2. Information Measure

Vaughan Sohn

October 5, 2024

Contents

Entropy

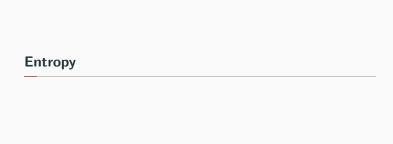
Mutual Information

KL-Divergence

Remarks about Informations measures

Convexity and Concavity of Information Measures

Data Processing Inequality and Fano's Inequality



Entropy

Represent INFORMATION

Definition 1 (entropy on event)

Entropy

Definition 2 (entropy on random variable)

By definition, entropy has following properties:



Definition 3 (joint entropy)

Definition 4 (conditional entropy on observable)

Multivariable Entropy

Definition 5 (conditional entropy on r.v.)

 $\sqrt{\frac{\text{meaning:}}{\text{Proof:}}}$

Examples

•

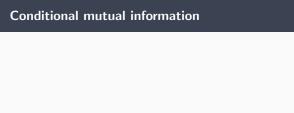


Mutual Information

Definition 6 (mutual information)

By definition, mutual information has following properties:

-



Definition 7 (conditional mutual information)

Chain rule

Theorem 8 (chain rule)



KL-Divergence

Definition 9 (KL-divergence)

•

KL-Divergence

By definition, KL-Divergence has following properties:



Useful facts for Information Measures

Theorem 10

Useful facts for Information Measures

Theorem 11

Theorem 12

Application of Information Measures

Hypothesis testing



Convexity and Concavity

Definition 13 (convexity)

Definition 14 (concavity)

•

Prerequisites: Log-sum inequality

Lemma 15 (log-sum inequality)

Convexity of KL-divergence

Theorem 16

•

√ meaning:

Convexity of KL-divergence

* <u>Proof</u>:

Concavity of Entropy

Corollary 17

√ meaning:

Concavity and Concavity of Mutual Information

Corollary 18

√ meaning:

Concavity and Concavity of Mutual Information

Corollary 19

√ meaning:



Prerequisites: Markov chain

Definition 20 (Markov chain)

Data Processing Inequality

Theorem 21 (data processing inequality)

√ meaning:



Corollary 22 (Data Processing Inequality on entropy)

Corollary 23 (Data Processing Inequality on KL-divergence)

Fano's Inequality

System:

Theorem 24 (Fano's inequality)

√ meaning:

Fano's Inequality

Appendix

Notations

- entropy of r.v. $X \sim P_X$: $H(X), H(P_X)$
- conditional entropy of Y conditioned on X = x: $H(Y|X = x), H(P_{Y|X}(\cdot|x))$
- conditional entropy of Y conditioned on X: $H(Y|X), H(P_{Y|X})$
- mutual information of X,Y: $I(X;Y),I(P_X\cdot P_{Y|X}),I(P_X,P_{Y|X})$ 주로 $W=P_{Y|X}$ 로 두고 $I(P_X\cdot W)$ 와 같이 표기하여 사용한다.
- KL divergence between two distribution P and Q: $D(P||Q), D(P_P||P_Q)$

References

- T. M. Cover and J. A. Thomas. Elements of Information Theory, Wiley, 2nd ed., 2006.
- Lecture notes for EE623: Information Theory (Fall 2024)