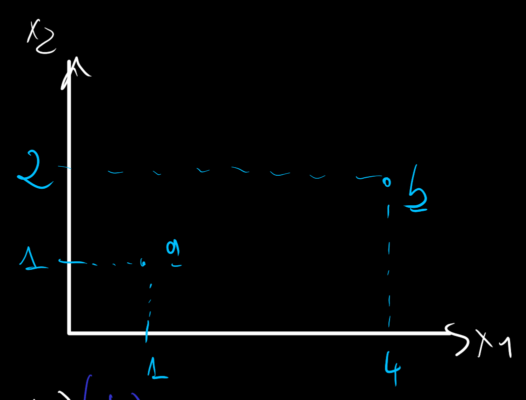
Kullback-Leibler Divergence (KL)

Introduction

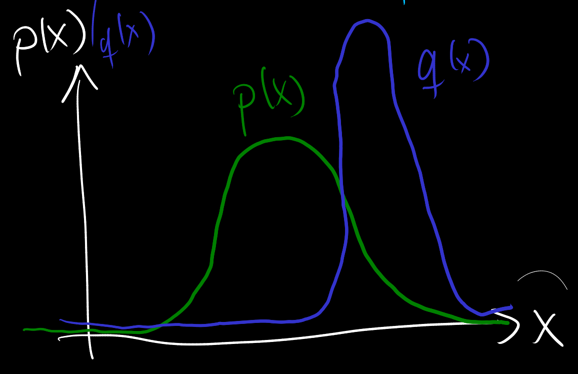
distance between a & b

$d(a,b) = |a-b|$



distance between $\underline{a}$ & $\underline{b}$

$d(\underline{a},\underline{b}) = \|\underline{a}-\underline{b}\|_2$



distance between p & q
(between two distributions)

↳ KL-Divergence

$$D_{KL}(p\|q) = \mathbb{E}_{X\sim p(x)}\left[\log\left(\frac{p(x)}{q(x)}\right)\right] = \begin{cases} \sum\limits_{X} p(x)\cdot\log\left(\frac{p(x)}{q(x)}\right) & \text{if } X \text{ discrete} \\ \int\limits_{X} p(x)\cdot\log\left(\frac{p(x)}{q(x)}\right) & \text{if } X \text{ continuous} \end{cases}$$

KL-Divergence does not have all properties of distance

Example

Weather (W) $\begin{cases} \to \text{bad} \\ \to \text{good} \end{cases}$ → Bernoulli distributions

$W \in \{\overset{0}{\text{Bad}}, \overset{1}{\text{Good}}\}$

$W \sim \text{Bernoulli}(\theta)$

prob of good weather

W

Person 1
$\theta_A = 0.8$
$W \sim \text{Bernoulli}(\theta_A) = p(w)$

Person 2
$\theta_B = 0.7$
$W \sim \text{Bernoulli}(\theta_B) = q(w)$

How far apart are the distributions?

$$D_{KL}(p\|q) = \sum_{W} p(w)\log\left(\frac{p(w)}{q(w)}\right)$$

$$= \sum_{w=0}^{1} p(W=w)\log\left(\frac{p(W=w)}{q(W=w)}\right)$$

$\left[\text{Bernoulli}(w;\theta) = \theta^{w}(1-\theta)^{1-w}\right.$

$$= \underbrace{\theta_A^{0}\cdot(1-\theta_A)^{1-0}}_{1-\theta_A}\cdot\log\left(\frac{\overbrace{\theta_A^{0}\cdot(1-\theta_A)^{1-0}}^{1-\theta_A}}{\underbrace{\theta_B^{0}\cdot(1-\theta_B)^{1-0}}_{1-\theta_B}}\right) + \underbrace{\theta_A^{1}\cdot(1-\theta_A)^{1-1}}_{\theta_A}\cdot\log\left(\frac{\overbrace{\theta_A^{1}\cdot(1-\theta_A)^{1-1}}^{\theta_A}}{\underbrace{\theta_B^{1}\cdot(1-\theta_B)^{1-1}}_{\theta_B}}\right)$$

$$= (1-\theta_A)\cdot\log\left(\frac{1-\theta_A}{1-\theta_B}\right) + \theta_A\cdot\log\left(\frac{\theta_A}{\theta_B}\right)$$

$= 0.0257$