1 Divergence

Two probability distributions can P and Q can be defined over the same space. Suppose the input domain is defined $X = \{x_1, ..., x_N\} \sim P$ and $X = \{\ x_1, ..., x_N\} \sim Q$ There are two metrics to compute the divergence $D(P, Q)$ we can compute the difference $D(P - Q)$ or the fractional density $D(\frac{P}{Q})$.

1.1 f-divergence

The f-divergence gives the distribution in terms of the density ration $r(x) = p(x)/q(x)$.

$$D_f(p||q) = \int q(x) f(\frac{p(x)}{q(x)})\, dx$$

The function in f divergence must be a convex function which satisfies the following properties $D_f(p||q) \leq 0$ and $D - f(p||p) = 0$.

1.1.1 KL divergence

If we compute the function $f(r)$ as $f(r) = r \log(r)$ we get the *Kullback Leibler divergence*.

$$D_{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)}\, dx$$

1.1.2 Alpha divergence

The alpha divergence is given by $f(x) = \frac{4}{1-\alpha^2}(1 - x^{\frac{1+\alpha}{2}})$ As $\lim_{\alpha \to \infty}$ the distribution q prefers to cover the whole of p whereas $\lim_{\alpha \to -\infty}$ the distribution of q prefers to cover a single mode of p. If $a \to 0$ the alpha divergence tends towards the KL $D_{KL}(p||q)$ divergence. If $\alpha \to 1$ then the alpha divergence approaches KL $D_{KL}(q||p)$ if $\alpha \to 0.5$ the alpha divergence is equal to the Hellinger distance.

1.1.3 Hellinger distance

The squared hellinger distance

$$D_H^2(P)$$