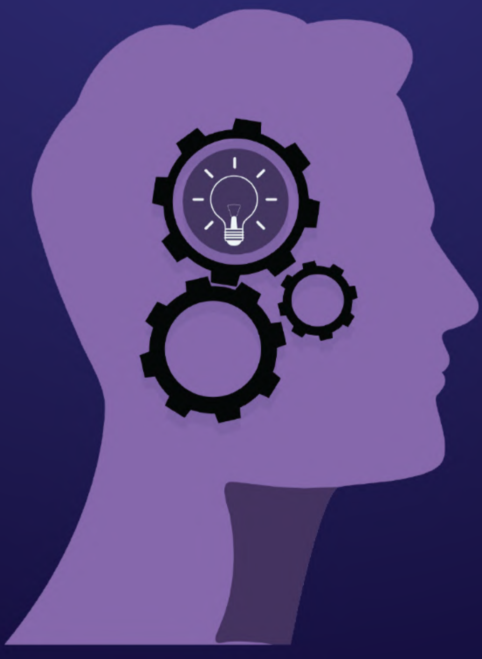
Information theory



بسم الله الرحمن الرحيم

Lecture 04

في البداية كده هنراجع علي اخر حاجة اتكلمنا عنها المحاضرة الي فاتت .

وكان الrelative entropy , mutual information

وطبعا عرفنا ان ال mutual information حالة خاصة من ال relative entropy .

بعدين هنتكلم عن العلاقة بين ال mutual information , relative entropy .

بعدين هنتكلم عن الchain rule for entropy, information , relative entropy .

وبعدين هنعرف امتي نقدر نحصل علي ال maximum entropy .

وهنختم بال Jensens inequality , data-processing inequality .

Relative entropy and mutual information

Mutual information I(X ; Y)

1- is the reduction in uncertainty due to another random variable.

Mutual information => *H(XIY).* بيقلل ال entropy لاني بطرح منه

2- is a measure of the amount of information that one random variable

contains about another random variable.

هجيبهالك بمثال , لو انا قولتلك الجو انهارده معتدل احتمال اني اخرج هتزيد, ولو قولتلك الجو ممطر احتمال اني افضل في البيت هتزيد ,عشان كده بقولك ان بقيس المعلومات الي بيوفرها متغير عشوائي بمعرفة متغير عشوائي اخر

3- *I(X ; Y)* is a measure of the dependence between the two random variables.

لما اقولك مثلا انت لو نجحت هجيبلك هدية هنا في علاقة بين الهدية والنجاح

4- It is symmetric in X and Y and always nonnegative.

Symmetric => *I(X ; Y) = I(Y ; X)*

5- It is equal to zero if and only if X and Y are independent.

I(X ; Y) = H(X) - H(XIY).

Relative Entropy

Relative entropy, or Kullback-Leibler distance D(p||q), measures how different two probability distributions p and q are from each other.

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

↔Relative entropy ≥ and is zero if and only if p = q.

It is not a true distance between distributions because it is neither symmetric nor satisfies the triangle inequality.

1- not symmetric=> $D(p||q) \neq D(q||p)$

2- Triangle Inequality => $D(p||q) + D(q||r) \geq D(p||r)$ وده مش بتحقق فيها

Mutual information /(X; Y) is the relative entropy between the joint distribution and the product distribution p(x)p(y).

$$I(X;Y) = \sum_{x\in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

↔Relative entropy D(p‖q) measures how much extra effort is needed when assuming distribution q instead of the true distribution p

↔D(p || q) measures the extra bits needed to represent the random variables using the q distribution instead of the true distribution p.

Relationship between entropy and mutual information

1- I(X ; Y) = H(X) - H(X|Y)

2- I(X ; Y) = H(Y) - H(Y|X)

3- I (X ; Y) = H(X) + H(Y) - H(X;Y)

4- I(X ; Y) = I(Y ; X)

5- I(X ; X) = H(X)

↔The mutual information of a random variable with itself is its entropy, which is why entropy is also called self-information => I(X ; X) = H(X)

سبحان الله وبحمده سبحان الله العظيم

↔The entropy of a collection of random variables is the sum of the conditional entropies

بقدر احسب Entropy الكلية عن طريق اني بضيف ال Conditional Entropy لكل متغير عشوائي بناءً على معرفتي بالمتغيرات الأخري

$$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_{n-1}, \ldots, X_1)$$

Conditional mutual information => Reduction in uncertainty of *X* due to knowledge of *Y* when *Z* is given

↔$I(X; Y \mid Z) = H(X \mid Z) - H(X \mid Y,Z)$

↔$D(p(x, y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x))$.

Maximum entropy

Uniform distribution has maximum entropy among all distributions

.

Uniform distribution=> كل عناصر المتغير العوائي ليهم نفس البروبيرتي

Uniform distribution قيمة ال Entropyاكبر ما يمكن لما يكون المتغير بتاعي

Conditioning reduces entropy (Information can't hurt) $H(X|Y) \le H(X)$, with equality if and only if X and Y are independent.

طلما عندي معلومات زيادة ال uncertainty هيقل فبالتالي ال entropy هيقل

هيتساوو بقا في حالة ان ال X,Y يكونوا independent

Jensen's inequality

A function f(x) is convex over an interval (a , b) =>

1- if for every x1, *x2* ∊*(a, b)* and 0 ≤ λ ≤ 1 ,

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

2- If the second derivative of a function f is nonnegative (positive) over an interval, then the function is convex (strictly convex) on that interval.

3- A function is concave if it lies above any chord, and convex if it lies below any chord. A function is concave if its negative is convex.

If *f* is a convex function and *X* is a random variable, **E(f (X)) ≥ f(E(X))**

متنساش تصلي علي النبي

لما تجي تعمل applay لميثود معينة في حالة ال convex function علي ال expected value هيكون اقل من او يساوي ال excected value لما تعمل applay علي الميثود نفسها

Information inequality

$D(p||q) \geq 0$, with equality if and only if $p(x) = q(x)$

Nonnegativity of mutual information

$I(X; Y) > 0$, with equality if and only if X and Y are independent.

Data-processing inequality

X ➔ Y ➔ Z is a Markov chain in that order if : The conditional distribution of Z depends only Y and is conditionally independent of X.

Markov chain=> عندي مجموعة من ال States واقدر انتقل من State ل State

بس عشان انتقل لل Next state محتاج اعرف المعلومات الي في ال Current state

ومليش دعوة بال Previous state

↔$p(x, y, z) = p(x)p(y|x)p(z|y)$.

↔Markovity implies conditional independence.

Conditional independence**:** X→Y→Z forms a Markov chain if, and only if, X and Z are conditionally independent given Y.

↔ $p(x, z|y) = p(x|y)p(z|y)$

↔ Past and future are conditionally independent given the present.

↔If X ➔ Y ➔ Z, then $I(X; Y) \geq I(X; Z)$.

$I(X; Y, Z) = I(X; Z) + I(X;Y |Z)$

$= I(X; Y) + I(X; Z|Y)$.

↔ Since X and Z are conditionally independent given Y , we have

$I(X; Z|Y) = 0$. Since $I(X; Y |Z) \geq 0$, we have $I(X; Y) \geq I(X; Z)$.

سبحان الله ... الحمد لله ... لا إله إلا الله ... الله اكبر