

[转] 不同损失函数的对比

作者：凯鲁嘎吉 - 博客园 <http://www.cnblogs.com/kailugaji/>

本章介绍的三种两类分类模型：Logistic 回归、感知器和支持向量机。虽然它们的决策函数相同，但由于使用了不同的损失函数以及相应的优化方法，导致它们之间在实际任务上的表现存在一定的差异。

为了比较这些损失函数，我们统一定义类别标签 $y \in \{+1, -1\}$ ，并定义 $f(\mathbf{x}; \mathbf{w}) = \mathbf{w}^T \mathbf{x} + b$ 。这样对于样本 (\mathbf{x}, y) ，若 $yf(\mathbf{x}; \mathbf{w}) > 0$ ，则分类正确，相反则分类错误。这样为了方便比较这些模型，我们可以将它们的损失函数都表述为定义在 $yf(\mathbf{x}; \mathbf{w})$ 上的函数。

Logistic 回归的损失函数可以改写为

$$\mathcal{L}_{LR} = -\log p(y|\mathbf{x}) \quad (3.100)$$

$$= -I(y = 1) \log \sigma(f(\mathbf{x}; \mathbf{w})) - I(y = -1) \log \sigma(-f(\mathbf{x}; \mathbf{w})) \quad (3.101)$$

$$= \log \left(1 + \exp(-yf(\mathbf{x}; \mathbf{w})) \right). \quad (3.102)$$

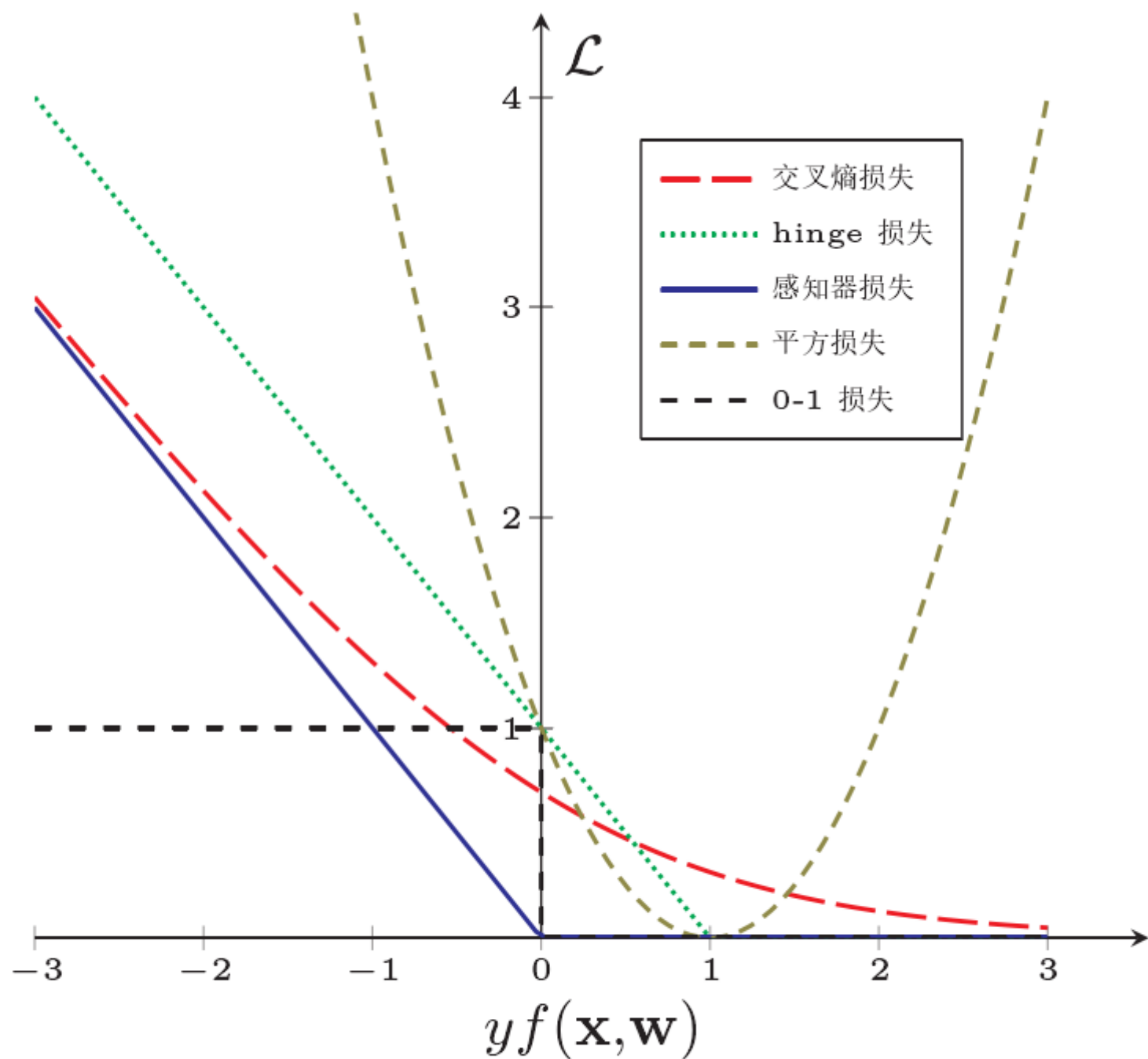
感知器的损失函数为

软

平

图

分类器
类器预



3.103)

3.104)

3.105)

3.106)

3.107)

> 0 时,
时,分
失函数

	激活函数	损失函数	优化方法
线性回归	-	$(y - \mathbf{w}^T \mathbf{x})^2$	最小二乘、梯度下降
Logistic 回归	$\sigma(\mathbf{w}^T \mathbf{x})$	$y \log \sigma(\mathbf{w}^T \mathbf{x})$	梯度下降
Softmax 回归	$\text{softmax}(W^T \mathbf{x})$	$y \log \text{softmax}(W^T \mathbf{x})$	梯度下降
感知器	$\text{sgn}(\mathbf{w}^T \mathbf{x})$	$\max(0, -y\mathbf{w}^T \mathbf{x})$	随机梯度下降
支持向量机	$\text{sgn}(\mathbf{w}^T \mathbf{x})$	$\max(0, 1 - y\mathbf{w}^T \mathbf{x})$	二次规划、SMO 等

表 3.1 几种不同的线性模型对比

参考文献:

[1] 邱锡鹏, [神经网络与深度学习](#)[M]. 2019.

[2] [交叉熵代价函数 \(作用及公式推导\)](#)