Cross-Entropy

$H_c(X_1, X_2) = -\sum_{x \in D} p_1(x) \log(p_2(x))$

$= -\int_{x \in D} p_1(x) \log(p_2(x))\, dx$ — continuous random var x

$= -\int_{\bar{x} \in D} p_1(\bar{x}) \log(p_2(\bar{x}))\, d\bar{x}$ — vector

Both take value x from D

majority contribution from dissimilarity

- → $p_{gt} = 1$, $p_{pred} \neq 1$ ⟹ does not contribute to H_c
- → p_{gt} ↑, p_{pred} ↓ ⟹ high H_c
- → p_{gt} ↓ ⟹ will not contribute a lot

> uniform distri → high entropy
>
> steep peaks + low elsewhere → low entropy

> minimum no. of bits needed to communicate a message if we know where message is drawn from

K-L Divergence

Cross entropy = entropy + divergence in distri

$H_c(X_1, X_2) = H(X_1) + D(X_1, X_2)$

$D(X_1, X_2) = H_c(X_1, X_2) - H(X_1)$

$D(X_1, X_2) = -\sum_{x \in D} p_1(x) \log(p_2(x)) + \sum_{x \in D} p_1(x) \log p_1(x)$

$= -\sum_{x \in D} p_1(x) \log(p_1(x) - p_2(x))$

$$KL = -\sum_{x \in D} p(x) \log\left(\frac{p_1(x)}{p_2(x)}\right) = -\int_{x \in D} p_1(x) \log\left(\frac{p_1(x)}{p_2(x)}\right)$$

- → multiclass classification: $\min(H_c) \rightleftharpoons \min(KL)$
- → KL not symmetric, depends on gt
- → always positive
- → KL divergence b/w identical random vars = 0

$\log\left(\frac{p_1(x)}{p_1(x)}\right) = \log(1) = 0$

JS Divergence

- KL asymm → depends on which distri is considered "reference"
- JS = symm

$$JSD(P||Q) = \frac{KL(P||M) + KL(Q||M)}{2}$$

where M = avg var $= \frac{P+Q}{2}$

order matters

$D_{JS}(P||Q) = \frac{1}{2} D_{KL}\left(P, \frac{P+Q}{2}\right) + \frac{1}{2} D_{KL}\left(Q, \frac{P+Q}{2}\right)$

→ JSD > 0 ⟹ diff distri

$P = [0.7, 0.2, 0.1]$ $Q = [0.3, 0.4, 0.3]$

$M = [0.5 \;\; 0.3 \;\; 0.2]$

$D_{KL}(P||M) = -\sum_{i=1}^{n} P_i \left(\frac{\log P_i}{\log Q_i}\right)$

$= -\left[0.7\left(\frac{\log 0.7}{\log 0.5}\right)\right] - \left[0.2\left(\frac{\log 0.2}{\log 0.3}\right)\right] - \left[0.1\left(\frac{\log 0.1}{\log 0.2}\right)\right]$

$= 0.1228$

$D_{KL}(Q||M) = 0.1204$

$D_{JS} = \frac{D_{KL}(P||M) + D_{KL}(Q||M)}{2} = 0.1216$ bits

P, Q are similar

Conditional Entropy

Entropy of one variable given that another is a constant

$H(w|H=h_i) = -\sum_{i=1}^{n} P(w_i|h_i) \log(P(w_i|h_i))$

$H(w|H) = \sum_{i=1}^{n} P(h_i) \left[-\sum_{i=1}^{n} P(w_i|h_i) \log(P(w_i|h_i))\right]$