Introduction: Machine learning for structure-based virtual screening

CS/CME/Biophys/BMI 371

Jan. 30, 2018

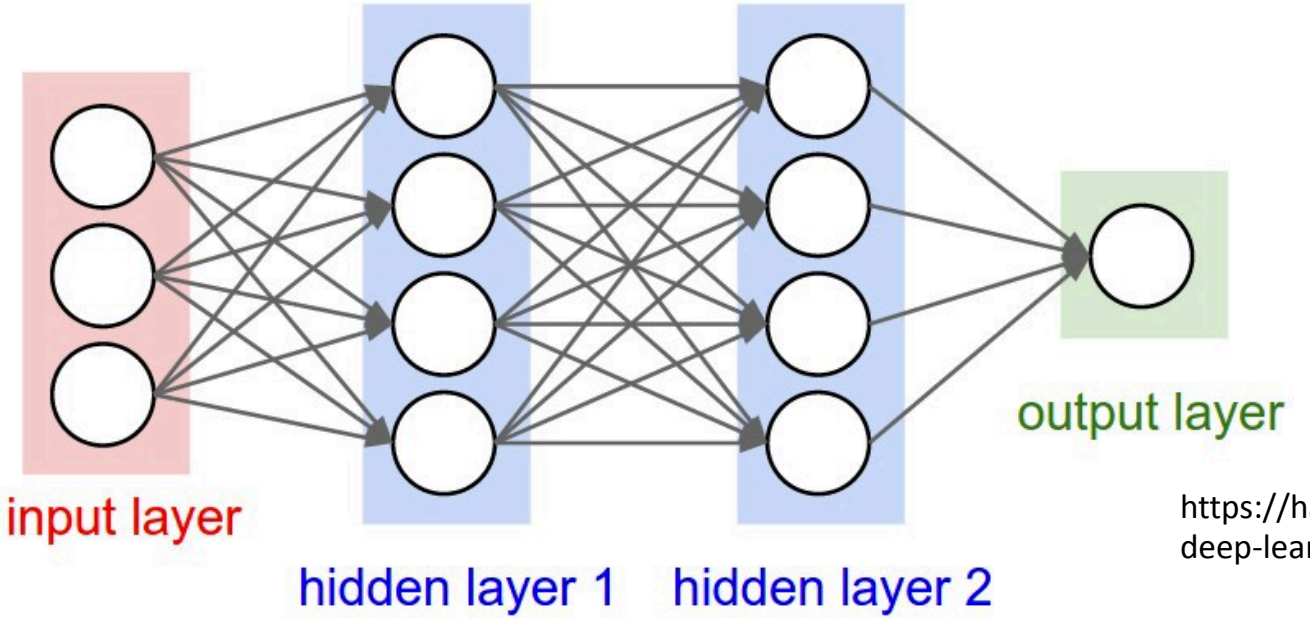
Ron Dror

Alternative to physics-based approach: machine learning

- Forget the physics and treat this instead as a statistical pattern recognition problem
- Learn from existing structures of protein-ligand complexes (and, perhaps, data on affinities of many ligands to proteins)

Deep learning

A traditional neural network



https://hackernoon.com/challenges-in-deep-learning-57bbf6e73bb

- Basic idea of deep learning: just add more hidden layers (and then train on large data sets)
- Thursday's papers rely on convolutional neural networks, which are particularly useful for spatial data sets

Why did I pick these three papers?

- Wallach et al. (from Atomwise, Inc.)
  - First paper applying deep learning to structure-based virtual screening
  - Tough to read and understand
- Gonczarek et al.
  - Also tough to read and understand
  - Intriguing point: machine learning is awfully good at cheating in this context!
- Ragoza et al.
  - Recent paper in a peer-reviewed journal
  - More complete description

Background material

- Ligand docking slides from CS/CME/BioE/ Biophys/BMI 279:
  - http://web.stanford.edu/class/cs279/lectures/lecture7.pdf
- Deep learning tutorial
  - http://ufldl.stanford.edu/tutorial/

Please provide feedback on presentations!

- Anonymous survey to fill out at end of each class: cs371.stanford.edu/feedback.html

- Feedback won't affect grades. The goal is to help students improve future presentations.

- We're trying this in response to student suggestions. Thank you!