



Discussion

Jose Hernandez-Orallo^{1,2}

Take-aways

- ❑ Capability is not average performance
 - ❑ Capability allows to predict performance, at the instance level and even OOD
- ❑ System Capabilities and Task Demands are related through a “margin”
- ❑ Measurement layouts capture domain knowledge and intuitions
- ❑ Backward inference: estimate capabilities for one single system
- ❑ Forward inference: infer performance for new task instances.

Discussion

- ❑ How could capability-oriented evaluation benefit your own work?
- ❑ How can this approach be applied when there are too many relevant features?
- ❑ What if no meta-features are given?
- ❑ What can be done in scenarios where domain knowledge is limited?
- ❑ How can these limitations be addressed?