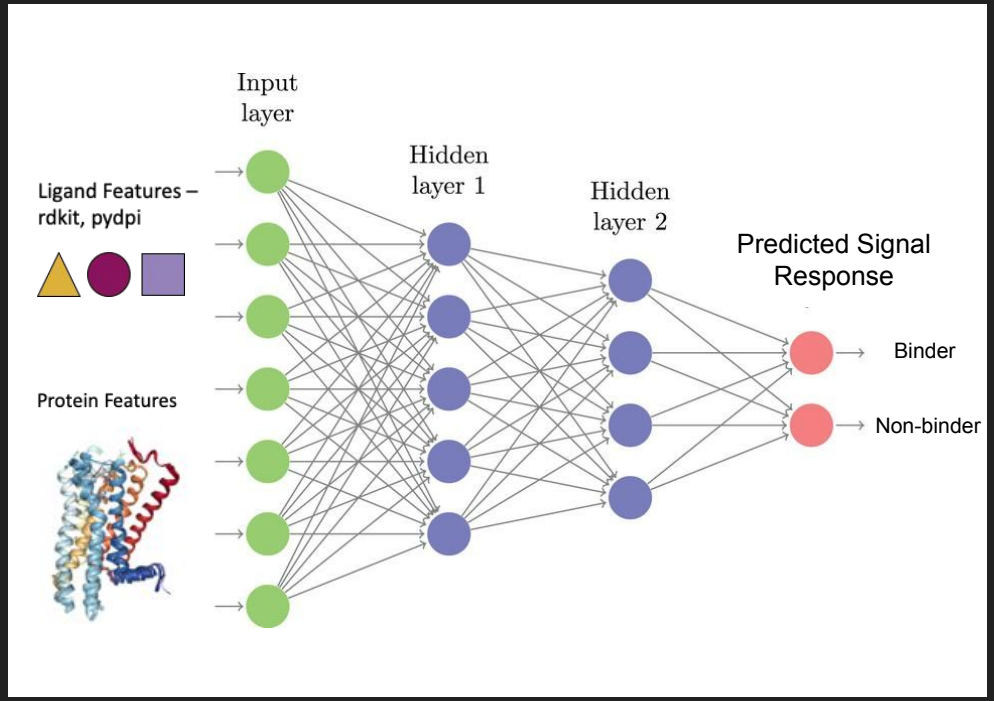
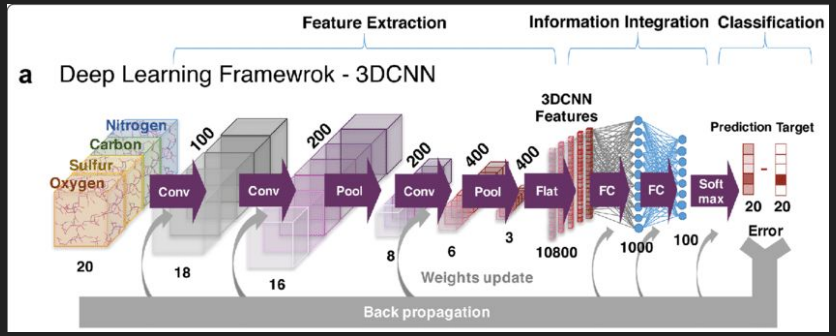
Technology Review

Machine Learning Frameworks

1. Predicting Protein-Ligand Binding





- Protein-ligand binding prediction
  - Drug discovery
  - Unsolved
- We require a machine learning API that is flexible, deep and parallelizable

2. Technologies Considered

Constraints to be considered:

- Ease of Deployment
- Level of Abstraction
- Visualization Options
- Debugging Flexibility

PyTorch

TensorFlow

Keras

Other available libraries:

- MXNet
- Theano
- CNTK

3. Our Choice: PyTorch

PyTorch as a scientific computing package serves two major purposes:

- A replacement for NumPy to use the power of GPUs.
- A deep learning research platform with focus on speed and flexibility.

```
import torch
```

It provides modules and classes to create and train neural networks such as torch.nn , torch.optim , Dataset , and DataLoader.

4. Appeal of Choice

- Very flexible
- Well documented, supported by Facebook, rapidly growing ecosystem
- Very Pythonic (intuitive, readable)
- Comparable in speed to TF
- 3D-CNNs package in PyTorch

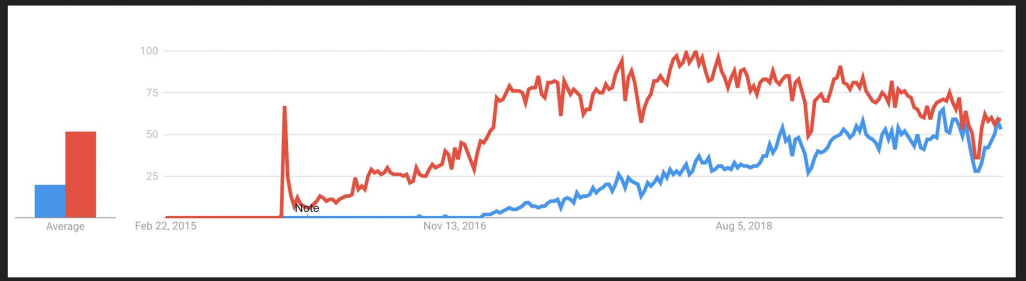
Pythonic manipulation of tensors and layers

```
x = torch.rand(5, 3)
print(x)
```

Out:

```
tensor([[0.8987, 0.7482, 0.2768],
        [0.0514, 0.4693, 0.2699],
        [0.5284, 0.9734, 0.2294],
        [0.0184, 0.9836, 0.8537],
        [0.3716, 0.6091, 0.1335]])
```

Google Search Interest in Tensorflow and PyTorch the past 5 years – U.S.



5. Drawbacks of Choice

- Relatively new (2017)
- High flexibility is potentially confusing/intimidating
- Minimal model visualization tools for tracking and debugging

Examples of TF's "TensorBoard"