Entropy

1 Recall

Mean, Median, Variance, Standard Deviation...

2 Definition : $H(P)$

Q : How can you describe the probability distribution with just "a" number?

We are given the probability distribution P. (We have $p_1, p_2, ..., p_n$ such that $\forall i, p_i \geq 0$ and $\sum p_i = 1$).

$$H(P) \triangleq \sum_i p_i(-\log p_i) = \Big(p_1(-\log p_1) + p_2(-\log p_2) + ... + p_n(-\log p_n)\Big).$$

You might recall that the entropy is the average of bits that is needed to send the information.

3 Cross Entropy : $H(P, Q)$

Q : What is the average of bits when we do not know the true probability distribution P?

Since we do not know the true probability distribution P, we choose some distribution Q. The bits corresponding to p_i is $-\log q_i$. Therefore, the answer is $\sum_i p_i(-\log q_i)$.

$$H(P, Q) \triangleq \sum_i p_i(-\log q_i).$$

4 KL divergence : $D_{KL}(P||Q)$

Q : What is the loss of bits when we do not know the true probability distribution P?

A : Cross Entropy - Entropy

$$D_{KL}(P||Q) \triangleq H(P,Q) - H(P) = \sum_i p_i(-\log q_i) - \sum_i p_i(-\log p_i) = \sum_i p_i \log \frac{p_i}{q_i}.$$

5 Log Loss

Assume that estimated click probability is p. (non-click probability is $1-p$.)

We calculate KL-divergence between P(click event "or" non-click event) with Q(estimated click probability distribution).

1. Click event

	P	Q
non-click	0%	$1-p$
click	100%	p

$$D_{KL}(P||Q) \triangleq H(P,Q) - H(P) = \sum_i p_i(-\log q_i) - \sum_i p_i(-\log p_i)$$
$$= \Big(0(-\log(1-p)) + 1(-\log p)\Big) - \Big(0(-\log 0) + 1(-\log 1)\Big) = -\log p.$$

2. Non-Click event

	P	Q
non-click	100%	$1-p$
click	0%	p

$$D_{KL}(P||Q) \triangleq H(P,Q) - H(P) = \sum_i p_i(-\log q_i) - \sum_i p_i(-\log p_i)$$
$$= \Big(1(-\log(1-p)) + 0(-\log p)\Big) - \Big(1(-\log 1) + 0(-\log 0)\Big) = -\log(1-p).$$

5.1 Definition

Summing up, $C \cdot (-\log p) + (1-C) \cdot (-\log(1-p))$ where C is 1 if click event and 0 otherwise.

$$LogLoss(C) \triangleq C \cdot (-\log p) + (1-C) \cdot (-\log(1-p)).$$