2.5 Maximum entropy principle

5. Maximum entropy principle.

Setting:

Given n data points $x^{(1)}, \cdots, x^{(n)}$.

A feature function $\phi: \mathcal{X} \to \mathbb{R}^d$.

Define the empirical moments: $\hat{\mu} := \frac{1}{n} \sum_{i=1}^{n} \phi(x^{(i)})$

Define $Q := \{ q \in \Delta_{|\mathcal{X}|} : E_q[\phi(x)] = \hat{\mu} \}$, where $E_q[\phi(x)] := \sum q(x)\phi(x)$

Note: $|\mathcal{X}|$ can be large but q only have d constraints.

[Definition 2] maximum entropy principle.

Choose the distribution $\hat{q}$ with the highest entropy:

$$\hat{q} := \arg\max_{q \in Q} H(q)$$

where $H(q) := E_q[-\log q(x)]$

We will show that Maximum Likelihood is equivalent to Maximum Entropy, in the following theorem.

[Theorem 1] maximum entropy duality:

Assume Q is non-empty, then

$$\arg\max_{q \in Q} H(q) = \arg\max_{p \in P} \sum_{i=1}^{n} \log p(x^{(i)})$$

Pf: Straightforward application of Langrangian duality.

$$\max_{q \in Q} H(q) = \max_{q \in \Delta_{|\mathcal{X}|}} \min_{\theta \in \mathbb{R}^d} H(q) - \theta \cdot (\hat{\mu} - E_q[\phi(x)]) \quad (i)$$

Since Q is non-empty (Slater's condition), we can switch the min & max.

$$\min_{\theta \in \mathbb{R}^d} \max_{q \in \Delta_{|\mathcal{X}|}} -\sum_{x \in \mathcal{X}} q(x) \log q(x) - \theta \cdot \left(\hat{\mu} - \sum_{x \in \mathcal{X}} q(x)\phi(x)\right) \quad (ii)$$

Next, differentiate w.r.t. q and set it to some constant c:

$$-(1+\log q(x)) + \theta\cdot\phi(x) = c \quad , \quad \text{for each } x\in\mathcal{X}.$$

(Since $\Sigma q(x) = 1$)

Solving for q (rewrite as q_θ) , then $q_\theta(x) \propto \exp(\theta\cdot\phi(x))$.

By (ii) , we have

$$\min_{\theta\in\mathbb{R}^d} -(\theta\cdot E_\theta[\phi(x)] - A(\theta)) - \theta\cdot(\hat{\mu} - E_\theta[\phi(x)])$$

$$\Longleftrightarrow \quad \max_{\theta\in\mathbb{R}^d} \theta\cdot\hat{\mu} - A(\theta) ,$$ which is the maximum likelihood objective.

Using the fact $\Delta A(\theta) = E_\theta[\phi(x)]$, we have

$$0 = \hat{\mu} - \Delta A(\theta) = \hat{\mu} - E_\theta[\phi(x)].$$

which means $q_\theta \in Q$ $\square$