Announcements

1. Scribes for lecture notes
2. HW2 released Friday
3. Next Thursday → Examples all lecture from Kia

Recap

f-divergences

Def: $D_f(P\|Q) = \mathbb{E}_Q f\left(\frac{dP}{dQ}\right) = \int f\left(\frac{dP}{dQ}(x)\right) dQ(x)$

where f is convex w/ $f(1)=0$

f strictly convex around 1

Examples: KL, TV, χ^2-divergence

$x \log x$ $\quad \frac{1}{2}|x-1|$ $\quad x^2-1$ $\quad (x-1)^2$

Properties

(i) Non-negativity: $D_f(P\|Q) \geq 0$ w/ equality iff $P=Q$

(ii) Convexity $(P,Q) \mapsto D_f(P\|Q)$ is jointly convex

(iii) Conditioning

$$P_X \to \boxed{P_{Y|X}} \to P_Y(\cdot) = \mathbb{E}_{P_X} P_{Y|X}(\cdot|X)$$

$$P_X \to \boxed{Q_{Y|X}} \to Q_Y(\cdot) = \mathbb{E}_{P_X} Q_{Y|X}(\cdot|X)$$

$$D_f(P_Y\|Q_Y) \leq D_f(P_{Y|X}\|Q_{Y|X}|P_X)$$

$$:= \int D_f(P_{Y|X}(\cdot|x)\|Q_{Y|X}(\cdot|x))\, dP_x(x)$$

(iv) Joint vs. Marginal

$$
\begin{tikzcd}
P_X \arrow[dr] & & P_Y(\cdot) := \mathbb{E}_{P_X} P_{Y|X}(\cdot|x) \\
& \boxed{P_{Y|X}} \arrow[ur] \arrow[dr] & \\
Q_X \arrow[ur] & & Q_Y(\cdot) := \mathbb{E}_{Q_X} P_{Y|X}(\cdot|x)
\end{tikzcd}
$$

$$D_f(P_X \| Q_X) = D_f(\underset{P_X P_{Y|X}}{P_{XY}} \| \underset{Q_X P_{Y|X}}{Q_{XY}})$$

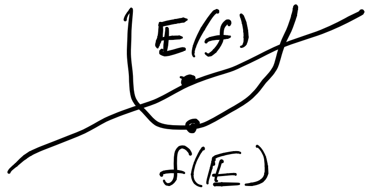
Sketch of Proof

(i) $D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right]$

$$\geq f\left(\mathbb{E}_Q\left[\frac{dP}{dQ}\right]\right)$$

$$\geq f(1)$$

$$= 0$$



Jensen's Inequality

equality iff follows by strict convexity

(ii) For any convex f, we may define "the perspective of f" which is given by

$$(x, y) \mapsto y f\left(\frac{x}{y}\right)$$

If f is convex, then perspective of f is jointly convex.

Recalling that $D_f(P\|Q) = \int f\left(\frac{dP}{dQ}\right) dQ$ and combining w/ the above leads to the conclusion.

(iii)

$$D_f(P_Y \| Q_Y) = D_f\left(\mathbb{E}_{P_X}[P_{Y|X}(\cdot|X)] \,\|\, \mathbb{E}_{P_X}[Q_{Y|X}(\cdot|X)]\right)$$
$$\leq \mathbb{E}_P \, D_f(P_{Y|X}(\cdot|X) \,\|\, Q_{Y|X}(\cdot|X))$$
$$= D_f(P_{Y|X} \| Q_{Y|X} | P_X)$$

(iv) $$D_f(P_{XY} \| Q_{XY}) = \mathbb{E}_{Q_{XY}} f\left(\frac{dP_{XY}}{dQ_{XY}}\right)$$

$$= \mathbb{E}_{Q_{XY}} f\left(\frac{dP_X \, dP_{Y|X}}{dQ_X \, dP_{Y|X}}\right)$$

$$= \mathbb{E}_{Q_X} f\left(\frac{dP_X}{dQ_X}\right) = D_f(P_X \| Q_X)$$

Data Processing Inequality

$$P_X \to \boxed{P_{Y|X}} \to P_Y(\cdot) = \mathbb{E}_{P_X} P_{Y|X}(\cdot|X)$$
$$Q_X \to \boxed{P_{Y|X}} \to Q_Y(\cdot) = \mathbb{E}_{Q_X} P_{Y|X}(\cdot|X)$$

Theorem: Let $P_X, Q_X \in \mathcal{P}(\mathcal{X})$ and $P_{Y|X}$ be a transition kernel. Then:

$$D_f(P_X \| Q_X) \geq D_f(P_Y \| Q_Y)$$

conditioning does NOT increase

Examples

① Gaussian Convolutions

$X \sim P_X \quad X' \sim Q_X$ and $\overset{\text{iid}}{\underset{Z' \sim}{Z \sim}} \underbrace{N(0, \sigma^2 I_d)}_{N_\sigma} \perp\!\!\!\perp X, X'$

Let

$$Y = X + Z \quad \text{and} \quad Y' = X' + Z'$$

(where the transition kernel is
$P_{Y|X}(\cdot|x) = N(x, \sigma^2 I_d)$)

Recall:

if $A \sim \mu \perp\!\!\!\perp B \sim \nu$

then $A + B \sim \mu * \nu$ where

elaborated below

$$(\mu * \nu)(A) = \boxed{\int_A \int_x \mathbb{1}_{\{x = y\}} d\mu(x)\, d\nu(y)}$$

$\Longrightarrow Y \sim P_X * N_\sigma$ and $Y' \sim Q_X * N_\sigma$

$\Longrightarrow$ <u>DPI</u>: $D_f(P_X \| Q_X) \geq D_f(\underbrace{P_X * N_\sigma}_{Y} \| \underbrace{Q_X * N_\sigma}_{Y'})$

$\left.\begin{array}{l}\mu \text{ has pdf } f \\ \nu \text{ has pdf } g\end{array}\right\} f * g$

$\mu * \nu$ is probability measure when pdf is $f * g$

② Deterministic Functions

$X \sim P_X$, $X' \sim Q_X$ and set $Y = g(X)$ for a deterministic measurable g.

here $P_{Y|X}(\cdot|x) = \delta_{g(x)}$ where δ_a is the

Dirac measure at a: $\delta_a(A) = \begin{cases} 1, & a \in A \\ 0, & o/w \end{cases}$

Also notice $\delta_a(A) = \mathbb{1}_A(a)$

(i) Take any event E and set

$$Y = \mathbb{1}_{\{X \in E\}}$$

• Y is a binary rv w/ $P_Y(1) = \mathbb{E}_{P_X}\left[\mathbb{1}_{\{X \in E\}}\right]$

$$= P_X(E)$$

$$\Rightarrow Y = \mathbb{1}_{\{X \in E\}} \sim \mathrm{Ber}(P_X(E))$$

$$Y' = \mathbb{1}_{\{X' \in E\}} \sim \mathrm{Ber}(Q_X(E))$$

$\Rightarrow$ DPI: For any measurable E

$$D_f(P_X \| Q_X) \geq D_f(\mathrm{Ber}(P_X(E)) \| \mathrm{Ber}(Q_X(E)))$$

easy to compute → allows for optimizations

(ii) When $X = (X_1, X_2) \sim P_{X_1, X_2}$

$X' = (X'_1, X'_2) \sim Q_{X_1, X_2}$

and $Y = g(X_1, X_2) = X_1$ and $Y' = g(X'_1, X'_2) = X'_1$

$Y \sim P_{X_1}$ $Y' \sim Q_{X_1}$

$$\xrightarrow{DPI} D_f(P_{X_1, X_2} \| Q_{X_1, X_2}) \geq D_f(P_{X_1} \| Q_{X_1})$$

Proof DPI

$$D_f(P_X \| Q_X) \overset{\text{prop (iv)}}{=} D_f(\overset{P_X P_{Y|X}}{P_{XY}} \| \overset{Q_X P_{Y|X}}{Q_{XY}})$$

$$= \mathbb{E}_{Q_{XY}}\left[f\left(\frac{dP_{XY}}{dQ_{XY}} \right) \right]$$

Law of Total Probability

$$= \mathbb{E}_{Q_Y}\left[\mathbb{E}_{Q_{X|Y}}\left[f\left(\frac{dP_{XY}}{dQ_{XY}} \right) \Big| Y \right] \right]$$

Jensen's

$$\geq \mathbb{E}_{Q_Y}\left[f\left(\underbrace{\mathbb{E}_{Q_{X|Y}}\left[\frac{dP_{XY}}{dQ_{XY}} \Big| Y \right]}_{*} \right) \right]$$

look at ⊛

$$= \mathbb{E}_{Q_{X|Y=y}}\left[\frac{dP_{XY}}{dQ_{XY}} \,\Big|\, Y=y\right] = \mathbb{E}_{Q_{X|Y=y}}\left[\frac{dP_Y(y)P_{X|Y}(\cdot|y)}{dQ_Y(y)Q_{X|Y}(\cdot|y)} \,\Big|\, Y=y\right]$$

$$= \mathbb{E}_{P_{X|Y=y}}\left[\frac{dP_Y(y)}{dQ_Y(y)} \,\Big|\, Y=y\right]$$

$$\sum_{x\in\mathcal{X}} \cancel{q_{X|Y}}(x|y)\, \frac{P_Y(y)P_{X|Y}(x|y)}{q_Y(y)\,\cancel{q_{X|Y}(x|y)}}$$

$$= \sum_{x\in\mathcal{X}} P_{X|Y}(x|y)\frac{P_Y(y)}{q_Y(y)} = \frac{P_Y(y)}{q_Y(y)}$$

$$= \frac{dP_Y(y)}{dQ_Y(y)}$$

$\Rightarrow$ from ⊛ $\quad h(y) = \dfrac{dP(y)}{dQ(y)}$

$$h(Y) = \frac{dP(Y)}{dQ(Y)} \quad !$$

$$D_f(P_X\|Q_X) \geq \mathbb{E}_{Q_Y}\left[f\left(\frac{dP_Y(Y)}{dQ_Y(Y)}\right)\right]$$

$$= D_f(P_Y\|Q_Y)$$

Q.E.D