12. Estimating the Kullback-Leibler (KL) Divergence

Estimating KL Divergence

Estimating the KL

$$\mathsf{KL}(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta) = \mathbb{E}_{\theta^*}\Big[\log\Big(\frac{p_{\theta^*}(X)}{p_\theta(X)}\Big)\Big]$$

Oops.

$$= \mathbb{E}_{\theta^*}\big[\log p_{\theta^*}(X)\big] - \mathbb{E}_{\theta^*}\big[\log p_\theta(X)\big]$$

So the function $\theta \mapsto \mathsf{KL}(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta)$ is of the form:

"constant" $-\ \mathbb{E}_{\theta^*}\big[\log p_\theta(X)\big]$

Can be estimated: $\mathbb{E}_{\theta^*}[h(X)] \rightsquigarrow \frac{1}{n}\sum_{i=1}^n h(X_i)$ (by LLN)

$$\widehat{\mathsf{KL}}(\mathbb{P}_{\theta^*}, \mathbb{P}_\theta) = \text{"constant"} - \frac{1}{n}\sum_{i=1}^n \log p_\theta(X_i)$$

/ blue curve

5:34 / 5:42 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Concept check: Properties of KL Divergence

1/1 point (graded)

Which of the following are properties of the **Kullback-Leibler KL divergence**?
(Choose all that apply.)

- [ ] The KL divergence is symmetric, *i.e.*, $\mathrm{KL}(\mathbf{P},\mathbf{Q}) = \mathrm{KL}(Q,P)$ for all distributions P,Q.
- [ ] The KL divergence is, strictly speaking, a distance function between probability distributions.
- [x] The KL divergence $\mathrm{KL}(P_{\theta^*},P_\theta)$ can be written as an expectation with respect to the distribution P_{θ^*}.
- [x] In general, it is easier to build an estimator for the KL divergence than it is to build an estimator for the total variation distance.

✔

Solution:

- The first choice is incorrect. The second problem in this section shows that the KL divergence is not symmetric.
- The second choice is also incorrect. A distance function, strictly, speaking must be symmetric and satisfy the triangle inequality. The KL divergence is not symmetric and does not satisfy the triangle inequality in general, so it is not a proper distance.
- The third choice is correct. Suppose that the distributions P_θ and P_{θ^*} are discrete and have pmfs p_θ and p_{θ^*}, respectively. Then

$$\mathrm{KL}(P_{\theta^*},P_\theta) = \sum_{x\in E} p_{\theta^*}(x)\ln\left(\frac{p_{\theta^*}(x)}{p_\theta(x)}\right) = \mathbb{E}_{\theta^*}\left[\ln\left(\frac{p_{\theta^*}}{p_\theta}\right)\right].$$

  Notation: Here we use the notation $\mathbb{E}_{\theta^*}$ to denote the expectation with respect to the distribution P_{θ^*}.

- The fourth choice is also correct. In general, it is very hard to compute the total variation between two distributions– even for two Gaussians this is a difficult computation. As a result, it is also difficult to build an estimator for total variation. The KL divergence is easier to compute, and since it can be written as an expectation, we can estimate the KL by taking averages.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 3 Methods of Estimation:Lecture 8: Distance measures between distributions / 12. Estimating the Kullback-Leibler (KL) Divergence

Add a Post

Show all posts ▼ by recent activity ▼

? KL divergence as an "approximation" for TV? 2