5 最大似然估计和逻辑回归

概要

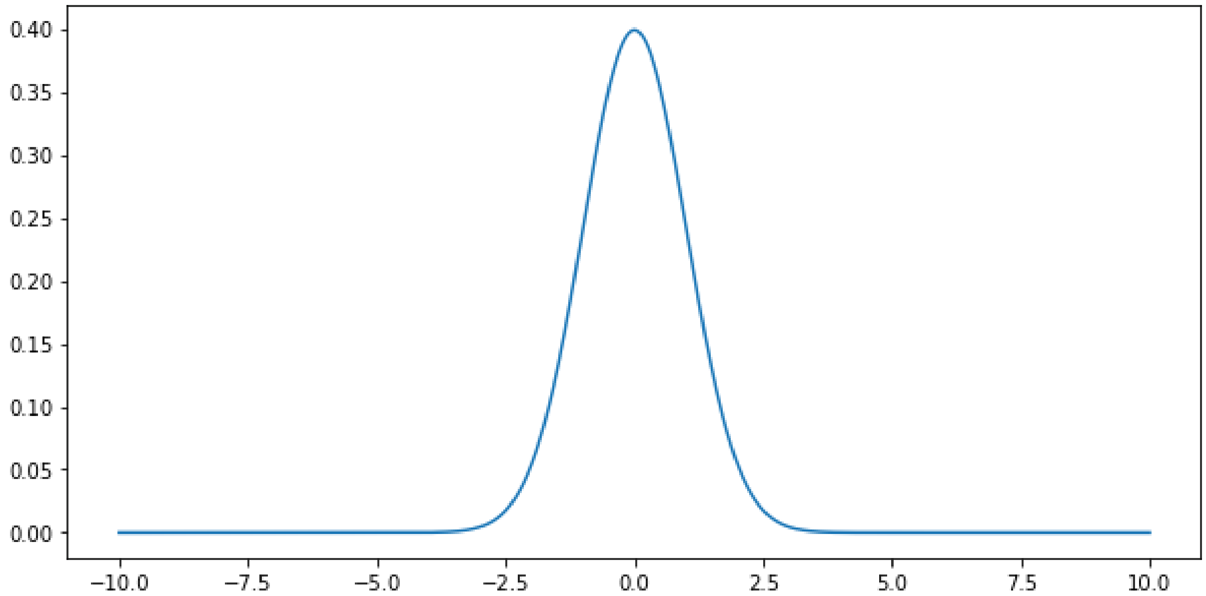
- 最大似然估计
- 损失函数
  - l_2 损失
  - l_1 损失
  - Huber 损失

最大似然估计

正态分布

- 概率密度函数

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



估计正态分布的参数

- 均值

$$\mu = \mathbf{E}[x]\text{，hence } \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- 方差

$$\sigma^2 = \mathbf{E}[(x-\mu)^2]\text{，hence } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

可能性

- 观测记录数据 $X = \{x_1, \dots x_n\}$
- 假设数据是从高斯分布生成的

$$p(X;\mu,\sigma^2) = \prod_{i=1}^{n} p(x_i;\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

- 拟合参数相当于使 $p(X;\mu,\sigma^2)$ 关于 μ,σ^2 最大化
- 实用简化版本:

$$\underset{\mu,\sigma^2}{\text{maximize}}\, p(X;\mu,\sigma^2) \Leftrightarrow \underset{\mu,\sigma^2}{\text{minimize}} - \log p(X;\mu,\sigma^2)$$

最大似然估计

- 通过可解释数据来估算参数

$$\underset{\mu,\sigma^2}{\text{minimize}} - \log p(X;\mu,\sigma^2)$$

- 展开式子

$$-\log p(X;\mu,\sigma^2) = \sum_{i=1}^{n} \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x_i-\mu)^2$$

$$= \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

当$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$时，可最小化

最大似然估计

➢估计方差

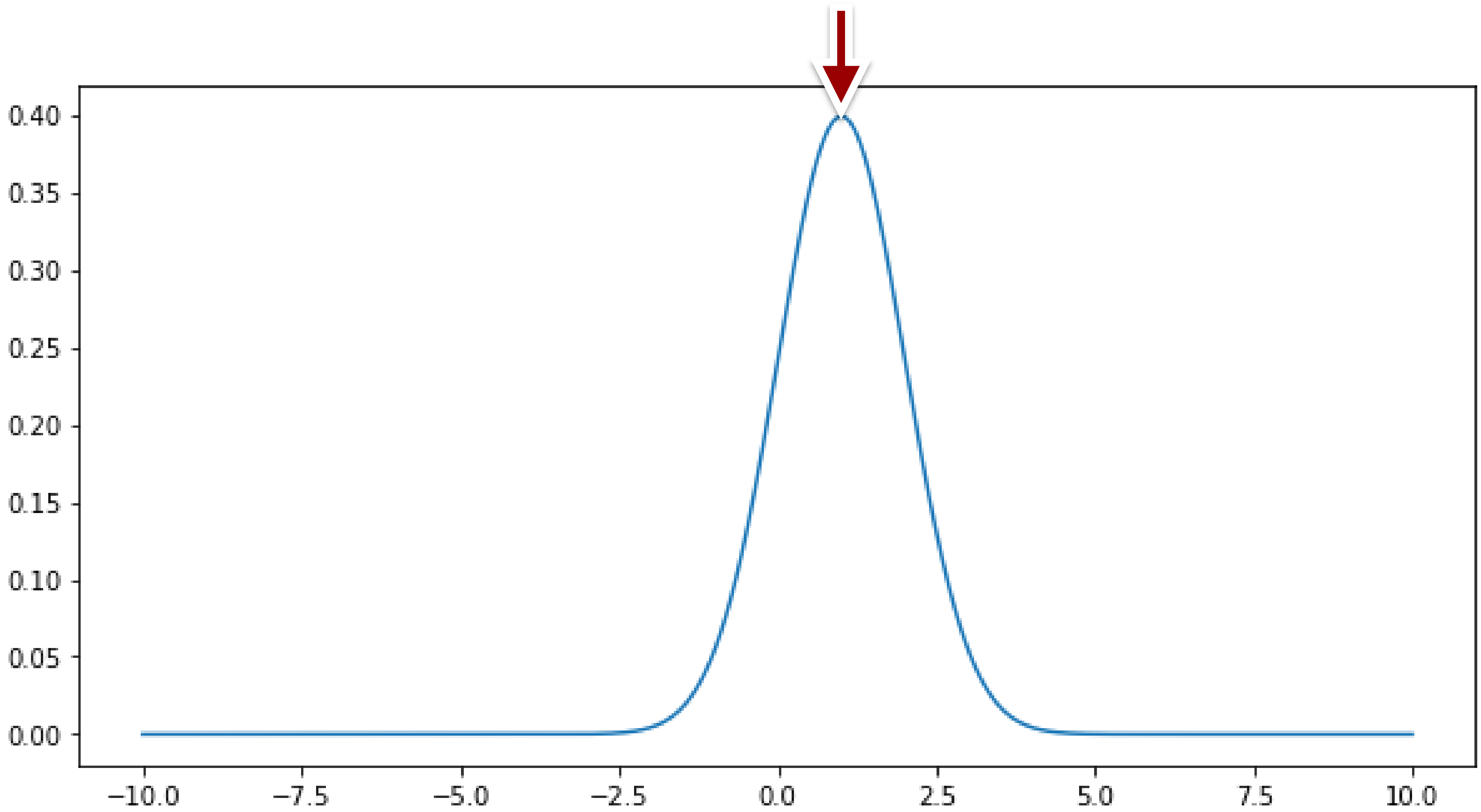
$$\frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

➢对基求导

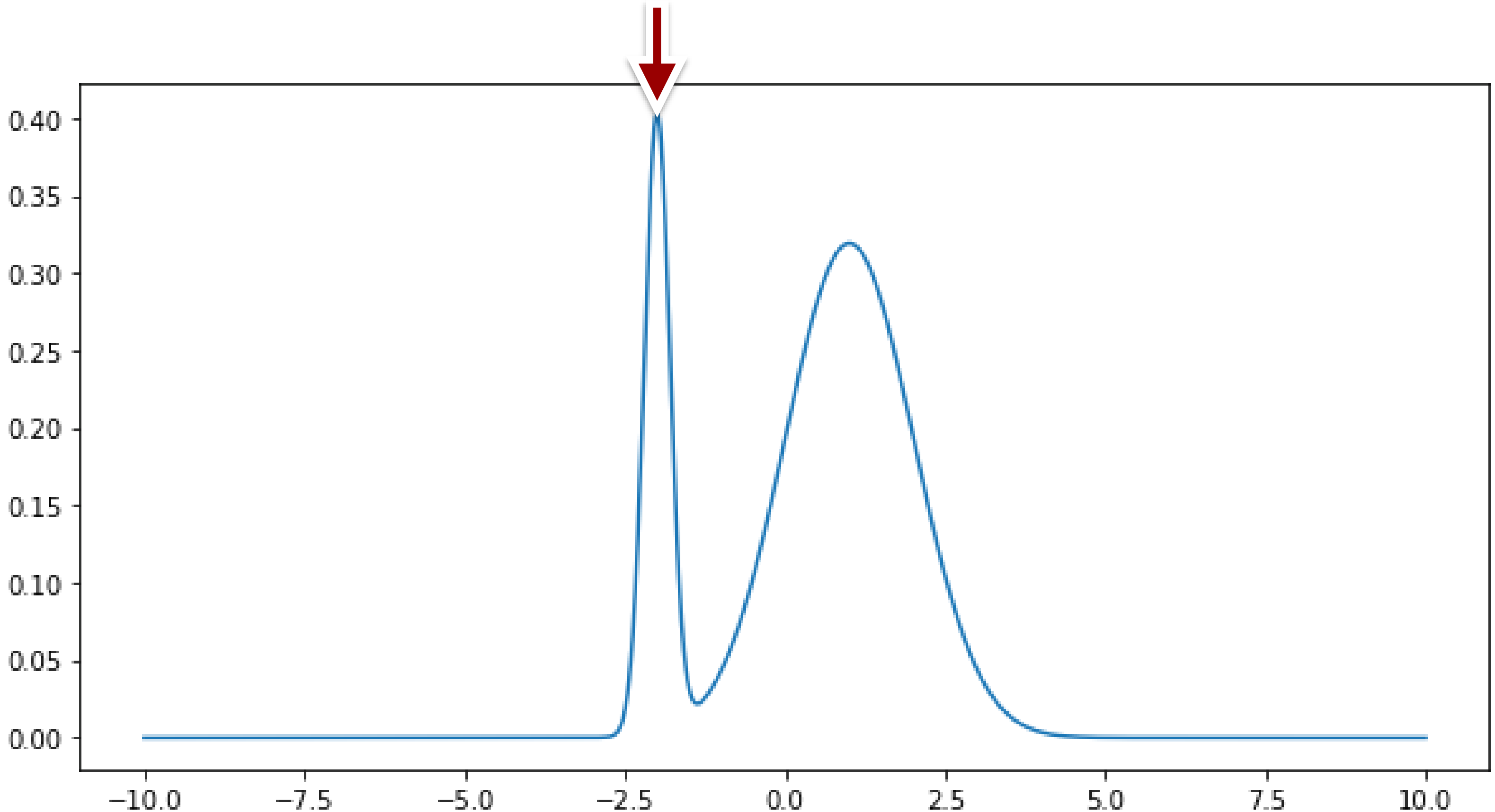
$$\partial_{\sigma^2}[\cdot] = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

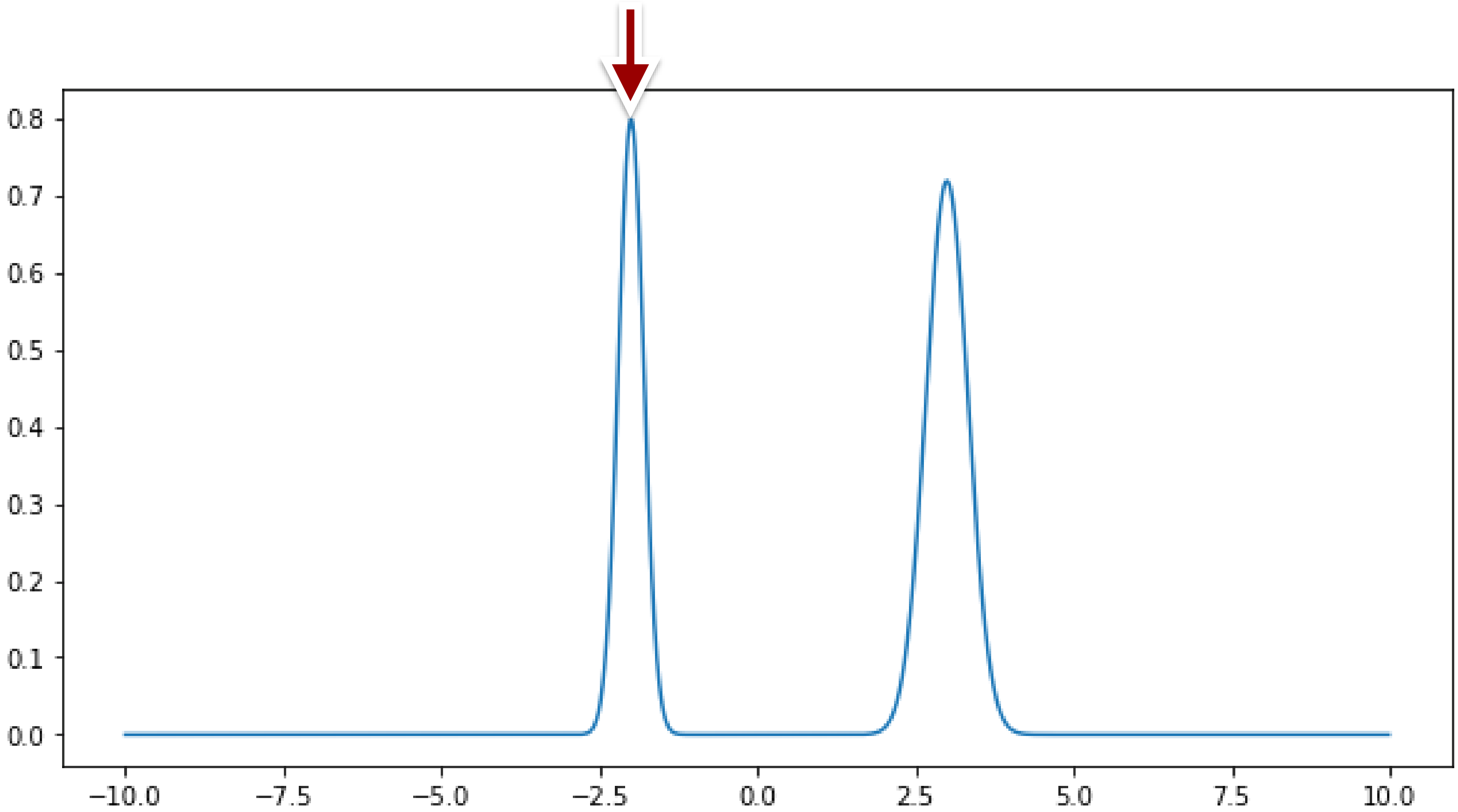
最大似然估计



最大似然估计



最大似然估计



最大似然估计

- 已有数据 - ‘学生未交作业’
- 可能的参数
  - ‘狗吃了作业’
  - ‘外星人拿走了作业’
  - ‘学生太懒’
  - ‘祖母生病’

- 所有参数可解释数据

最大后验估计

- 后验概率

$$p(w \mid X) \propto p(X \mid w)p(w)$$

$$\text{hence} \; -\log p(w \mid X) = -\log p(X \mid w) - \log p(w) + c$$

- 惩罚 c
- 最大后验估计

$$\underset{w}{\text{minimize}} - \log p(X; w) - \log p(w)$$

- 例子
  - P(未完成作业|合理解释) = P(合理解释|未完成作业）P(未完成作业)

lazy student	grandma sick	dog ate it	alien abduction
0.8	0.19	0.0099	0.0001

这与回归有什么关系？

回归

- 回顾 - 优化问题

$$\underset{w}{\text{minimize}} \sum_{i=1}^{n} (y_i - f(x_i, w))^2 + \text{penalty}(w)$$

  - 该模型有效吗？
  - 添加高斯噪声

- 数据生成模型

$$y_i = f(x_i, w) + \epsilon_i \text{where} \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- 高斯先验函数 $p(w)$

$$-\log p(w) = \frac{1}{2\overline{\sigma}^2} \| w \|^2 + \text{const.}$$

回归

➢最大后验估计

$$\underset{w}{\text{minimize}} - \log p(w \mid X, Y)$$

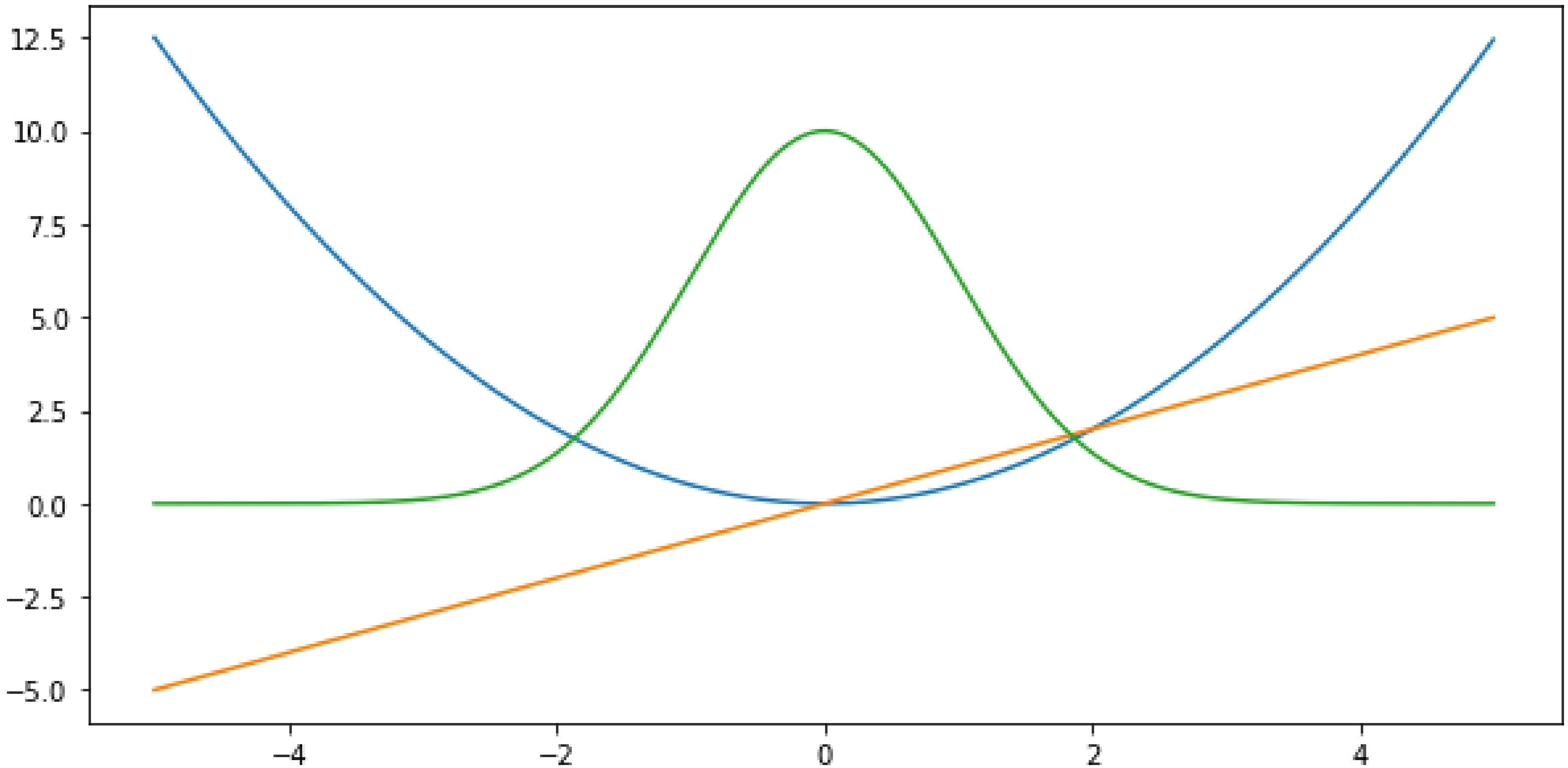
$$\Leftrightarrow \quad \underset{w}{\text{minimize}} \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - f(x_i, w)\right)^2 + \frac{1}{2\bar{\sigma}^2} \parallel w \parallel^2 + \text{const}\,.$$

$$_{winimize} \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - f(x_i, w)\right)^2 + \frac{\lambda}{2} \parallel w \parallel^2$$

损失函数

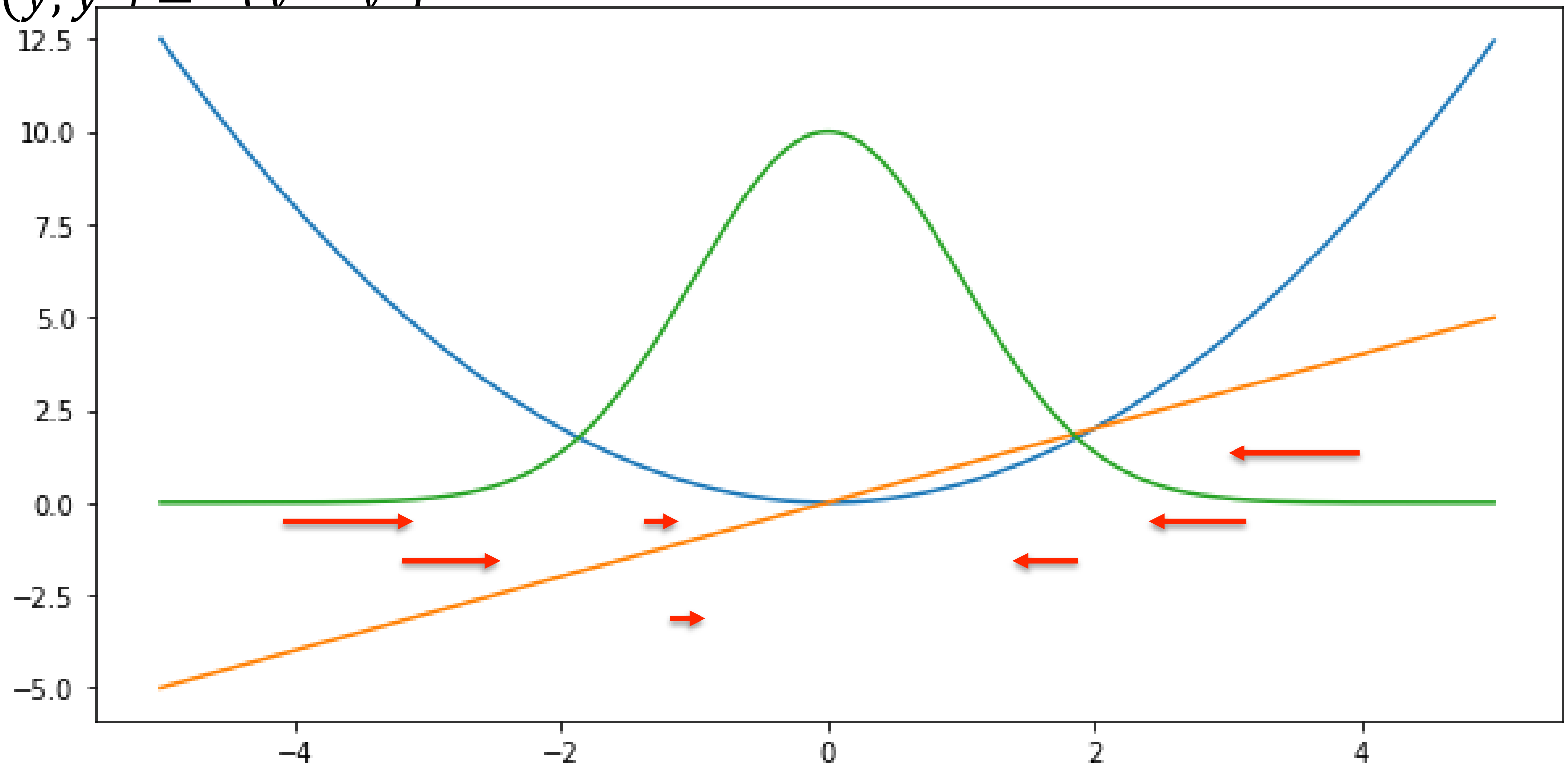
平方损失（L2 损失）

- $l(y, y') = \frac{1}{2}(y - y')^2$



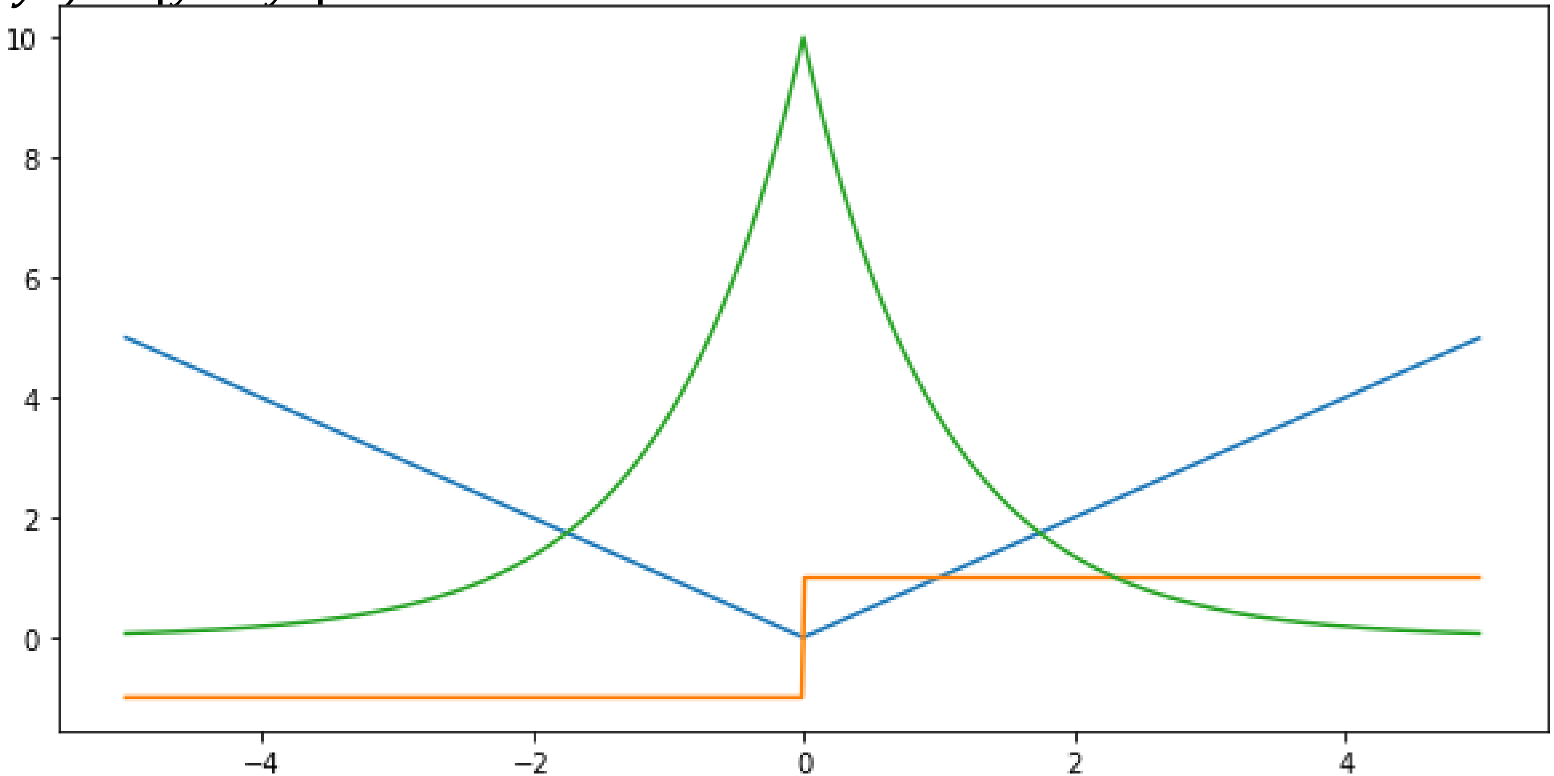
平方损失（L2 损失）

➢ $l(y, y') = \frac{1}{2}(y - y')^2$



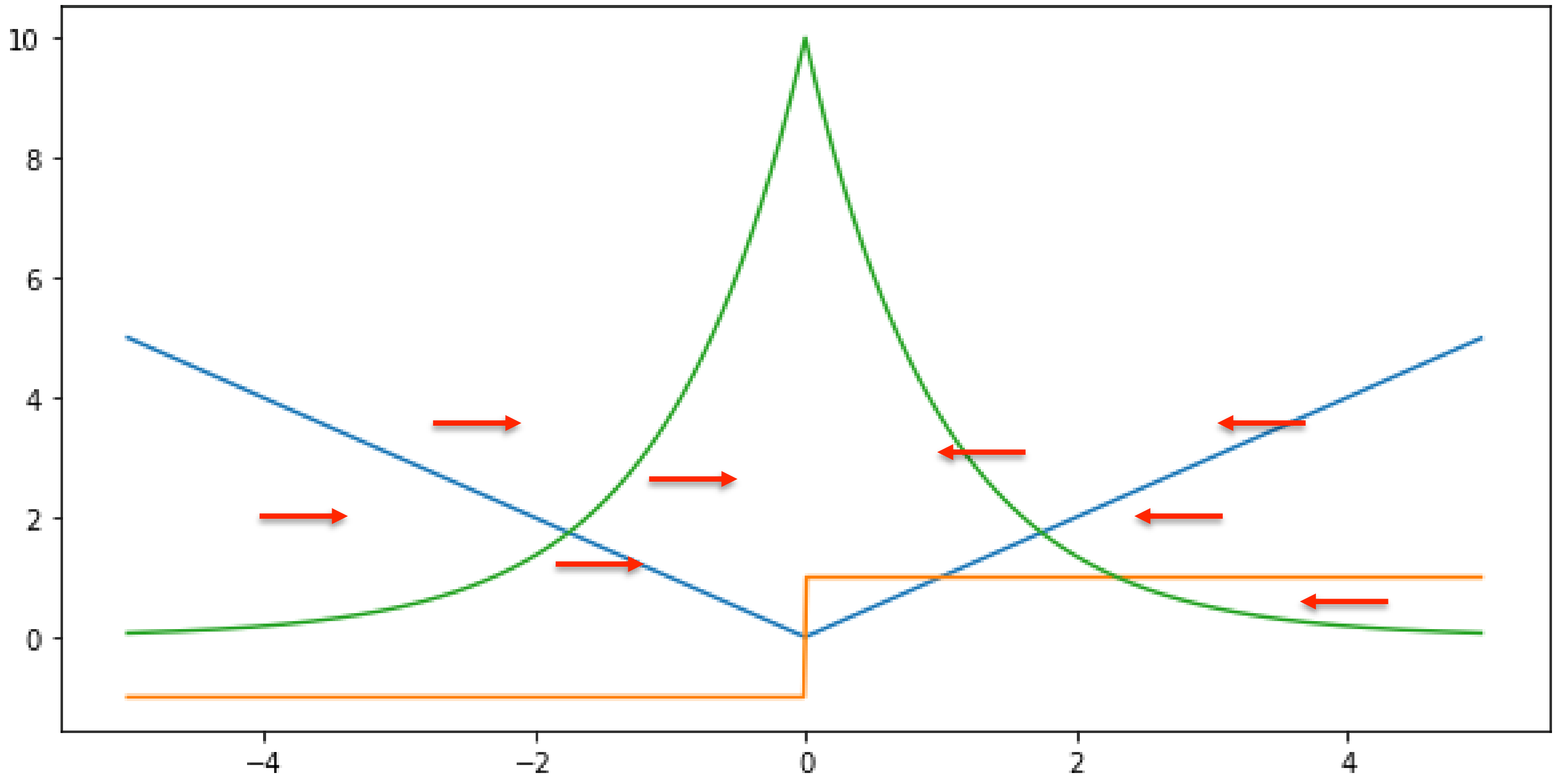
L1 损失

- $(y, y') = |y - y'|$



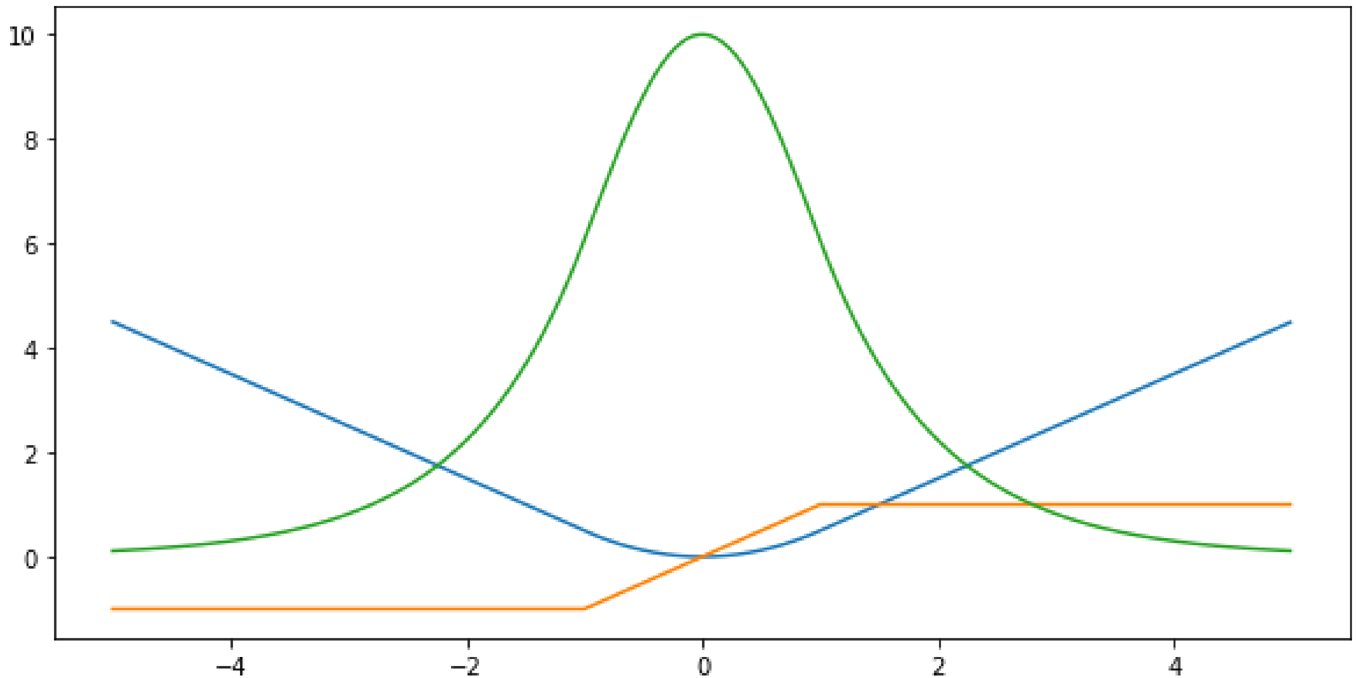
L1 损失

- $l(y, y') = |y - y'|$

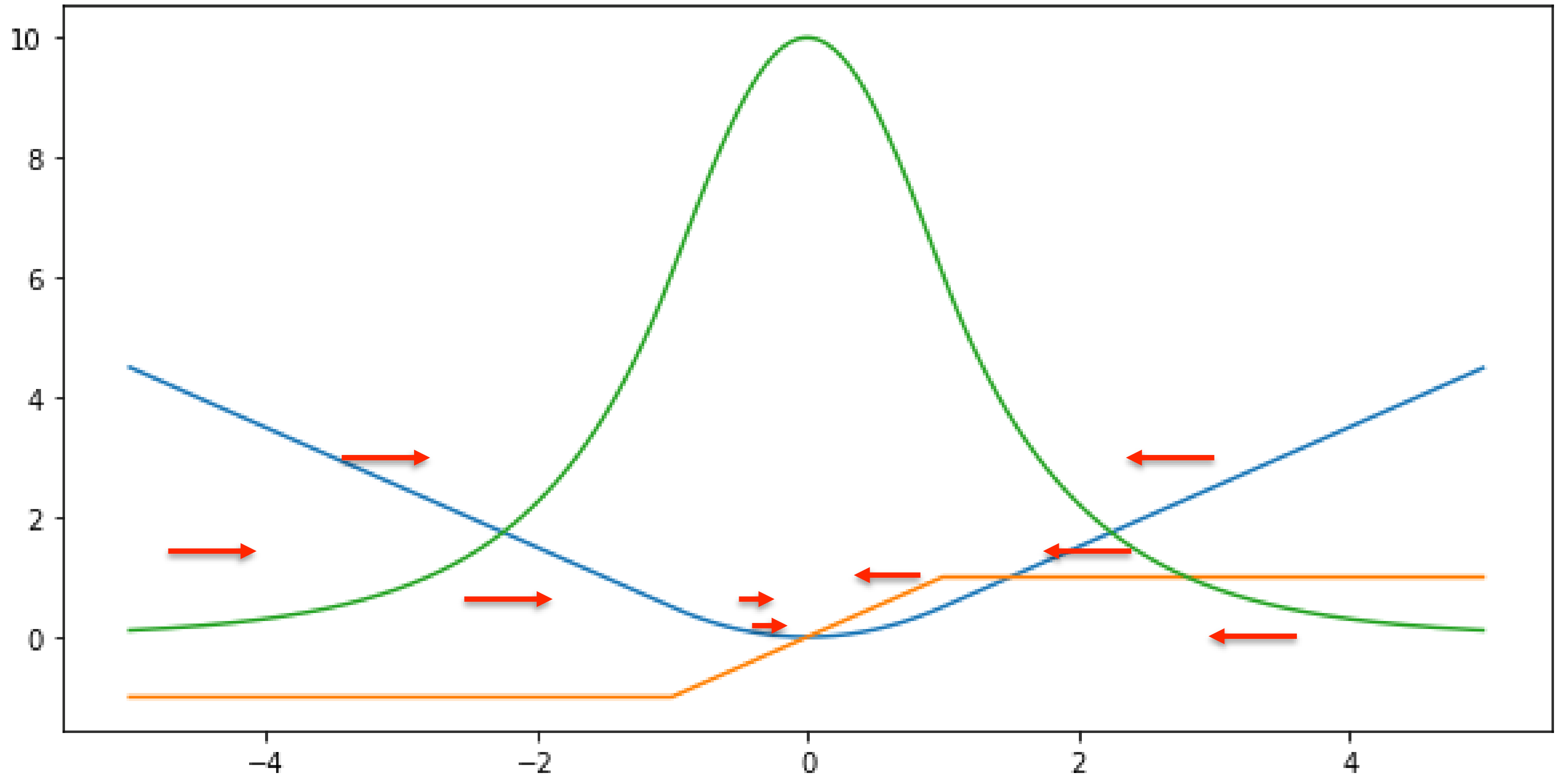


Huber 损失

- $l(y, y') = \begin{cases} |y - y'| - \frac{1}{2} & \text{if } |y - y'| > 1 \\ \frac{1}{2}(y - y')^2 & \text{otherwise} \end{cases}$



Huber 损失



总结

- 最大似然估计
- 损失函数
  - l_2 损失
  - l_1 损失
  - Huber 损失