ECE 5630: Information Theory for Data Transmission, Security and Machine Learning 02/11/20

Lecture 5: Divergence and Convexity

Lecturer: Prof. Ziv Goldfeld

Scriber: Pengzhi Huang, *Net ID*: ph448

Assistant Editor: Kia Khezeli

5.1 Divergence

Our goal is to develop means to measure (a reasonable notion of) distance between probability measures.

Definition 5.1 (Divergence) *Consider a functional $\delta : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}_{\geq 0} \bigcup \{\infty\}$. We say that δ is a divergence if it satisfies: $\delta(P, Q) = 0 \iff P = Q$.*

Definition 5.2 (Metric) *A divergence δ is a metric, if for all $P, Q, R \in \mathcal{P}(\mathcal{X})$ it also satisfies:*

1. *Symmetry: $\delta(P, Q) = \delta(Q, P)$*
2. *Triangle inequality $\delta(P, Q) \leq \delta(P, R) + \delta(R, Q)$*

A large class of divergences falls under the framework of f-divergences, which is the focus of this chapter. This includes:

- Kullback–Leibler divergence
- Total variation (TV) divergence
- Hellinger distance
- χ^2-divergence
- Le Cam divergence
- Jensen-Shannon divergence

5.2 Primer: Convexity

Definition 5.3 (Convex Set) *A subset $\mathcal{K}$ of a vector space $\mathcal{V}$ is convex, if*

$$\alpha x + (1 - \alpha) y \in \mathcal{K}, \quad \forall x, y \in \mathcal{K},\ \alpha \in [0, 1].$$

Example 5.1 (Set of Probability Measures) *The set $\mathcal{P}(\mathcal{X})$ of all probability measures on $\mathcal{X}$ is convex.*

Proof: $\forall P_1, P_2 \in \mathcal{P}(\mathcal{X})$, $\forall \alpha \in [0, 1]$, define $P_\alpha := \alpha P_1 + (1 - \alpha) P_2$. Observe that:

(i) $P_\alpha(\mathcal{X}) = (\alpha P_1 + (1 - \alpha) P_2)(\mathcal{X}) = \alpha P_1(\mathcal{X}) + (1 - \alpha) P_2(\mathcal{X}) = \alpha + (1 - \alpha) = 1$

(ii) For $A_1, A_2, ... \in \mathcal{F}$ disjoint (i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$), we have

$$\begin{aligned} P_\alpha\left(\bigcup_{n=1}^{\infty} A_n\right) &= (\alpha P_1 + (1 - \alpha) P_2)\left(\bigcup_{n=1}^{\infty} A_n\right) = \alpha P_1\left(\bigcup_{n=1}^{\infty} A_n\right) + (1 - \alpha) P_2\left(\bigcup_{n=1}^{\infty} A_n\right) \\ &= \alpha \sum_{n=1}^{\infty} P_1(A_n) + (1 - \alpha) \sum_{n=1}^{\infty} P_2(A_n) = \sum_{n=1}^{\infty} P_\alpha(A_n) \end{aligned}$$

Thus, P_α is also a probability measure on $\mathcal{X}$, i.e., $P_\alpha \in \mathcal{P}(\mathcal{X})$, as claimed. ■

Definition 5.4 (Convex Function) *A function $f : \mathcal{K} \to \mathbb{R}$, for a convex set $\mathcal{K} \subseteq \mathbb{R}^d$, is convex if it satisfies:*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \quad \forall x, y \in \mathcal{K},\ \forall \alpha \in [0,1]$$

Remark 5.1 (Strict Convexity and Concavity)

- *f is strictly convex if the above inequality is strict;*
- *f is concave if $-f$ is convex.*

Definition 5.5 (Epigraph) *The epigraph of a function $f : \mathcal{K} \to \mathbb{R}$ is:*

$$\mathsf{epi}(f) := \big\{(x, y) \in \mathcal{K} \times \mathbb{R} : \ y \geq f(x)\big\}.$$

Proposition 5.1 *f is a convex function if and only if $\mathsf{epi}(f)$ is a convex set*

Proposition 5.2 (Convexity Preserving Operations) *Let $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$ be convex, then the following functions are also convex:*

1. *Sum:* $f_1 + f_2$
2. *Maximum:* $\max(f_1, f_2)$
3. *Composition with linear: $g : \mathbb{R}^n \to \mathbb{R}$ defined as $g(x) = f_1(\mathrm{A}x)$, where Let $\mathrm{A} \in \mathbb{R}^{d \times n}$ is a matrix.*

Proposition 5.3 (2nd Derivative Test) *$f : \mathcal{K} \to \mathbb{R}$, $\mathcal{K} \subseteq \mathbb{R}^d$ is convex, if and only if its Hessian:*

$$\mathsf{Hess}(f) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_d \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

is positive semi-definite ($\mathrm{A} \in \mathbb{R}^{n \times n}$ is positive semi-definite if $x^\top \mathrm{A} x \geq 0$, for all $x \in \mathbb{R}^n$).

Remark 5.2 (Strict Convexity and Concavity)

- *f is strictly convex if its Hessian is positive definite;*
- *f is concave if its Hessian is semi-negative definite;*
- *f is strictly concave if its Hessian is negative definite.*

Corollary 5.1 (Real-Valued Functions) *For a function $f : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval, we have:*

1. *f is convex, if and only if $\dfrac{\partial^2 f}{\partial x^2} \geq 0$;*
2. *f is strictly convex, if and only if $\dfrac{\partial^2 f}{\partial x^2} > 0$;*
3. *f is concave, if and only if $\dfrac{\partial^2 f}{\partial x^2} \leq 0$;*
4. *f is strictly concave, if and only if $\dfrac{\partial^2 f}{\partial x^2} < 0$.*

Theorem 1 (Jensen's Inequality) *Let* $X \sim P_X \in \mathcal{P}(\mathcal{X})$ *and* $f : \mathcal{X} \to \mathbb{R}$. *The following hold:*

1. *If* f *is convex, then* $\mathbb{E}[f(X)] \geq f\big(\mathbb{E}[X]\big)$;
2. *If* f *is strictly convex, then* $\mathbb{E}[f(X)] > f\big(\mathbb{E}[X]\big)$ *(unless* X *is almost surely a constant);*
3. *If* f *is concave, then* $\mathbb{E}[f(X)] \leq f\big(\mathbb{E}[X]\big)$.
4. *If* f *is strictly concave, then* $\mathbb{E}[f(X)] < f\big(\mathbb{E}[X]\big)$ *(unless* X *is almost surely a constant);*