

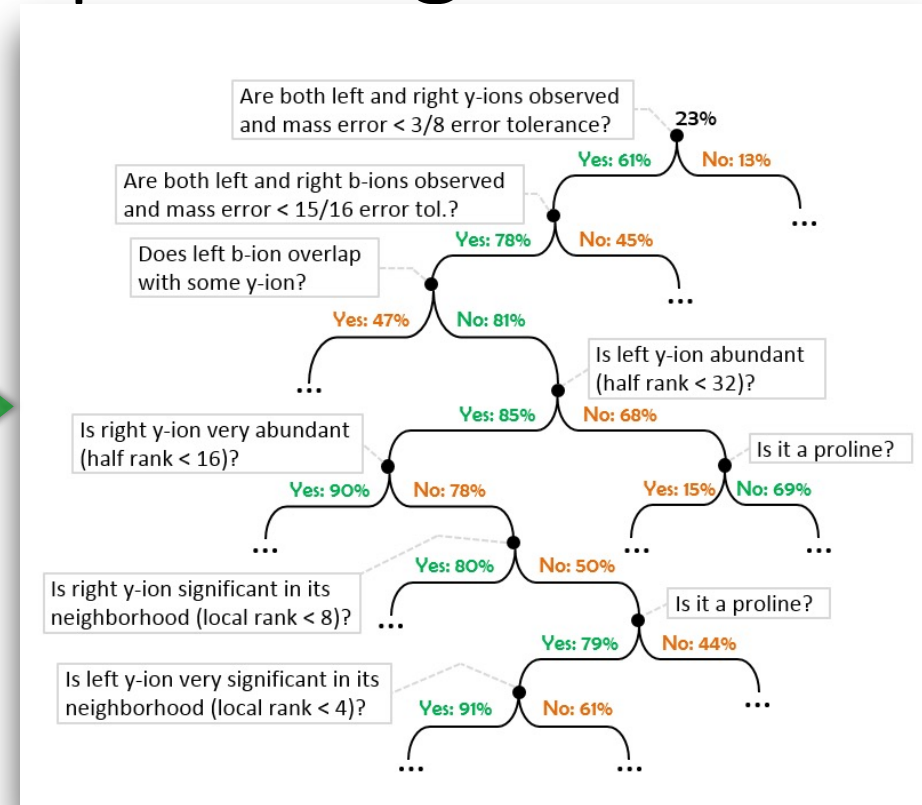
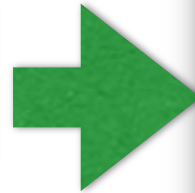
# Spectrum Prediction with DNN

# 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.

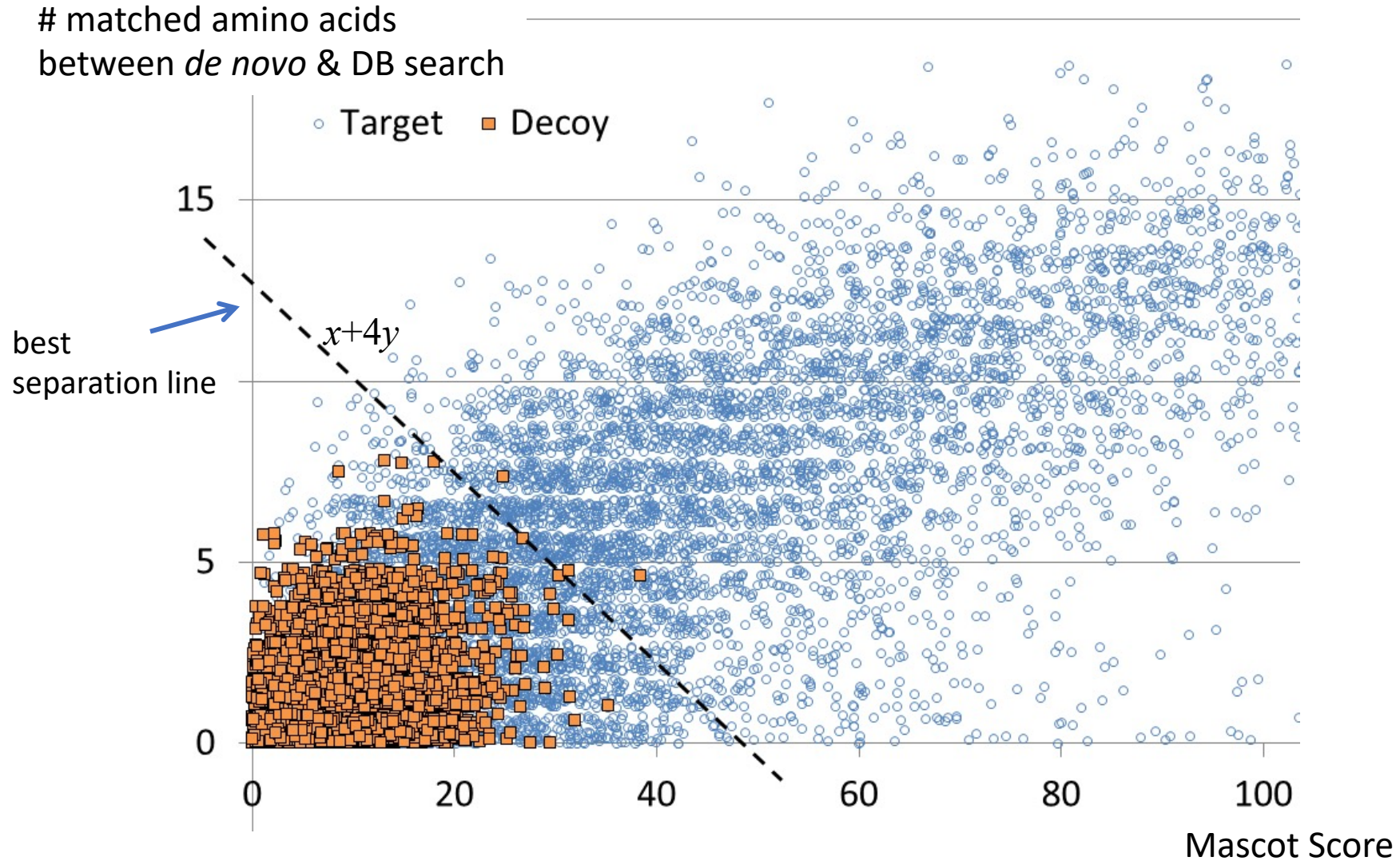
# Use a decision tree to determine the amino acid score in de novo sequencing.

*NIST Spectrum Library*  
*340,000 spectra*

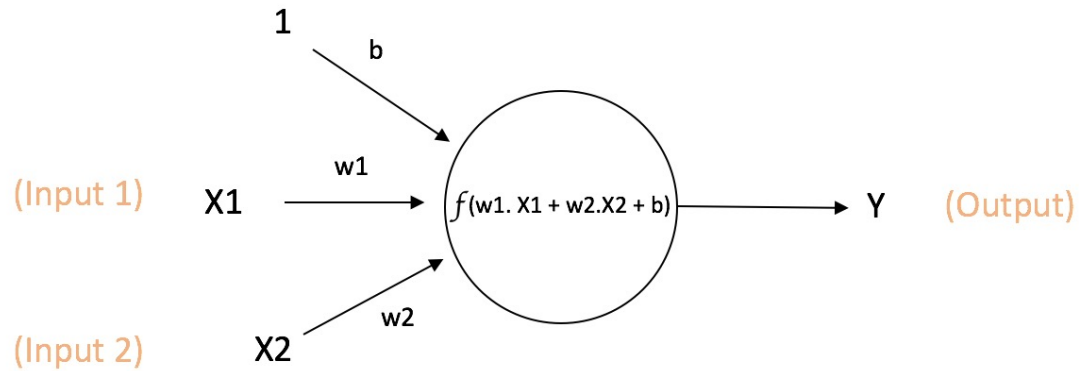


169 features  
14,000 internal nodes  
average depth 18.4

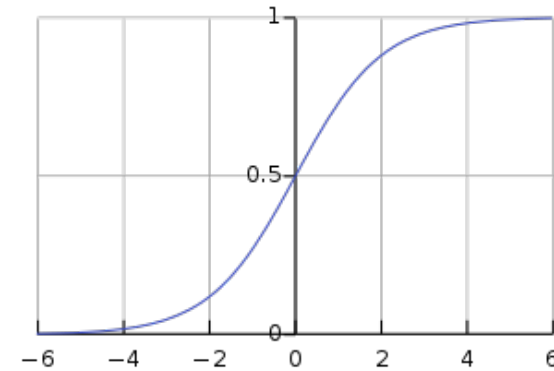
# Use multiple features in PSM score



# Combine Multiple Features Together



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



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

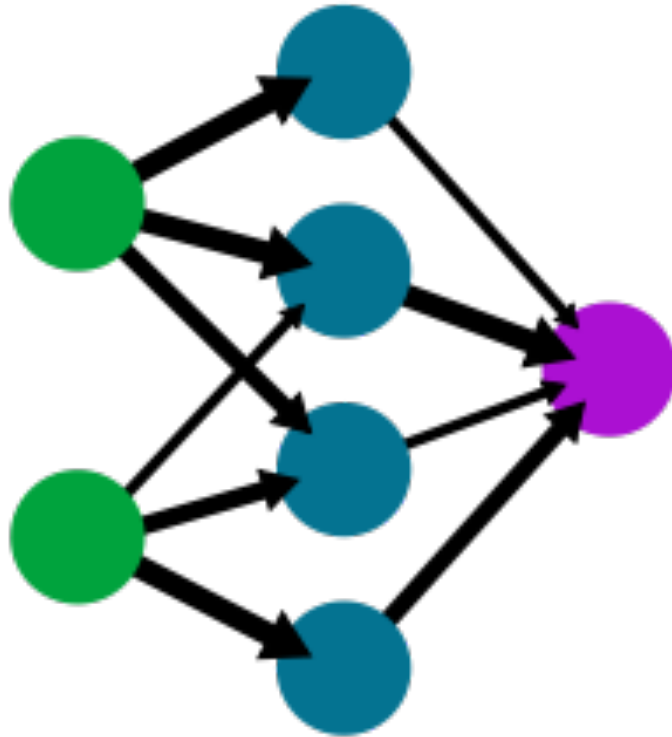
Activation function

- 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 certain cost function and sigmoid activation, this is equivalent to the logistic regression.

# Neural Network

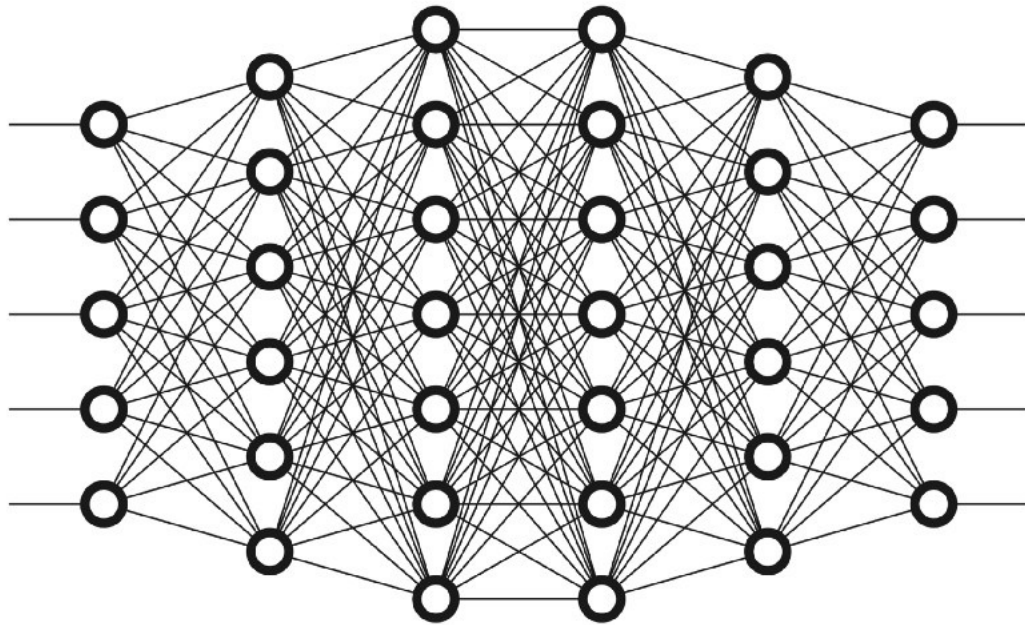
## A simple neural network

input layer      hidden layer      output layer



- A neural network just combines many neurons together to fit a more complex nonlinear function.
- Often these neurons are organized in layers.
- The algorithm 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.