Optimization

Mahdi Roozbahani
Georgia Tech

Outline

Motivation

Entropy

Conditional Entropy and Mutual Information

Cross-Entropy and KL-Divergence

Let's work on this subject in our Optimization lecture

Cross Entropy

Cross Entropy: The expected number of bits when a wrong distribution Q is assumed while the data actually follows a distribution P

$$H(p,q) = -\sum_{x \in \mathcal{X}} p(x) \log q(x) = H(P) + KL[P][Q]$$

This is because:

$$H(p,q) = \mathrm{E}_p[l_i] = \mathrm{E}_p\left[\log \frac{1}{q(x_i)}\right]$$

$$H(p,q) = \sum_{x_i} p(x_i) \log \frac{1}{q(x_i)}$$

$$H(p,q) = -\sum_{x} p(x) \log q(x).$$

Labeling target values
Label encoding (ordinal) and One-hot encoding

Why Cross entropy and not simply use dot product?

Kullback-Leibler Divergence

Another useful information theoretic quantity measures the difference between two distributions.

$$\mathbf{KL}[P(S)\|Q(S)] = \sum_s P(s)\log\frac{P(s)}{Q(s)}$$

$$= \underbrace{\sum_s P(s)\log\frac{1}{Q(s)}}_{\text{cross entropy}} - \mathbf{H}[P] = H(P,Q) - H(P)$$

Excess cost in bits paid by encoding according to Q instead of P.

KL Divergence is a **KIND OF** distance measurement

$$-\mathbf{KL}[P\|Q] = \sum_s P(s)\log\frac{Q(s)}{P(s)}$$

log function is concave or convex?

$$\sum_s P(s)\log\frac{Q(s)}{P(s)} \leq \log\sum_s P(s)\frac{Q(s)}{P(s)}$$

By Jensen Inequality

$$= \log\sum_s Q(s) = \log 1 = 0$$

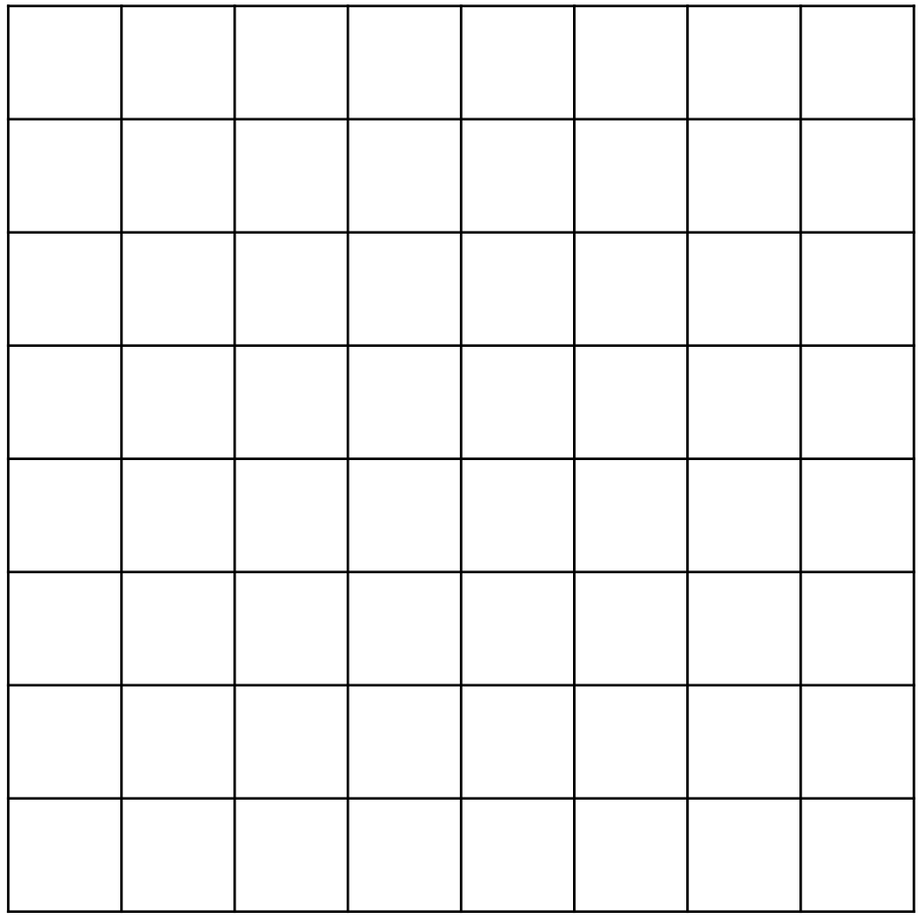
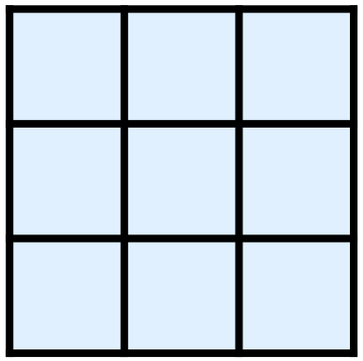
So $\mathbf{KL}[P\|Q] \geq 0$. Equality iff $P = Q$

When $P = Q,\ KL[P||Q] = 0$

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 2 \\ 2 & 1 & 3 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -2 \\ 2 & 0 & -3 \\ 1 & 0 & -1 \end{bmatrix}$$

$$\text{sum}\left(\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \right) =$$

$$0 + 0 + -2 + 2 + \cdots = -3$$

0	1	2	4	5	6	0	2
2	1	3	8	9	255	72	83
1	2	1	4	79	65	53	33
43	97	15	67	104	77	163	43
36	173	13	76	205	89	174	34
63	163	113	86	209	91	185	98
84	153	123	96	134	101	196	121
96	143	133	79	135	103	216	211

	-3