Maximum Likelihood Estimate by Automatic Differentiation for Bernoulli Distribution

Weather (W) → good / bad

$W \sim \text{Bernoulli}(\theta)$

have some data $D = \{G, G, G, B, G, B, \dots\}$

but no θ

log-likelihood $\ell(D;\theta) = \sum_{i=0}^{N-1} \left(w^{(i)} \log\theta + (1-w^{(i)}) \log(1-\theta) \right)$

$\theta^* = \underset{\theta \in [0,1]}{\operatorname{argmax}}\ \ell(D;\theta)$

$\frac{\partial \ell}{\partial \theta}$ … tedious 

↳ solve gradient-based optimization by Automatic Differentiation

(we do not have to provide a derivative)

$\theta^* = \underset{\theta \in [0,1]}{\operatorname{argmin}} -\ell(D;\theta)$ — negative log likelihood

≙ "loss function"