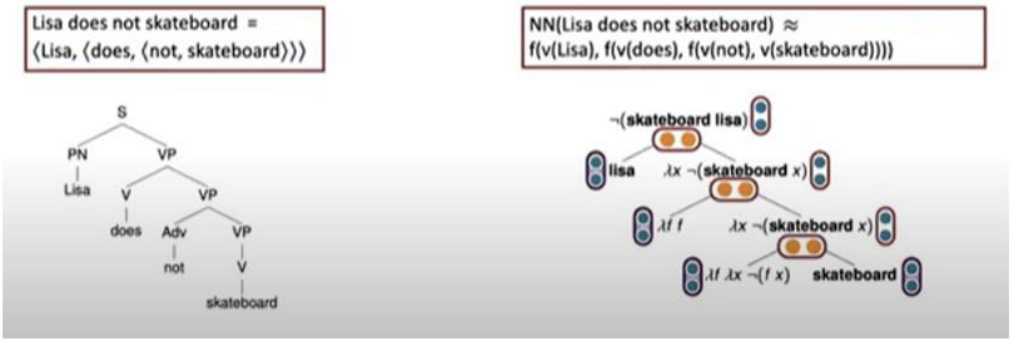
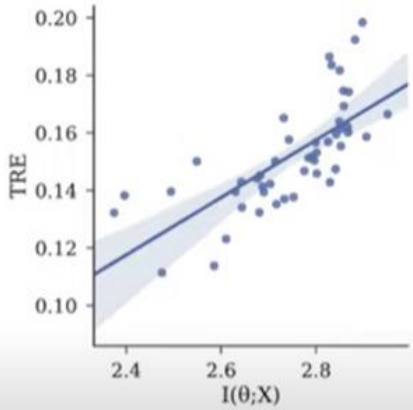
1. Extremely large models and GPT3
    - What's new about GPT-3
        - better than other models at language modeling and related tasks such as story completion
        - in the few-shot setting, we can clarify this by proving an example as a prompt
        - Apart from better language modeling, scaling also helps with knowledge intensive tasks and reading comprehension
        - List
    - Flexible "in-context" learning
        - GPT-3 demonstrates some level of fast adaptation to completely new tasks. This happens via "in-context" learning
        - The language model training is learning how to learn from the context [inner loop]
    - Limitations and Open questions
        - Seems to do poorly on more structured problems that involve decomposing into atomic / primitive skills: RTE / arithmetic / word problems / analogy making
        - Performing permanent knowledge updates interactively is not well studied
        - Doesn't seem to exhibit human like generalization (systematicity)
        - Language is situated and GPT-3 is merely learning from text without being exposed to other modalities
    - Copycat analogy problems
2. Compositional Representations and Systematic Generalization
    - Compositionality: closely related to the idea of systematicity is the principle of compositionality.
    - According to Montague, compositionality is about the existence of a homomorphism from syntax to semantics
    - Tree Reconstruction Error(TRE)
        - Compositionality of representations is about how well the representation approximates an explicitly homomorphic function in a learnt representation space



Are neural representations compositional? [Andreas 2019]

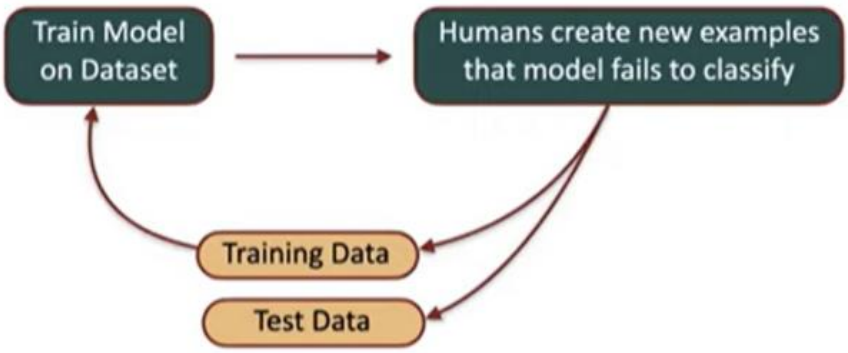


- **This graph plots the mutual information between the input and the representation I(θ; X) against TRE.**

- Maximize compound divergence to create challenging train / test splits!
  - Atoms: primitive elements (entity words, predicates)
  - Compounds: compositions of primitive elements.

3. Improving how we evaluate models in NLP

   instead of testing models on static venchmarks, evaluate on an ever changing dynamic benchmark



Overview of dynamic benchmarks

- main challenges : ensuring that humans are able to come up with hard examples and we are not limited by creativity
- current approaches use examples from other datasets for the same task as prompts

Grounding Language to other modalities

- Many have articulated the need for using modalities other than text
- Bender and Koller [2020]: Impossible to acquire "meaning" (communicative intent of the speaker) from form (text / speech signal) alone
- Bisk et al [2020]: Training on only web-scale data limits the world scope of models.