

## 数据热力学

信息熵:  $H(X) = -\sum_{x \in X} p(x) \log p(x)$

联合熵:  $H(X, Y) = -\sum_x \sum_y P(x, y) \log P(x, y)$

条件熵:  $H(Y|X) = -\sum_x p(x) H(Y|X=x) = -\sum_{x,y} p(x,y) \log \frac{p(x)}{p(x,y)}$

互信息:  $I(X, Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) = -\sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x,y)}$

信息增益:  $GI(P, Q) = H(P) - H(P|Q) = -\sum P(i) \log \frac{Q(i)}{P(i)}$

相对熵 (KL 散度):  $D_{KL}(P \parallel Q) = -\sum P(i) \log \frac{Q(i)}{P(i)}$

交叉熵:  $H(p, q) = -\sum p(x) \log q(x) = H(p) + D_{KL}(p \parallel q)$

(在信息论中, 基于相同事件测度的两个概率分布  $p$  和  $q$  的交叉熵是指, 当基于一个“非自然”(相对于“真实”分布  $p$  而言) 的概率分布  $q$  进行编码时, 在事件集合中唯一标识一个事件所需要的平均比特数。)

互信息 Venn 图

