INFORMATION CONTENT | SELF - INFORMATION (I)

* Basic quantity derived from probability of particular event occuring from a r.v.

INFORMATION CONTENT | SELF - INFORMATION (I)

* Basic quantity derived from probability of particular event occuring from a r.v.

* 3 AXIOMS

i) Event with 100% probability yields <u>no</u> information

INFORMATION CONTENT | SELF - INFORMATION (I)

* Basic quantity derived from probability of particular event occurring from a r.v.

* 3 AXIOMS

1) Event with 100% probability yields <u>no</u> information

2) Less probable the event, more informatⁿ it yields

INFORMATION CONTENT | SELF - INFORMATION (I)

* Basic quantity derived from probability of particular event occuring from a r.v.

* 3 AXIOMS

1) Event with 100% probability yields <u>no</u> information

2) Less probable the event, more informatⁿ it yields

3) If 2 independent events are measured separately, total information is sum of self-informatⁿ of individual events

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{I(x) = -\log_b [P_r(x)] = -\log_b P}$$

Event x w/ probability P

$b = 2$, $I(x)$ unit = Shannon/bit

$b = e$, $I(x)$ unit = NAT

* 3 AXIOMS

1) Event with 100% probability yields <u>no</u> information

2) Less probable the event, more informat^n it yields

3) If 2 independent events are measured separately, total information is sum of self-informat^n of individual events

INFORMATION CONTENT | SELF-INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_r(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

↓↓↓ ADHERES TO ALL LAWS

* 3 AXIOMS

1) Event with 100% probability yields <u>no</u> information

2) Less probable the event, more informat^n it yields

3) If 2 independent events are measured separately, total information is sum of self-informat^n of individual events

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_r(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

↓ ↓ ↓ ADHERES TO ALL LAWS

i) Event with 100% probability yields ~~no~~ information

Event with 100% probability

$$I(x) = -\log_2 P_r(1) = 0$$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$I(x) = -\log_b [P_r(x)] = -\log_b p$$

Event x w/ probability p

2) Less probable the event, more information it yields

$P(x) = 0$

$\Rightarrow I(x) = -\log_2 (0) \rightarrow \infty$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_X(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

3) If 2 independent events are measured separately, total information is sum of self-informat^n of individual events

2 Independent r.v. X, Y with pmfs $p_X(x)$, $p_Y(y)$

JOINT PMF $\quad p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{I(x) = -\log_b [P_r(x)] = -\log_b p}$$

Event x w/ probability p

3) If 2 independent events are measured separately, total information is sum of self-informat^n of individual events

2 Independent r.v. X, Y with pmfs $p_X(x), p_Y(y)$

JOINT PMF $\quad p_{X,Y}(x,y) = p_X(x) . p_Y(y)$

$$I_{X,Y}(x,y) = -\log_2 [p_{XY}(X=x, Y=y)] = -\log_2 p_X(x) - \log_2 p_Y(y)$$

$$= I_X(x) + I_Y(y)$$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_r(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

Example: FAIR TOSS COIN

$$p_X(H) = p_X(T) = 0.5$$

$$I_X(H) = -\log_2 p_X(H) = -\log_2 \left(\frac{1}{2}\right) = 1 \text{ shannon}$$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_x(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

Example: FAIR TOSS COIN

$p_x(H) = p_x(T) = 0.5$

$I_x(H) = -\log_2 p_x(H) = -\log_2 \left(\frac{1}{2}\right) = 1$ shannon

Example: Biased coin

$p_x(H) = 0.9$; $p_x(T) = 0.1$

$I_x(H) = -\log_2 (0.9) = 0.15$; $I_x(T) = -\log_2 (0.1) = 3.321$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_r(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

Example: TWO FAIR DICE

$X, Y \sim$ Discrete Uniform $[1, 6]$

$I_{X,Y}(X=4, Y=2) = ?$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_r(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

Example: TWO FAIR DICE

$X, Y \sim$ Discrete Uniform $[1, 6]$

$$I_{X,Y}(X=4, Y=2) = ?$$

$$= I_X(X=4) + I_Y(Y=2)$$

$$= -\log_2\left(\frac{1}{6}\right) - \log_2\left(\frac{1}{6}\right)$$

$$= 5.169 \text{ shannons}$$

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b [P_r(x)] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

Example: TWO FAIR DICE

$X, Y \sim$ Discrete Uniform $[1, 6]$

$I_{X,Y}(X = 4, Y = 2) = ?$

$= I_{XY}(X = 4, Y = 2)$

$= -\log_2 \left(\frac{1}{36} \right)$

$= 5.169$ shannons

INFORMATION CONTENT | SELF - INFORMATION (I)

$$\boxed{\begin{array}{l} I(x) = -\log_b \left[P_r(x) \right] = -\log_b p \\ \text{Event } x \text{ w/ probability } p \end{array}}$$

Example: Categorical Distribution

$$P_X(k) = \begin{cases} p_i \,; & k = s_i \in S \\ 0 \,; & \text{o/w} \end{cases}$$

$$S = \{ s_i \}_{i=1}^{N} = \text{Support}$$

$$I_X(x) = -\log_2 P_X(x)$$

ENTROPY

* Basic quantity in Information Theory
* Interpreted as <u>average</u> level of informational surprise

* $H(X) = E[I_X]$

ENTROPY

* Basic quantity in Information Theory
* Interpreted as <u>average</u> level of informational surprise
* $H(X) = E[I_X]$

Example: FAIR TOSS COIN

$$p_X(H) = p_X(T) = 0.5$$

$$I_X(T) = I_X(H) = -\log_2 p_X(H) = -\log_2\left(\frac{1}{2}\right) = 1 \text{ shannon}$$

$$H(X) = p_X(H) * I_X(H) + p_X(T) * I_X(T) = 1$$

ENTROPY

* Basic quantity in Information Theory
* Interpreted as <u>average</u> level of informational surprise
* $H(X) = E[I_X]$

$$= \sum_i P_X(x_i)\, I_X(x_i)$$

$$= -\sum_i P_X(x_i) \log_b P_X(x_i)$$

or $\sum -p_{x_i} \log p_{x_i}$

ENTROPY

$$H(X) = -\sum_{i} P_x(x_i) \log_b P_x(x_i)$$

Dataset

... Temp.	Play
Hot	No
Hot	No
Cold	Yes
Normal	Yes

ENTROPY

$$H(X) = -\sum_i P_x(x_i) \log_b P_x(x_i)$$

Dataset

Temp.	Play
Hot	NO
Hot	NO
Cold	yes
Normal	yes

$$H(Play) = -p_{NO} \log_2 p_{NO} - p_{yes} \log_2 p_{yes}$$

$$= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}$$

$$= 1$$

CONDITIONAL ENTROPY

* CONDITIONAL ENTROPY of Y GIVEN X

$$H(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)}$$

$\mathcal{X}$, $\mathcal{Y}$ are support for X and Y

CONDITIONAL ENTROPY

* CONDITIONAL ENTROPY of Y GIVEN X

$$H(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)}$$

DERIVATION:

$H(Y|X=x)$ = Entropy of Y conditioned on $X=x$

$$= -\sum_{y \in \mathcal{Y}} P(Y=y \mid X=x) \log P(Y=y \mid X=x)$$

CONDITIONAL ENTROPY

* CONDITIONAL ENTROPY of Y GIVEN X

$$H(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)}$$

DERIVATION:

$H(Y|X=x)$ = Entropy of Y conditioned on $X=x$

$$= -\sum_{y \in \mathcal{Y}} \underbrace{P(Y=y \mid X=x)}_{\text{EXPECTATION}} \underbrace{\log P(Y=y \mid X=x)}_{\text{INFORMATION CONTENT}}$$

CONDITIONAL ENTROPY

* CONDITIONAL ENTROPY of Y GIVEN X

$$H(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)}$$

DERIVATION:

$H(Y|X=x)$ = Entropy of Y conditioned on $X=x$

$$= -\sum_{y \in \mathcal{Y}} P(Y=y|X=x) \log P(Y=y|X=x)$$

$$\therefore H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X=x)$$

CONDITIONAL ENTROPY

DERIVATION:

$H(Y|X=x)$ = Entropy of Y conditioned on $X=x$

$$= -\sum_{y \in \mathcal{Y}} P(Y=y|X=x) \log P(Y=y|X=x)$$

$$\therefore H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X=x)$$

$$= \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$$

CONDITIONAL ENTROPY

DERIVATION:

$H(Y|X=x)$ = Entropy of Y conditioned on $X=x$

$$= -\sum_{y \in \mathcal{Y}} P(Y=y|X=x) \log P(Y=y|X=x)$$

$$\therefore H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X=x)$$

$$= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$$

$$= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(y|x)$$

CONDITIONAL ENTROPY

DERIVATION:

$$H(Y \mid X = x) = \text{Entropy of } Y \text{ conditioned on } X = x$$

$$= -\sum_{y \in \mathcal{Y}} P(Y = y \mid X = x) \log P(Y = y \mid X = x)$$

$$\therefore H(Y \mid X) = \sum_{x \in \mathcal{X}} p(x)\, H(Y \mid X = x)$$

$$= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y \mid x) \log p(y \mid x)$$

$$= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y \mid x)$$

$$H(Y \mid X) = -\sum_{x} \sum_{y} p(x, y) \log \left(\frac{p(x, y)}{p(x)} \right)$$

CONDITIONAL ENTROPY

CHAIN RULE

$$H(Y|X) = -\sum_x \sum_y p(x,y) \log \frac{(p(x,y))}{p(x)}$$

$$= -\sum_x \sum_y p(x,y) \log p(x,y) + \sum_x \sum_y p(x,y) \log p(x)$$

$$= H(X,Y) + \sum_x \left(\sum_y p(x,y) \right) \log p(x))$$

$$= H(X,Y) + \sum_x p(x) \log p(x)$$

$$\boxed{H(Y|X) = H(X,Y) - H(X)}$$

CONDITIONAL ENTROPY AND INFORMATION GAIN

Dataset

Temp.	Play
Hot	No
Hot	No
Cold	Yes
Normal	Yes

$$H(\text{Play}) = -p_{No}\log_2 p_{No} - p_{Yes}\log_2 p_{Yes}$$

$$= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$$

$$= 1$$

$$H(\text{Play} \mid \text{Temp}) = ?$$

CONDITIONAL ENTROPY AND INFORMATION GAIN

Dataset

... Temp.	Play
Hot	NO
Hot	NO
Cold	yes
Normal	yes

$$H(\text{Play}) = -p_{No}\log_2 p_{No} - p_{yes}\log_2 p_{yes}$$

$$= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$$

$$= 1$$

$$H(Y|X) = \sum_{x\in\mathcal{X}} p(x)\, H(Y|X=x)$$

$$\Rightarrow H(\text{Play}\mid\text{Temp}) = p(\text{Temp}=\text{Hot})\,H(\text{Play}\mid\text{Temp}=\text{Hot}) + p(\text{Temp}=\text{Cold})\,H(\text{Play}\mid\text{Temp}=\text{Cold}) + \ldots\ p(\text{Temp}=\text{Normal})\,H(\text{Play}\mid\text{Temp}=\text{Normal})$$

CONDITIONAL ENTROPY AND INFORMATION GAIN

Dataset

Temp.	Play
Hot	NO
Hot	NO
Cold	Yes
Normal	Yes

$$H(\text{Play}) = -p_{No}\log_2 p_{No} - p_{Yes}\log_2 p_{Yes}$$

$$= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$$

$$= 1$$

INFO. GAIN ON Play(Y) having observed

Temp (X) = Entropy (Play) − Entropy (Y|X)