# Multi-Class Neural Networks



Earlier, you encountered binary classification models that could pick between one of two possible choices, such as whether:

* A given email is spam or not spam.
* A given tumor is malignant or benign.

In this module, we'll investigate **multi-class** classification, which can pick from multiple possibilities. For example:

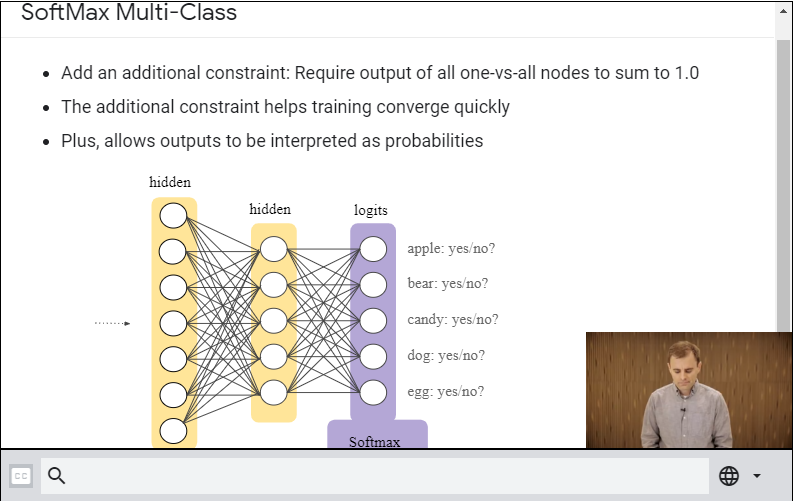
* Is this dog a beagle, a basset hound, or a bloodhound?
* Is this flower a Siberian Iris, Dutch Iris, Blue Flag Iris, or Dwarf Bearded Iris?
* Is that plane a Boeing 747, Airbus 320, Boeing 777, or Embraer 190?
* Is this an image of an apple, bear, candy, dog, or egg?

Some real-world multi-class problems entail choosing from millions of separate classes. For example, consider a multi-class classification model that can identify the image of just about anything.

**More than two classes?**

* Logistic regression gives useful probabilities for binary-class problems.
  + spam / not-spam
  + click / not-click
* What about multi-class problems?
  + apple, banana, car, cardiologist, ..., walk sign, zebra, zoo
  + red, orange, yellow, green, blue, indigo, violet
  + animal, vegetable, mineral

**SoftMax Multi-Class**



**What to use When?**

* Multi-Class, Single-Label Classification:
  + An example may be a member of only one class.
  + Constraint that classes are mutually exclusive is helpful structure.
  + Useful to encode this in the loss.
  + Use one softmax loss for all possible classes.
* Multi-Class, Multi-Label Classification:
  + An example may be a member of more than one class.
  + No additional constraints on class membership to exploit.
  + One logistic regression loss for each possible class.

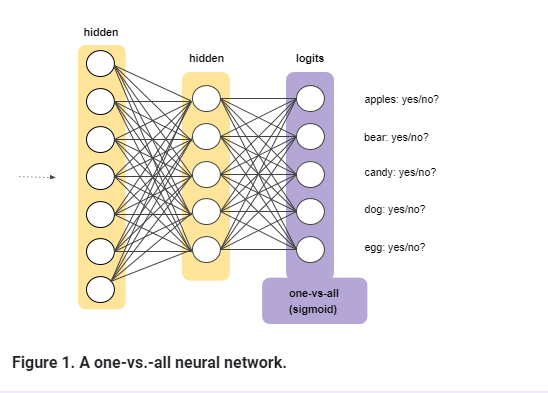
# Multi-Class Neural Networks: One vs. All

 **One vs. all** provides a way to leverage binary classification. Given a classification problem with N possible solutions, a one-vs.-all solution consists of N separate binary classifiers—one binary classifier for each possible outcome. During training, the model runs through a sequence of binary classifiers, training each to answer a separate classification question. For example, given a picture of a dog, five different recognizers might be trained, four seeing the image as a negative example (not a dog) and one seeing the image as a positive example (a dog). That is:

1. Is this image an apple? No.
2. Is this image a bear? No.
3. Is this image candy? No.
4. Is this image a dog? Yes.
5. Is this image an egg? No.

This approach is fairly reasonable when the total number of classes is small, but becomes increasingly inefficient as the number of classes rises.

We can create a significantly more efficient one-vs.-all model with a deep neural network in which each output node represents a different class. The following figure suggests this approach:



SoftMax:

