章	页	行	误	正	备注
7	3	61	↑9	由于式（3.33）为常数，因此二次型（即高斯密度）在曲面上也为常数。	在满足式（3.33）取同一常数值的曲面上，二次型（即对应的高斯密度）在曲面上处处相同。	 式(3.29)可知 $\boldsymbol{U}$ 是正交矩阵，满足 $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}$，其中 $\boldsymbol{I}$ 是单位矩阵（参见附录A）。由于式(3.33)为常数，因此二次型（即高斯密度）在曲面上也为常数。如果所有特征值 λ_i 均为正，则这些曲面表示以 $\boldsymbol{\mu}$ 为中心，以 $\boldsymbol{u}_i$ 的方向为坐标轴且以 $\lambda_i^{1/2}$ 为半轴长 在满足 取同一值的曲面上
8	3	69				 $p(\boldsymbol{x}_a\mid\boldsymbol{x}_b)=\mathcal{N}(\boldsymbol{x}\mid\boldsymbol{\mu}_{a\mid b},\boldsymbol{\Lambda}_{aa}^{-1})$ (3.80) $\boldsymbol{\mu}_{a\mid b}=\boldsymbol{\mu}_a-\boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)$ (3.81)
9	4	98	13	参数 w_0 代表数据中一个任意的固定偏移量，有时称为偏置参数（bias）（注意不要与统计学意义上的偏差弄混淆）。通过定义一个额外的虚拟基函数 $\phi_0(\boldsymbol{x})=1$，式（4.2）可以统一为（参见 4.3 节）	参数 w_0 代表数据中一个任意的固定偏移量，有时称为偏置参数（bias）（注意不要与统计学意义上的偏差弄混淆，参见 4.3 节）。通过定义一个额外的虚拟基函数 $\phi_0(\boldsymbol{x})=1$，式（4.2）可以统一为	 为 M。参数 w_0 代表数据中一个任意的固定偏移量，有时称为偏置参数（bias）（注意不要与统计学意义上的偏差弄混淆）。通过定义一个额外的虚拟基函数 $\phi_0(\boldsymbol{x})=1$，式(4.2)可以统一为（参见 4.3 节）
10	4	102	↑10	在实践中，当 $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$接近奇异值时	在实践中，当 $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$接近奇异时	 空间中的正交投影得到的，其中每个基函数都视为一个长度为 N 的向量 $\boldsymbol{\varphi}_j$，元素为 $\phi_j(\boldsymbol{x}_n)$ $\boldsymbol{\Phi}\boldsymbol{w}_{\mathrm{ML}}$ 给出的 很容易地验证 在实践中，当 $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$ 接近奇异值时，直接求解正 特别是当两个或两个以上的基向量 $\boldsymbol{\varphi}_j$ 共线或近似共
11	4	110	8	高斯基函数和一个常数“偏差”基函数的模型	高斯基函数和一个常数“偏置参数”基函数的模型	 产生。数据集的索引是 $l=1,\cdots,L$，其中 $L=100$。对于每个数据集 $\mathcal{D}^{(l)}$，拟合一含 $M=24$ 个高斯基函数和一个常数“偏差”基函数的模型，总共 25 个参数。如所示，通过最小化正则化误差函数式(4.26)，可以得到预测函数 $f^{(l)}(x)$。 偏置参数 $\ln\lambda=3$

12	5	111	14	可以看到，较小的λ值允许模型对每个单独数据集上的噪声进行微调，从而导致较大的方差。	可以看到，较小的λ值会使模型过度拟合每个单独数据集上的噪声，从而导致较大的方差。	 其中，由分布 $p(x)$ 加权的对 x 的积分，可通过该分布产生的数据点的有限和来近似。图 4.8 给出了这些量在不同 $\ln \lambda$ 取值时的曲线。可以看到，较小的 λ 值允许模型对每个单独数据集上的噪声进行微调，从而导致较大的方差。相反，较大的 λ 值会使权重参数向 0 靠近，从而导致较大的偏差。 会使 过度拟合