

Cross Entropy, KL-Div, and Likelihood

Random Var : $X \in \{x_1, \dots, x_n\}$ prob. given by a model : $\{p(X = x_i)\}$

Data Sample (i.i.d) : $Y = (y_1 y_2 \dots y_N)$, $y_i \in \{x_1, \dots, x_n\}$

$$\text{Empirical distribution : } \tilde{p}(X = x_i) = \frac{\text{count}(x_i, Y)}{N} = \frac{\sum_{j=1}^N \delta(y_j, x_i)}{N}$$

$$\text{loglikelihood : } \log L(Y) = \sum_{j=1}^N \log p(X = y_j) = \sum_{i=1}^n \text{count}(x_i, Y) \log p(X = x_i) = N \sum_{i=1}^n \tilde{p}(x_i) \log p(x_i)$$

$$\frac{1}{N} \log L(Y) = -H(\tilde{p}, p) = -D(\tilde{p} \parallel p) - H(\tilde{p})$$

Fix the data \Rightarrow fix Y , \tilde{p}

$$p^* = \arg \max_p \frac{1}{N} \log L(Y) = \arg \min_p H(\tilde{p}, p) = \arg \min_p D(\tilde{p} \parallel p) = \arg \min_p 2^{-\frac{1}{N} \log L(Y)}$$

**Example: $X \in \{\text{"H"}, \text{"T"}\}$
 $Y = (\text{HHTTH})$**

$$\tilde{p}(X = \text{"H"}) = \frac{c(\text{"H"}, Y)}{5} = 3/5$$

$$\delta(y, x) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

Equivalent criteria for selecting/evaluating a model

Perplexity(p)

