# Entropy

*Measure of uncertainty or randomness in a distribution*

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

**Terms:**

- $X$: Random variable
- $p(x_i)$: Probability of outcome $x_i$

**Use-case:** Decision trees, Random Forests, among others...

@AIinMinutes

# Gini Impurity

*Used in place of Entropy to speed up; caveat: tyranny of the majority*

$$G(X) = 1 - \sum_{i=1}^{n} p(x_i)^2$$

**Terms:**

- $X$: Random variable
- $p(x_i)$: Probability of outcome $x_i$

@AIinMinutes

# KL Divergence

*Measures how one probability distribution diverges from another*

$$D_{KL}(P||Q) = \sum_i P(x_i) \log \frac{P(x_i)}{Q(x_i)}$$

**Terms:**

- $P(x_i)$: Target probability distribution
- $Q(x_i)$: Approximating probability distribution

**Use-case:** Variational inference, among others...

@AIinMinutes

# Jensen's Inequality

*Used to derive Evidence Lower Bound (ELBO)*

$$f(E[X]) \leq E[f(X)]$$

**Terms:**

- $f(x)$: Convex function
- $E[X]$: Expected value of random variable X
- For concave functions, inequality is reversed

**Use-case:** Variational inference, Inferential statistics

@AIinMinutes

# Jensen-Shannon Divergence

*Symmetric and smoothed version of KL divergence*

$$JSD(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M)$$

**Terms:**

- $P, Q$: Two probability distributions
- $M = \frac{1}{2}(P + Q)$: Mixture distribution

@AIinMinutes

# Conditional Entropy

*Measures remaining uncertainty in X given Y*

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

**Terms:**

- $p(x,y)$: Joint probability
- $p(x|y)$: Conditional probability

**Use-case:** Information gain calculation

@AIinMinutes

# Mutual Information

*Measures information shared between two random variables*

$$I(X;Y) = D_{KL}(P_{XY} || P_X \otimes P_Y)$$

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

**Terms:**

- $P_{XY}$: Joint distribution of X and Y
- $P_X \otimes P_Y$: Product of marginal distributions

**Use-case:** Feature slection

@AIinMinutes

# Normalized Mutual Information

*Mutual information normalized to [0,1] range*

$$NMI(X;Y) = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}}$$

**Terms:**

- $I(X;Y)$: Mutual information
- $H(X), H(Y)$: Entropies of X and Y

@AIinMinutes

# Cross Entropy

*Measures how well a model's predicted distribution matches the true distribution*

$$H(P, Q) = -\sum_{i} P(x_i) \log Q(x_i)$$

**Terms:**

- $P(x_i)$: True probability distribution
- $Q(x_i)$: Predicted probability distribution

**Use-case:** Loss function in classification problems, Perplexity = exp(CE) used for LLM evaluation

@AIinMinutes

# Fisher Information

*Measures how much information a random variable carries about a parameter of its distribution*

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right]$$

**Terms:**

- $\theta$: Parameter of interest
- $f(X;\theta)$: Probability density function
- $E[\cdot]$: Expected value

**Use-case:** Natural gradient descent, CRLB (Only Stats. people can understand ;), among others...

@AIinMinutes