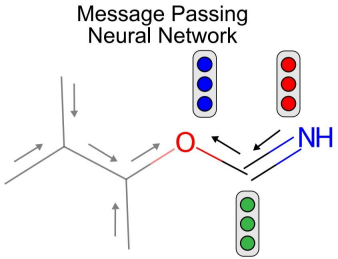
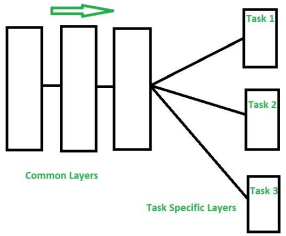
Team np.random

It's not just the organizers, even we used randomness

Approach 1 - Directed MPNN

- Used a DMPNN (Directed MPNN) to predict the 4 properties
  - Since there are only 5000 datapoints, we can't have a separate huge MPNN for each property
  - We have to utilize the fact that an MPNN can utilize common information of the molecules across the 4 properties
  - There are 3 common MPNN timesteps for all the properties
  - The final messages from the shared MPNN are passed to individual task-wise MPNN
  - The individual task-wise MPNN features are used to make the predictions
- Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., … Barzilay, R. (2019). Analyzing Learned Molecular Representations for Property Prediction. Journal of Chemical Information and Modeling, 59(8), 3370–3388. doi:10.1021/acs.jcim.9b00237





Approach 2 - XGBoost on Mordred Descriptors

- Used Mordred to calculate the descriptors for each molecule
  - Mordred descriptors are ~1600 length vectors that contain structural information about the molecule
- These Mordred features are used to predict the properties with XGBoost
  - XGBoost (Extreme Gradient Boosting) uses gradient boosting to learn an ensemble of random forests
  - XGBoost is the top performer for many properties in the TDC(Therapeutic Data Commons) leaderboard
- Our XGBoost implementation did not work as well since we didn't do any hyperparameter tuning of the algorithm

- Tian, H., Ketkar, R., & Tao, P. (2022). Accurate ADMET Prediction with XGBoost. arXiv preprint arXiv:2204.07532.