# Spetrum Prediction with DNN

#### Machine learning

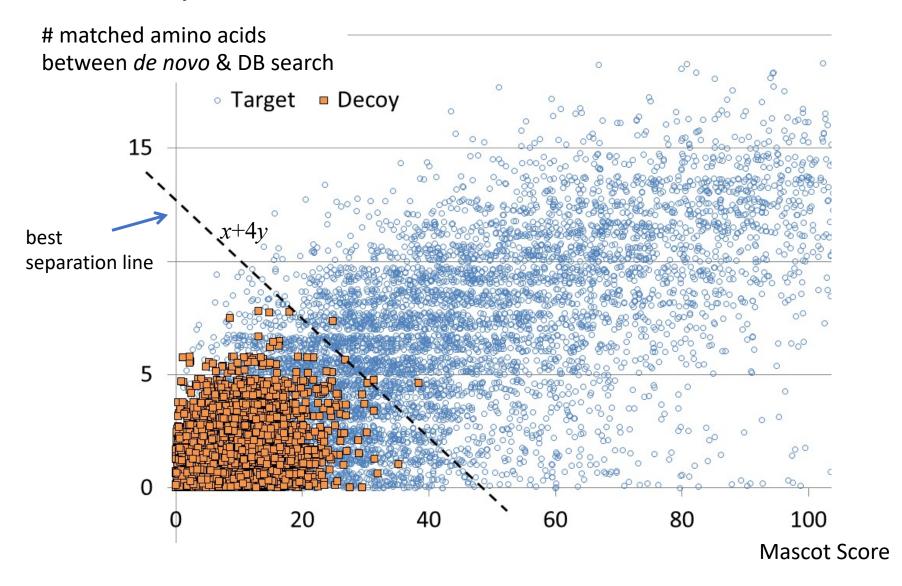
- We've briefly used machine learning twice now:
  - Use a decision tree to score an amino acid in a de novo sequence.
  - Combine multiple score features in database search.

Use a decision tree to determine the amion acid score in de novo sequencing.

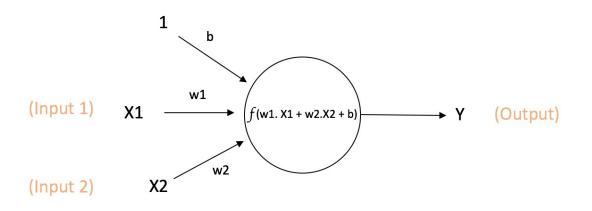
Are both left and right y-ions observed and mass error < 3/8 error tolerance? Are both left and right b-ions observed and mass error < 15/16 error tol.? Does left b-ion overlap with some y-ion? NIST Spectrum Library Is left y-ion abundant (half rank < 32)? 340,000 spectra Yes: 85% Is right y-ion very abundant Is it a proline? (half rank < 16)? Yes: 909 No: 78% Yes: 80% No: 50% Is right y-ion significant in its Is it a proline? neighborhood (local rank < 8)? ... Yes: 79% Is left y-ion very significant in its neighborhood (local rank < 4)? Yes: 91% No: 61%

> 169 features 14,000 internal nodes average depth 18.4

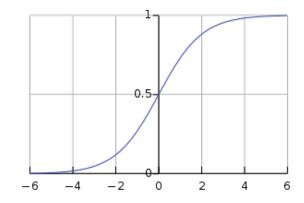
## Use multiple features in PSM score



## Combine Multiple Features Together



Output of neuron = Y= f(w1. X1 + w2. X2 + b)



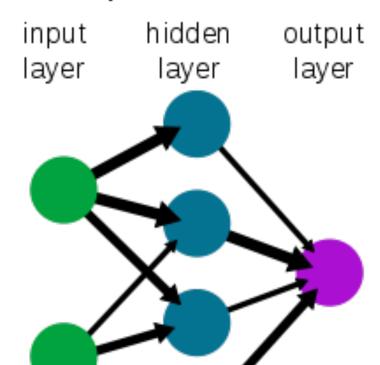
$$y = \frac{1}{1 + e^{-z}}$$

**Activiation function** 

- Train the coefficients w and b to maximize the separation of true and false data points.
- For training, a cost function is defined and there are optimization algorithms to minimize the cost.
- Under certin cost function and sigmoid activiation, this is equivalent to the logistic regression.

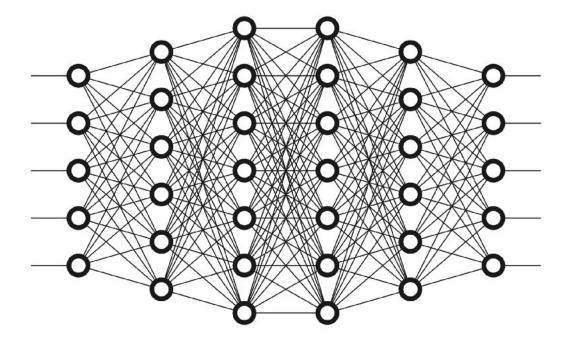
#### Neural Network

#### A simple neural network



- A neural network just combines many neuons together to fit a more complex nonlinear function.
- Often these neurons are organized in layers.
- The algoritm for training is usually the so-called backpropagation algorithm.
- Main idea behind backpropagation is gradient descent implemented in a nice way.

#### Deep Neural Network



- DNN is just a neural network with MANY layers.
- MANY coefficients (parameters) to train.
- Require new training algorithms to both learn fast and avoid overfitting.
- GPU, big data, and new learning algorithm contribute to the development.