III-Logics
Estimating the Parameters of an MLN

Foundations of Neuro-Symbolic AI

Alex Goessmann

University of Applied Science Würzburg-Schweinfurt

Summer Term 2026

Entropies

Definition

The Shannon entropy of a probability distribution is defined as

$$\mathbb{H}\left[\mathbb{P}\right] = \sum_{x_0,\dots,x_{d-1} \in \times_{k\in[d]}[m_k]} \mathbb{P}(x_0,\dots,x_{d-1}) \cdot \ln\left[\mathbb{P}(x_0,\dots,x_{d-1})\right]$$

Application in

- **Information Theory:** Minimal averaged code length to communicate samples
- **Machine Learning:** Measuring the degree of structure in a model

Further entropies

- Cross entropy: Likelihood of data for objective design
- Relative entropy (Kullback-Leibler divergence): Overhead by cross entropy as a distance measure of distributions

Likelihood of Data

Let there be a dataset of m samples in the form of a mapping

$$D : [m] \to \bigtimes_{k \in [d]} [2]$$

to observed states. The likelihood is the product

$$\mathbb{P}^{(\mathcal{F},\theta)}[D] = \prod_{j \in [m]} \mathbb{P}^{D(j)} \,.$$

The averaged negative log-likelihood is the cross entropy

$$\mathcal{L}_D\left(\mathcal{F},\theta\right) = -\frac{1}{m}\sum_{j \in [m]} \ln\left[\mathbb{P}^{D(j)}\right] = \mathbb{H}\left[\mathbb{P}^{D}, \mathbb{P}^{\mathcal{F},\theta}\right]$$

Optimization of the Likelihood

For each $f \in \mathcal{F}$ we have

$$\frac{\partial}{\partial \theta_f} \mathcal{L}_D(\mathcal{F}, \theta) = \mathbb{E}_{\mathbb{P}^D}[f] - \mathbb{E}_{\mathbb{P}^{\mathcal{F},\theta}}[f]$$

Algorithm: Alternating Weight Optimization

Iterate through $f \in \mathcal{F}$ until convergence and adjust θ_f such that $\frac{\partial}{\partial \theta_f} \mathcal{L}_D(\mathcal{F}, \theta) = 0$.

Entropy Maximization

Learning Task

Find a distribution with minimal structure reproducing key observed features.

Formalize

- minimal structure by maximal entropy
- key observed features by a set $\mathcal{F}$ such that

$$\forall f \in \mathcal{F} : \quad \mathbb{E}_{x \sim \mathbb{P}^{\theta}} [f(x)] = \mathbb{E}_{x \sim \mathbb{P}^{D}} [f(x)]$$

We state the **Entropy Maximation Problem**

$$\mathrm{argmax}_{\theta \in \Gamma} \mathbb{H}[\theta] \quad \text{subject to} \quad \forall f \in \mathcal{F} : \mathbb{E}_{x \sim \mathbb{P}^{\theta}} [f(x)] = \mathbb{E}_{x \sim \mathbb{P}^{D}} [f(x)] \qquad (P_{\Gamma, \mathcal{F}, D})$$

Characterization of the solution

By variational calculus one can proof the following:

Theorem

Among the distributions $\mathbb{P}$ satisfying

$$\forall f \in \mathcal{F} : \mathbb{E}_{x \sim \mathbb{P}} [f(x)] = \mathbb{E}_{x \sim \mathbb{P}^D} [f(x)]$$

the one with maximal entropy is the Markov Logic Network to $\mathcal{F}$ and the maximum likelihood weight θ.

- We did not assume that $\mathbb{P}$ is a Markov Logic Network, this is the result of the theorem!
- Removing the constraint that f is a map to [2], we can generalize to exponential families.