Alan Turing Course Summary

[*] Alan Turing: Key Contributions

- Turing Machines: Abstract computational devices that laid the foundation for computer science.

- Turing Test: A test to determine if a machine can exhibit intelligent behavior indistinguishable from a human.

- Cracking the Enigma Code: During WWII, Turing helped break the German Enigma code using statistical methods and the bombe machine, drastically reducing the number of tests needed through probabilistic reasoning.

[#] CAPTCHA

- Acronym for: Completely Automated Public Turing test to tell Computers and Humans Apart

- Used to differentiate humans from bots in online systems, inspired by Turing's ideas.

[P] Probability & Inference

- Conditional Probability: Probability of an event given another has occurred.

- Bayesian Methods: Used in Turing's codebreaking for updating probabilities based on new evidence.

- Pearson Correlation Coefficient: Measures the strength of a linear association (ranging from -1 to 1).

[C] Causal Inference & Bias

- Correlation does not imply causation.

- Judea Pearl's Work: Differentiates correlation from causality; emphasizes causal diagrams.

- Confounding Variables: Hidden variables that distort apparent relationships between studied variables.

[E] Experimental Design

- Randomized Controlled Trials (RCTs): Gold standard for establishing causal effects.

- Mediating & Moderating Variables:

Alan Turing Course Summary

- Mediator: Explains the relationship between two variables.

- Moderator: Alters the strength/direction of this relationship.

- Confounder: A third variable that creates a false association between two others.

[D] Data Science & Philosophy

- CRISP-DM: Data mining methodology with iterative cycles of modeling and testing.

- Leo Breiman vs. Traditional Science:

  - Breiman supports data-first, predictive modeling ("black box").

  - Others prefer hypothesis-driven, interpretable models.

[!] Statistical Pitfalls

- Spurious Correlations: Random patterns that may appear meaningful.

- Multiple Testing Problem: Increases the chance of finding false positives.

- Overfitting & Cheating in ML Models: Models may 'cheat' by using irrelevant clues in the data.