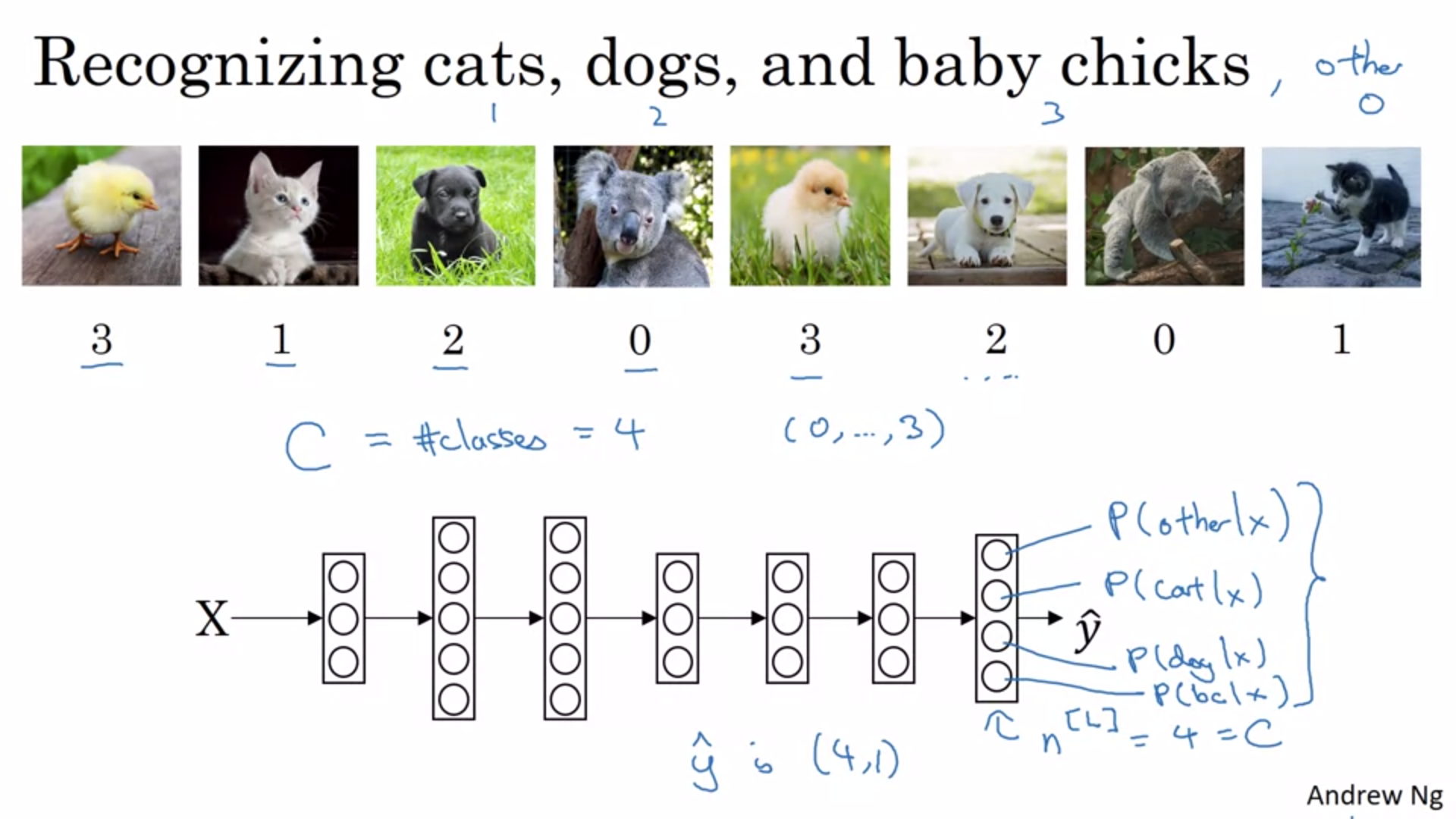
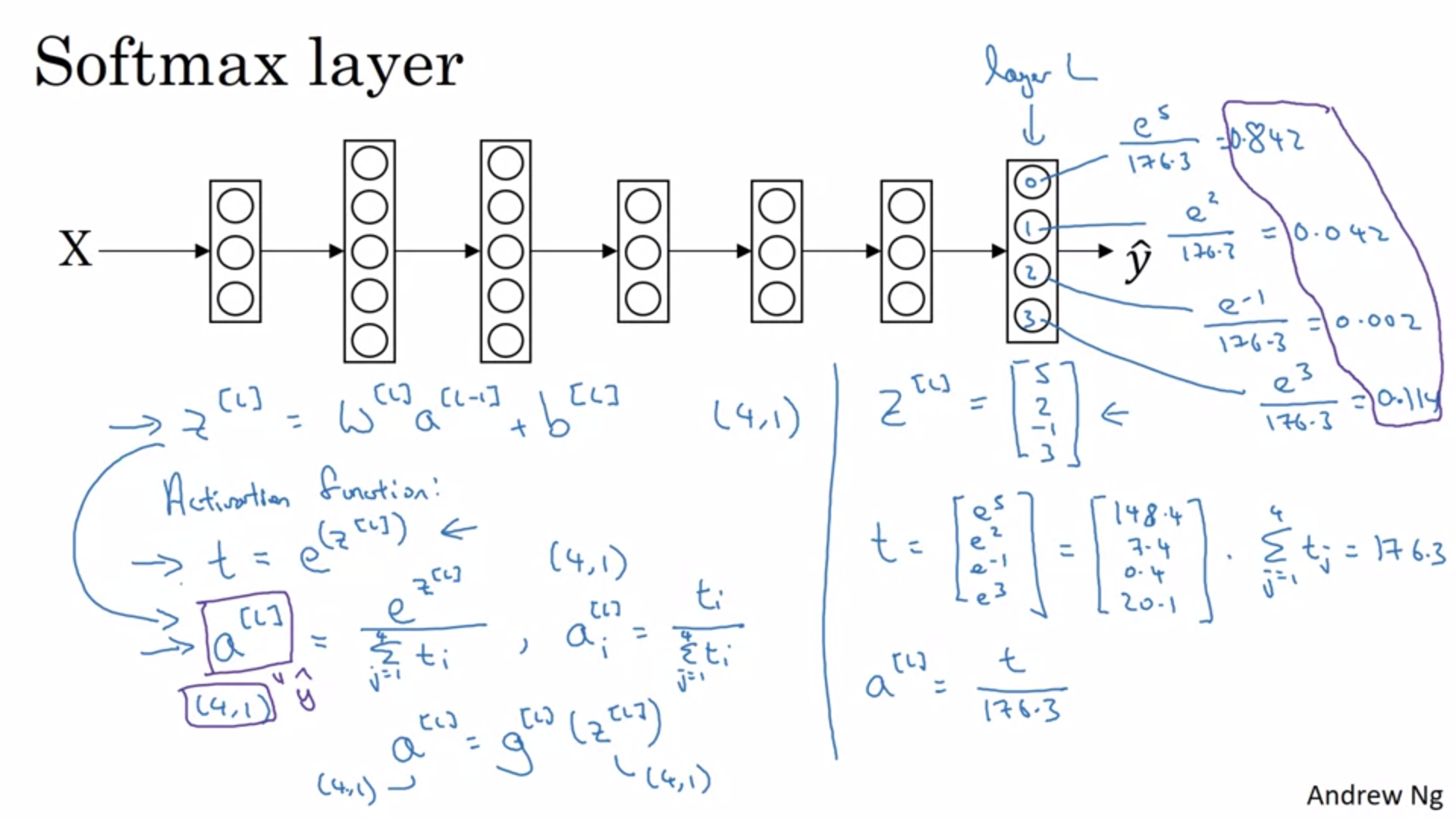


So far we have looked at binary classification. What if you have multiple classes? Softmax can be used for this.



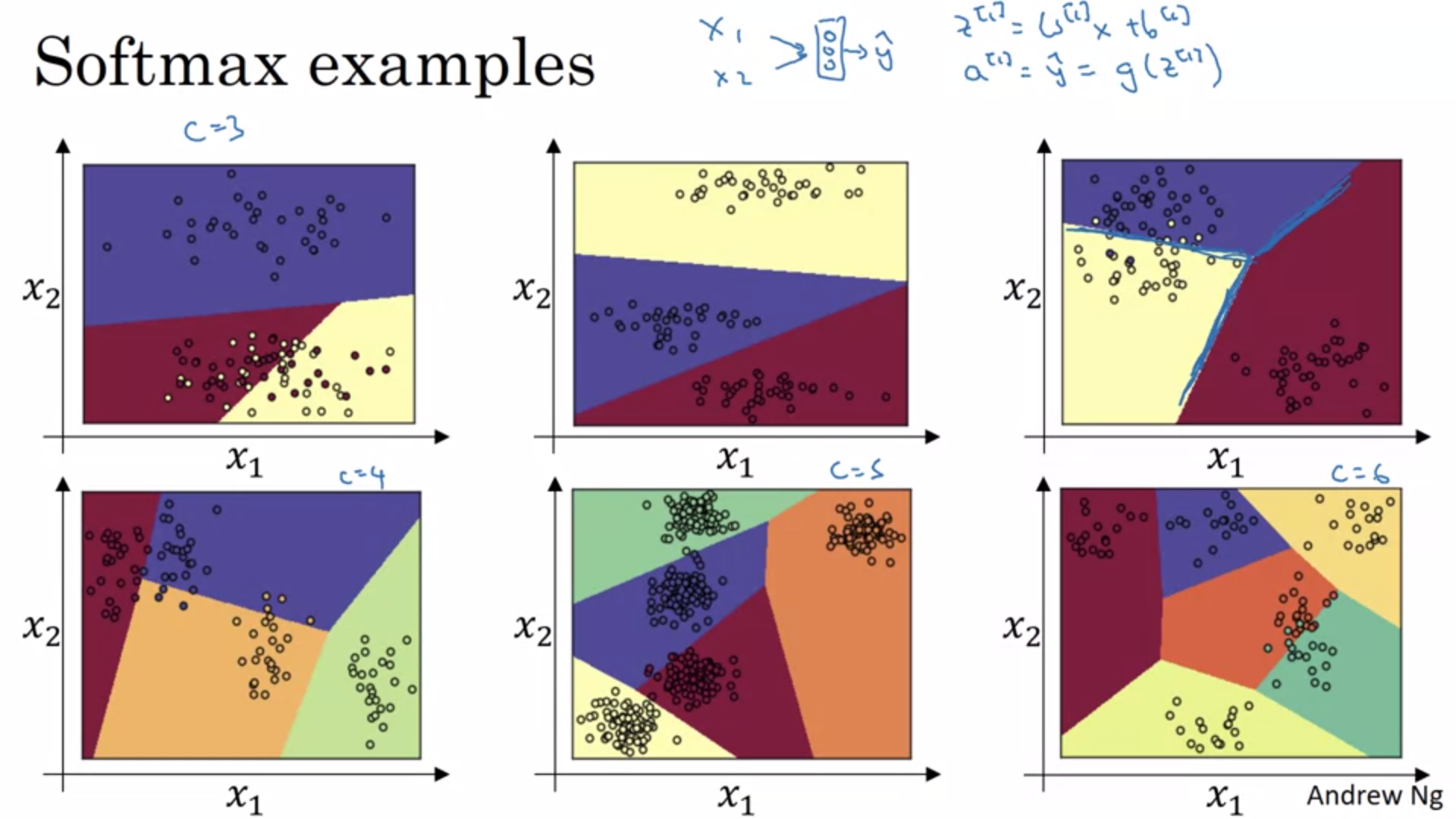
The output layer has now 4 units. The probabilities should sum to 1.



Temporary variable t: element-wise exponentiation.

Normalize to 1 by dividing by the sum of all t’s.

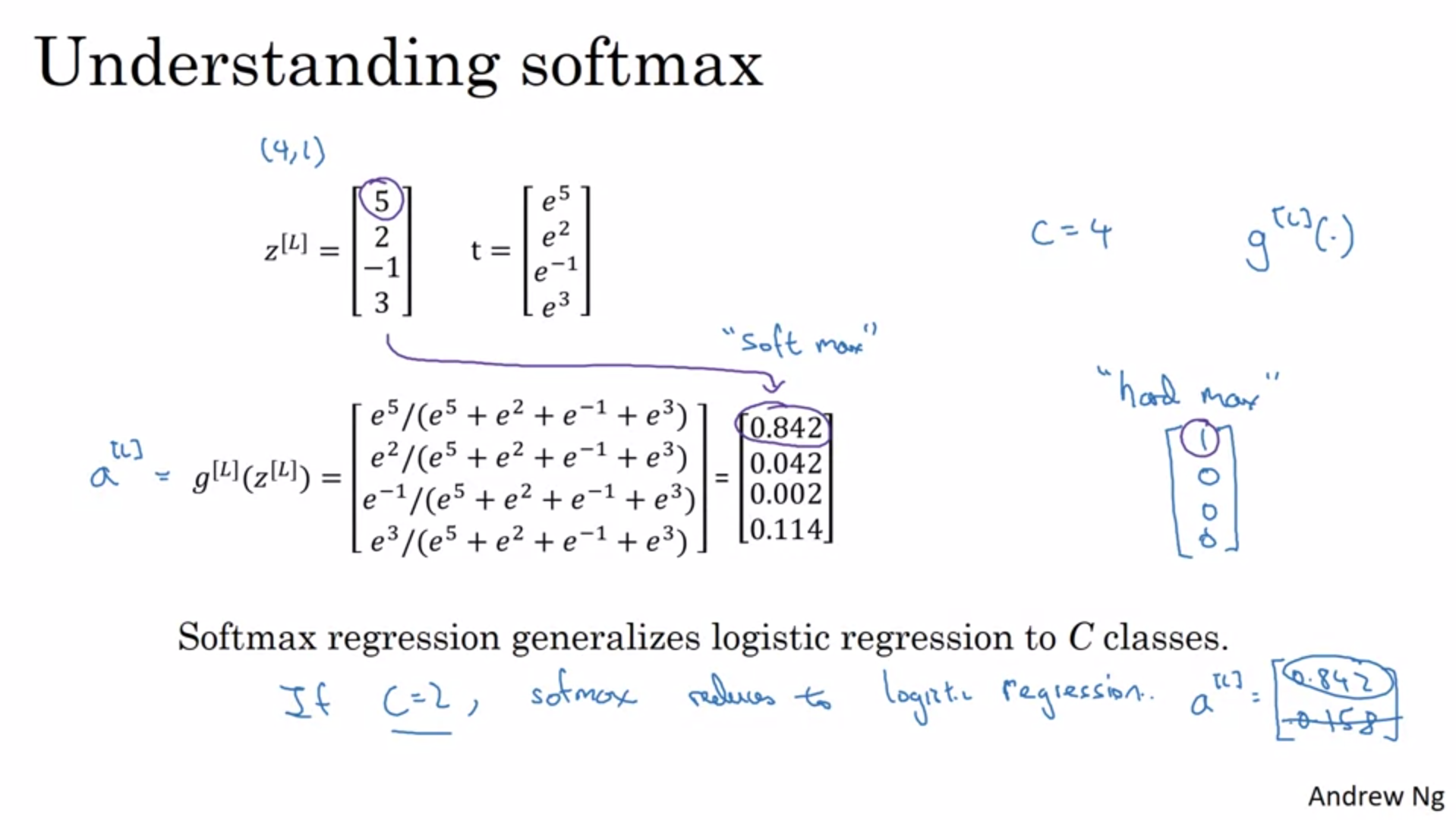
Our previous activation functions had 1 input and 1 output. This particular activation function takes as input a 4x1 vector and outputs a 4x1 vector.



The decision boundary between any 2 classes will be linear.

Here there is no hidden layer. With a deeper network we could learn more complex non-linear decision boundaries.

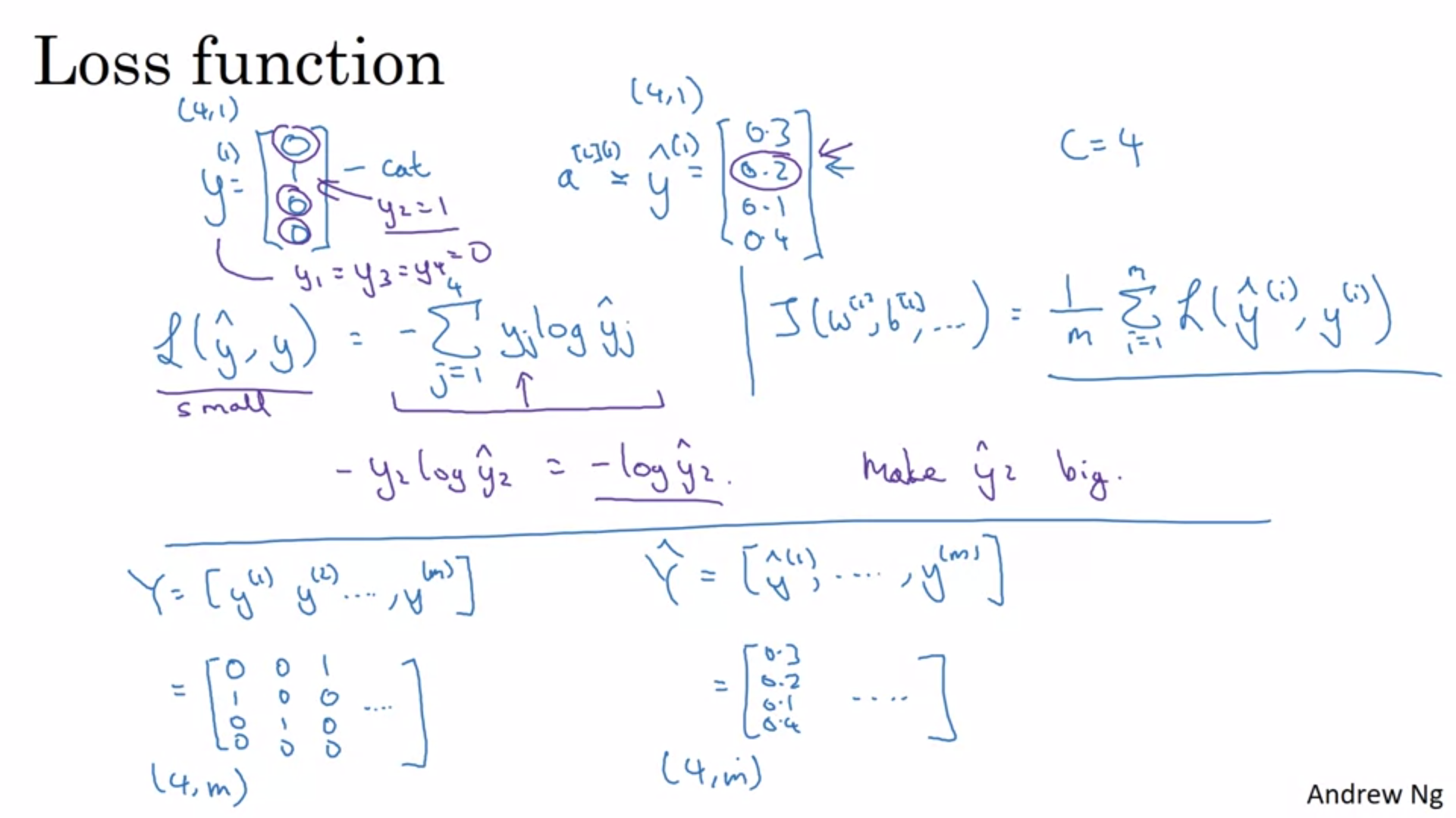




The name soft max comes in contrast to hard max.

The softmax activation generalizes the logistic activation function to multiple classes.

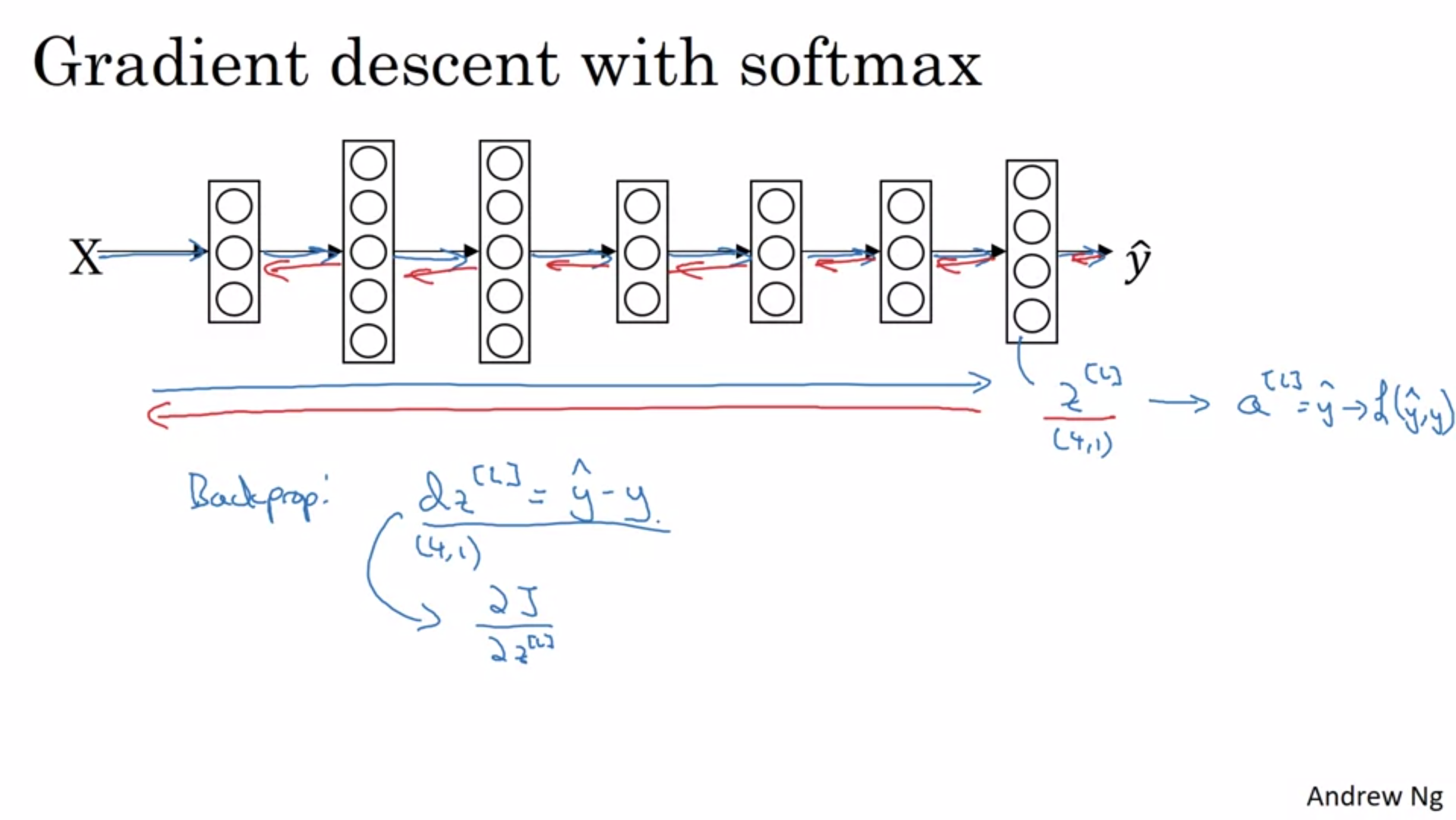
With C=2 softmax reduces to logistic regression. This can be proved.



The output label is now a vector too.

With this loss function all terms except for 1 will be 0.

This is a form of maximum likelihood estimation.



For backprop the start is y\_hat – y.

With deep learning frameworks we only need to figure out the forward pass and the framework will do the backward pass for us.