

第 4 章作业答案

1

1.1

对于 Logistic 回归, 设 x 在分类面上,

$$\begin{aligned} P(y = 1|x) &= P(y = 0|x) \\ \frac{1}{1 + e^{-w^T x}} &= \frac{e^{-w^T x}}{1 + e^{-w^T x}} \\ w^T x &= 0 \end{aligned}$$

分类面是线性分类面。

对于 Softmax 回归, 设 x 在类 i 和类 j 的分类面上,

$$\begin{aligned} P(y = i|x) &= P(y = j|x) \\ \frac{e^{w_i^T x}}{\sum_{k=1}^K e^{-w_k^T x}} &= \frac{e^{w_j^T x}}{\sum_{k=1}^K e^{-w_k^T x}} \\ w_i^T x &= w_j^T x \\ (w_i^T - w_j^T)x &= 0 \end{aligned}$$

分类面是线性分类面。

1.2

$$\begin{aligned} g(f(\mathbf{x})) &= g(\mathbf{w}^T \mathbf{x} + \mathbf{w}_0) \\ &= k(\mathbf{w}^T \mathbf{x} + \mathbf{w}_0) + b \end{aligned}$$

仍然是 \mathbf{x} 的线性组合。

1.3

如果多层感知机使用线性激活函数，多层感知机的输出仍为输入的线性函数，无法拟合线性不可分的函数。

2

2.1

$$\begin{aligned}f'(x) &= \frac{e^{-x}}{(1 + e^{-x})^2} \\g'(x) &= \frac{4}{(e^x + e^{-x})^2} \\h'(x) &= \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}\end{aligned}$$

2.2

Sigmoid 优点: 输出范围有限，可以作为概率。

Sigmoid 缺点: 存在梯度消失、输出均值不为零、幂运算耗时。

tanh 优点: 输出范围有限、输出均值为零。

tanh 缺点: 存在梯度消失、幂运算耗时。

ReLU 优点: $x > 0$ 时梯度不会消失、收敛快、计算简单。

ReLU 缺点: 输出均值不为零、 $x < 0$ 时梯度为零。