

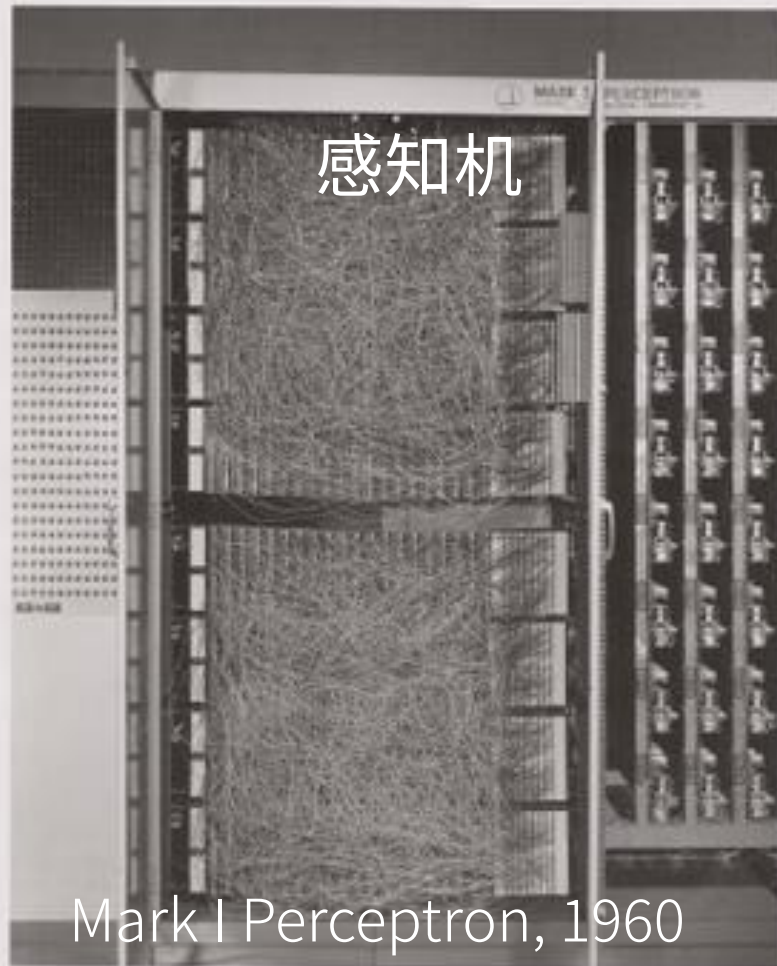
动手学深度学习 v2

李沐 · AWS



感知机

Mark I Perceptron, 1960

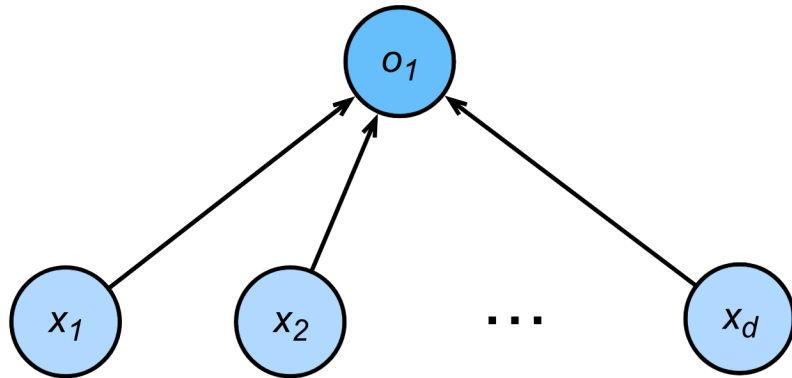


感知机



- 给定输入 \mathbf{x} , 权重 \mathbf{w} , 和偏移 b , 感知机输出:

$$o = \sigma(\langle \mathbf{w}, \mathbf{x} \rangle + b) \quad \sigma(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$





- 给定输入 \mathbf{x} , 权重 \mathbf{w} , 和偏移 b , 感知机输出:

$$o = \sigma(\langle \mathbf{w}, \mathbf{x} \rangle + b) \quad \sigma(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{otherwise} \end{cases}$$

- 二分类: -1 或 1
 - Vs. 回归输出实数
 - Vs. Softmax 回归输出概率



训练感知机

initialize $w = 0$ and $b = 0$

repeat

if $y_i [\langle w, x_i \rangle + b] \leq 0$ **then**

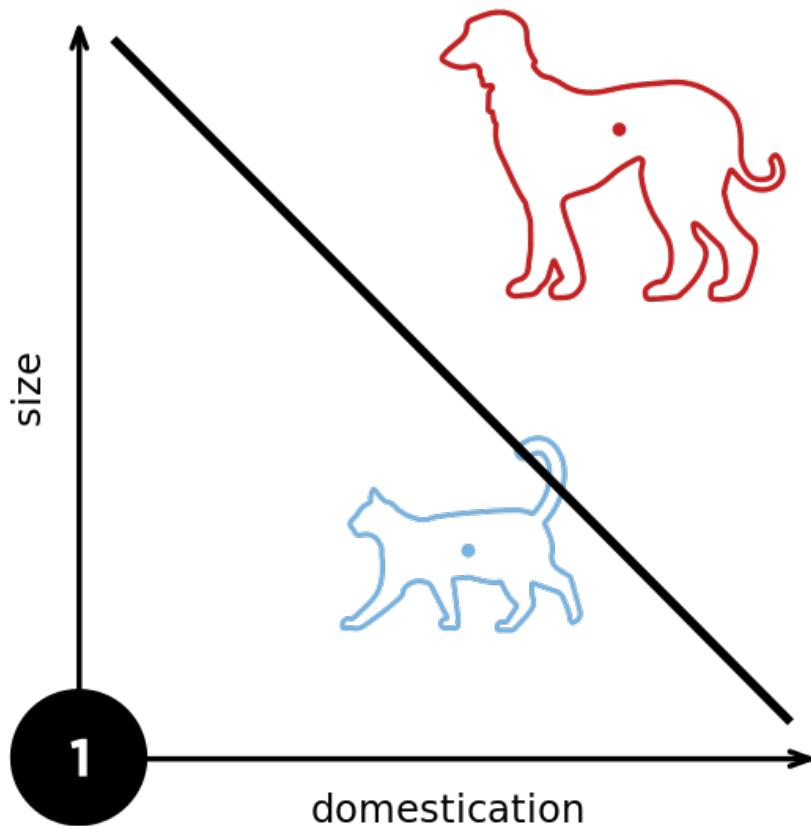
$w \leftarrow w + y_i x_i$ and $b \leftarrow b + y_i$

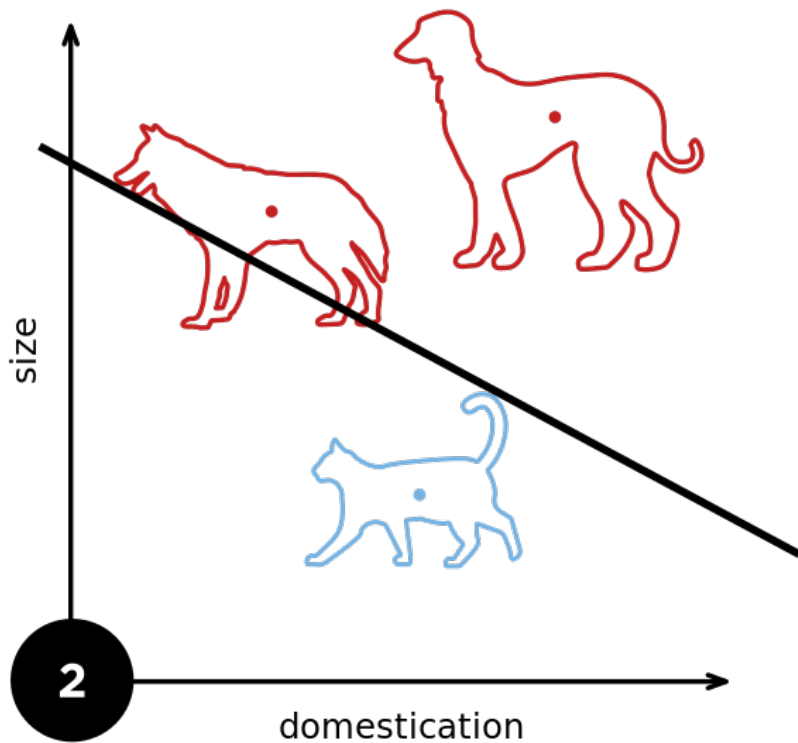
end if

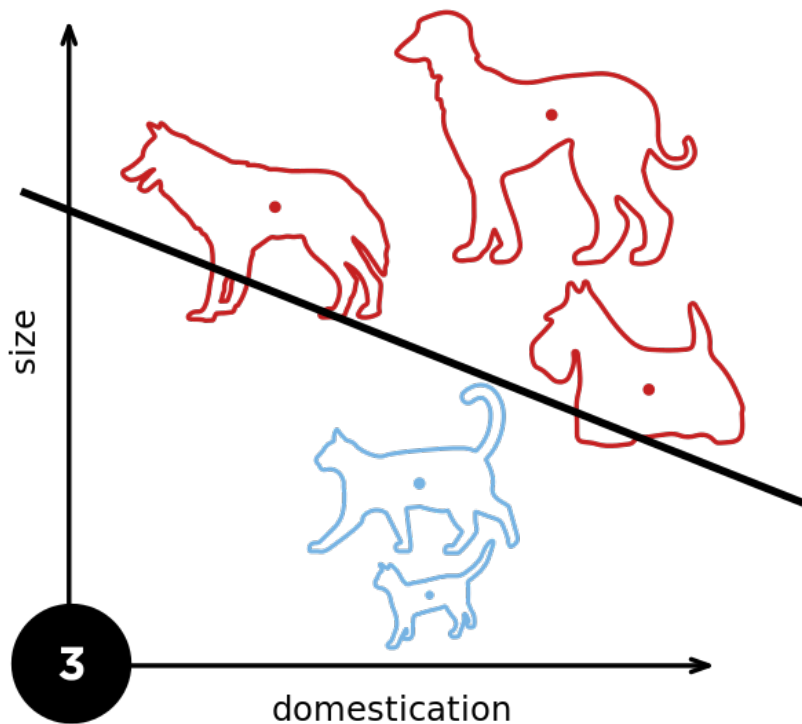
until all classified correctly

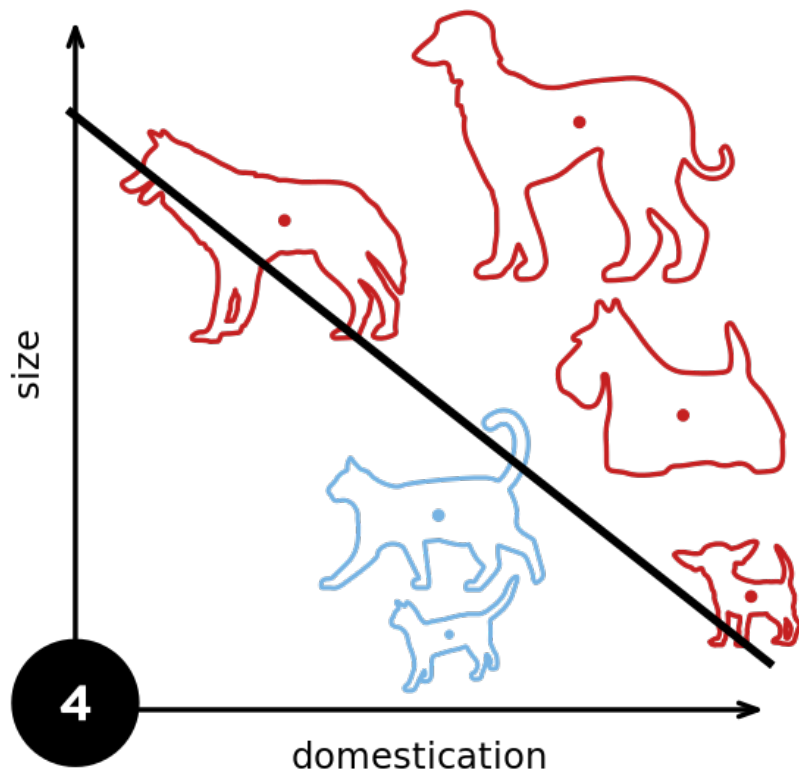
等价于使用批量大小为1的梯度下降，
并使用如下的损失函数

$$\ell(y, \mathbf{x}, \mathbf{w}) = \max(0, -y\langle \mathbf{w}, \mathbf{x} \rangle)$$









收敛定理

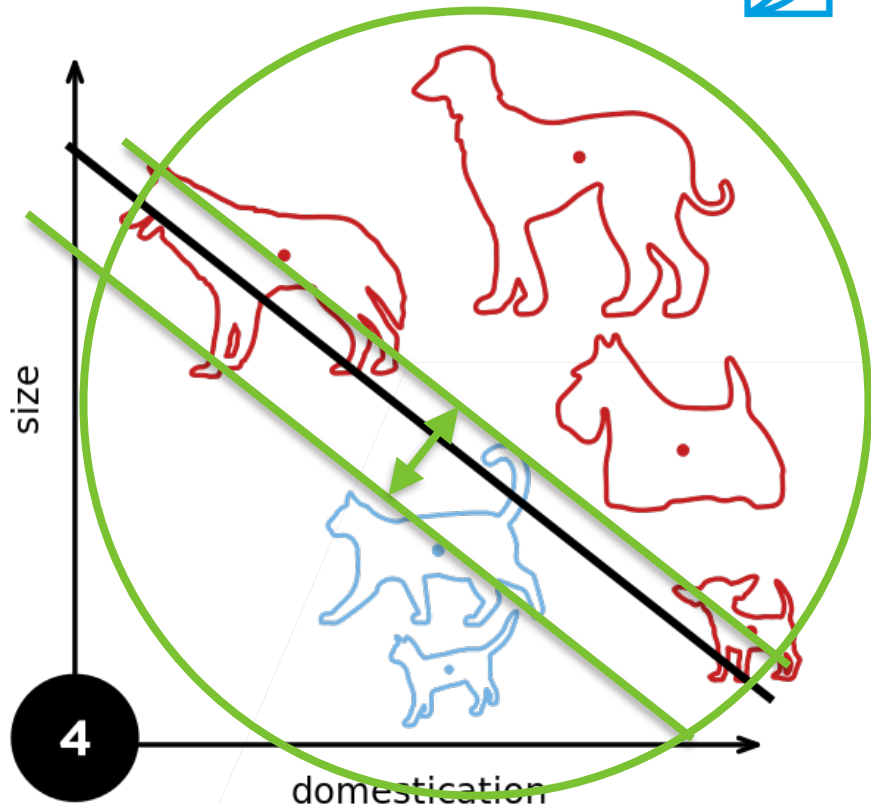


- 数据在半径 r 内
- 余量 ρ 分类两类

$$y(\mathbf{x}^\top \mathbf{w} + b) \geq \rho$$

对于 $\|\mathbf{w}\|^2 + b^2 \leq 1$

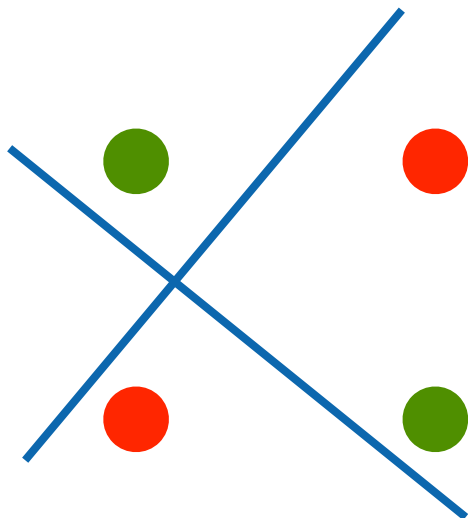
- 感知机保证在 $\frac{r^2 + 1}{\rho^2}$ 步后收敛



XOR 问题 (Minsky & Papert, 1969)



感知机不能拟合 XOR 函数，它只能产生线性分割面



总结



- 感知机是一个二分类模型，是最早的AI模型之一
- 它的求解算法等价于使用批量大小为1的梯度下降
- 它不能拟合 XOR 函数，导致的第一次 AI 寒冬