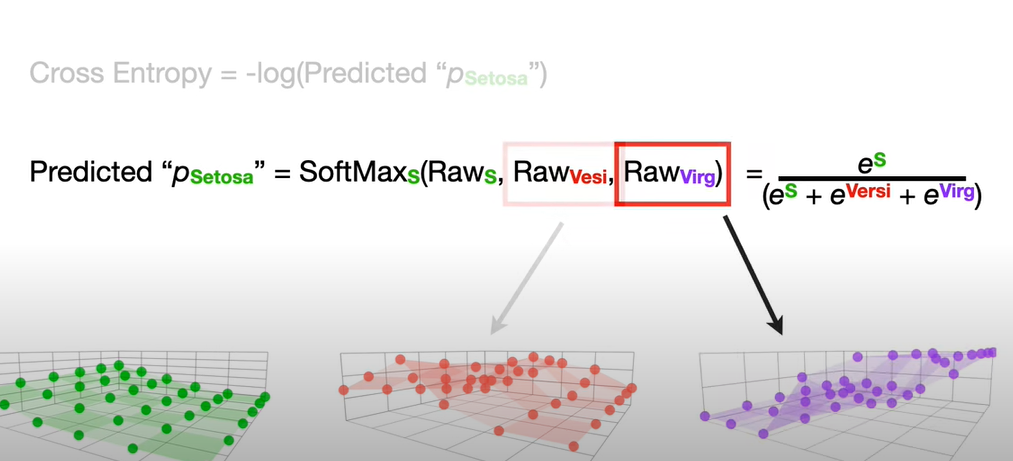
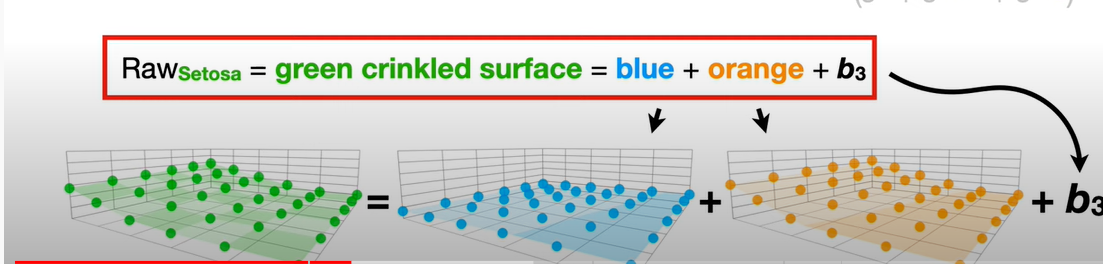
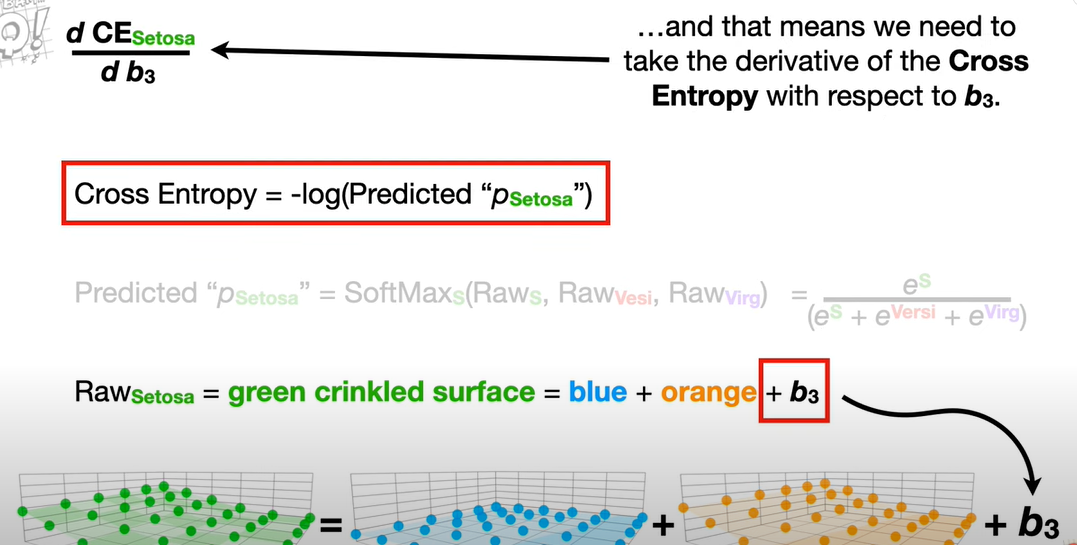
Neural Network



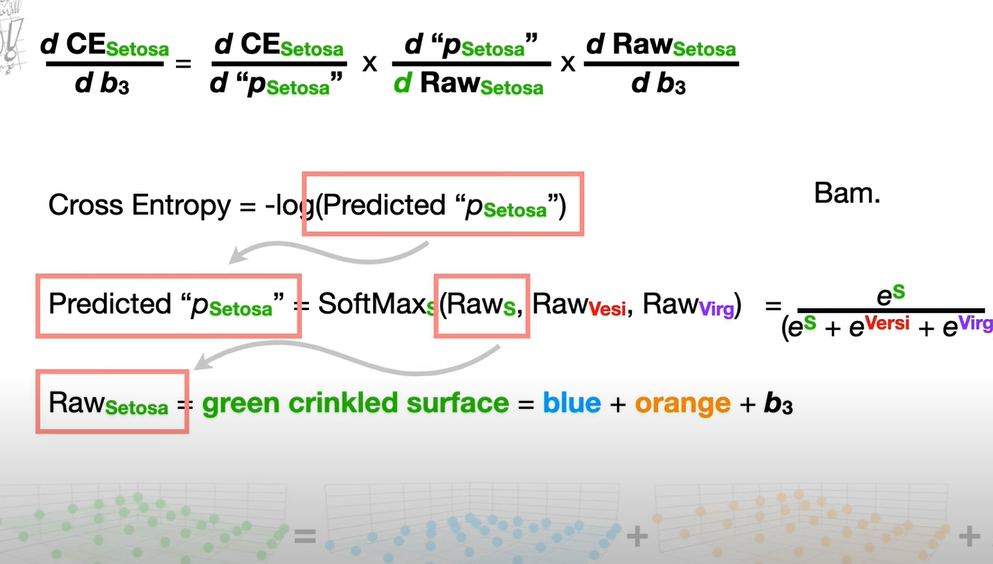
B3 is the BIAS

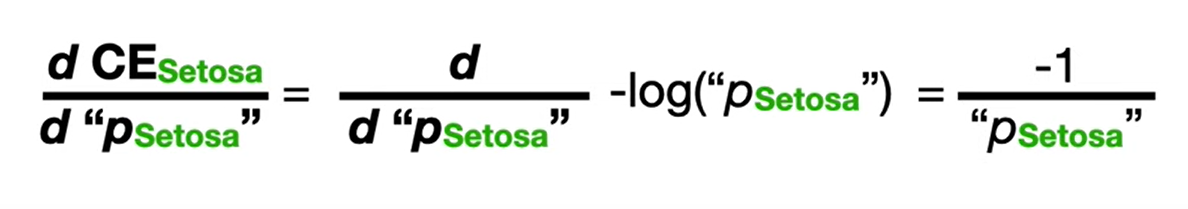


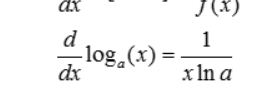
We want to use Gradient Descent to optimize b3. So we need to find the derivative (change in Cross Entropy when b3 changes)

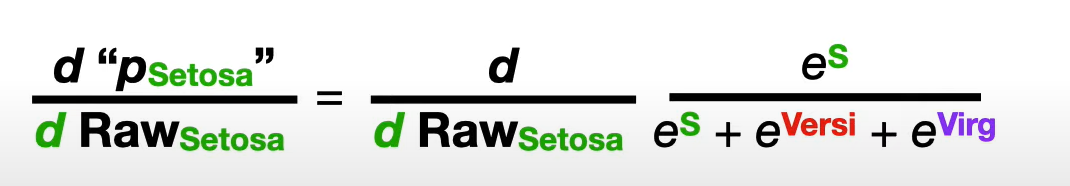


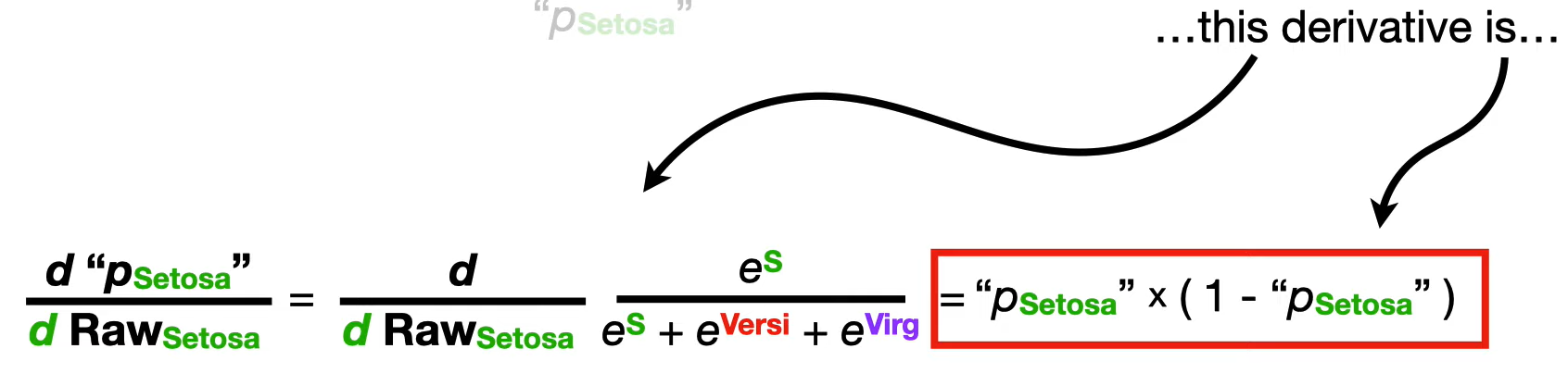
The above derivative follows a chain rule and is:

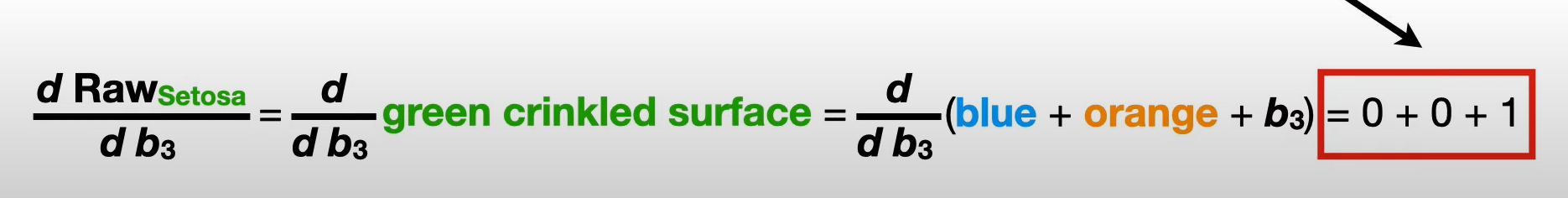


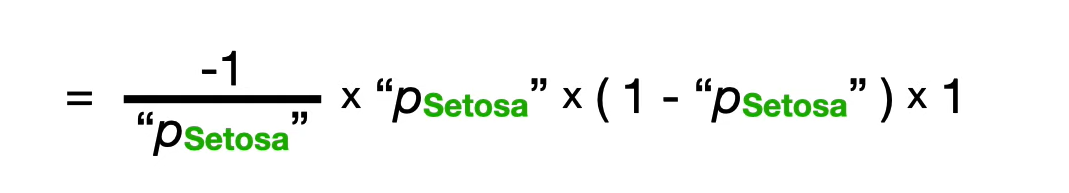


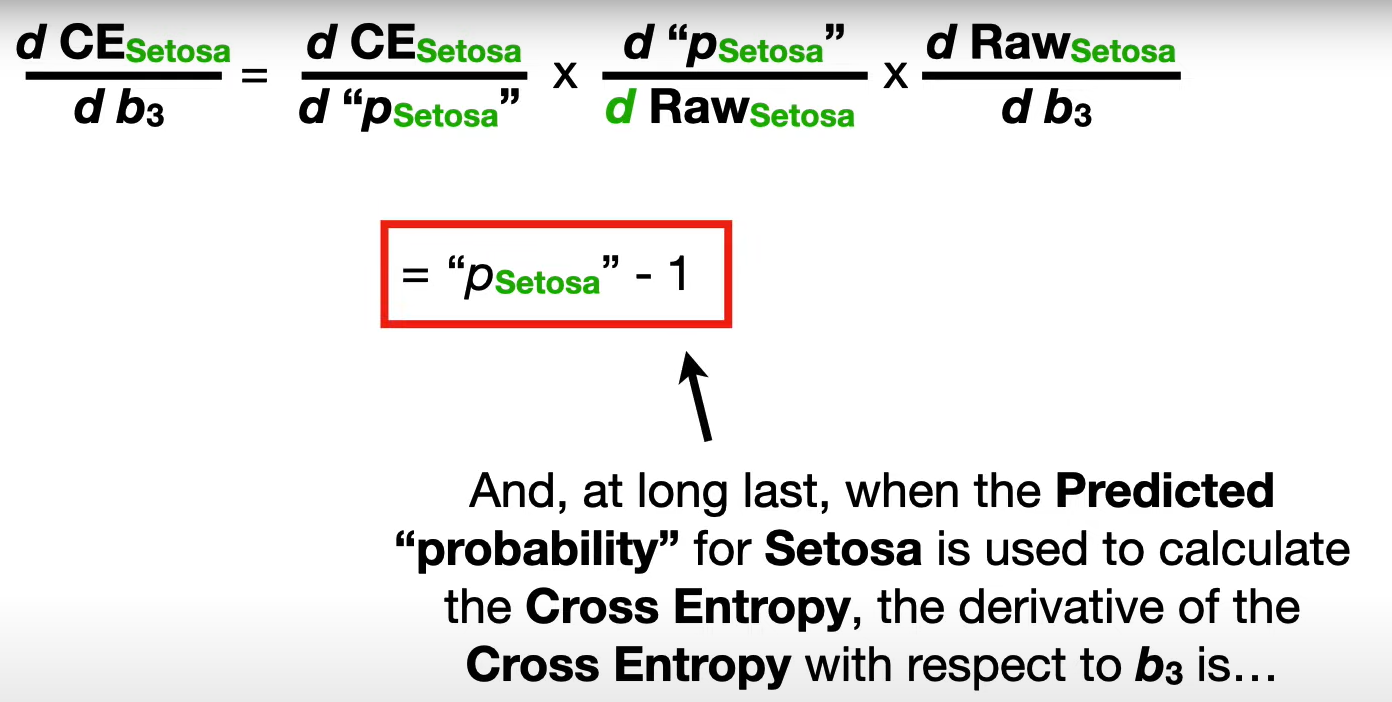




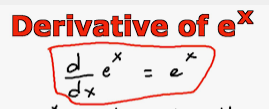






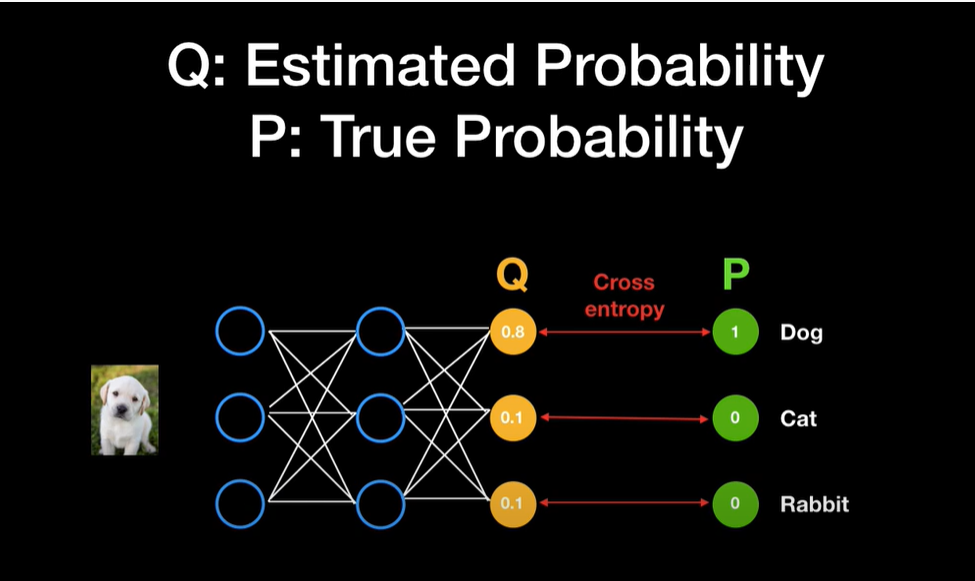


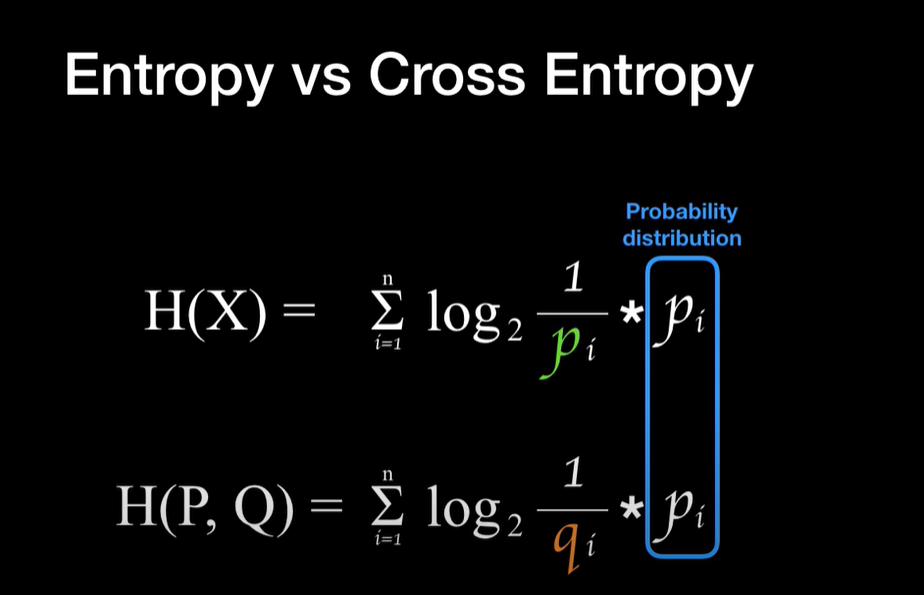
The quotient rule: low-D-high minus high-D-low square the bottom and away we go!

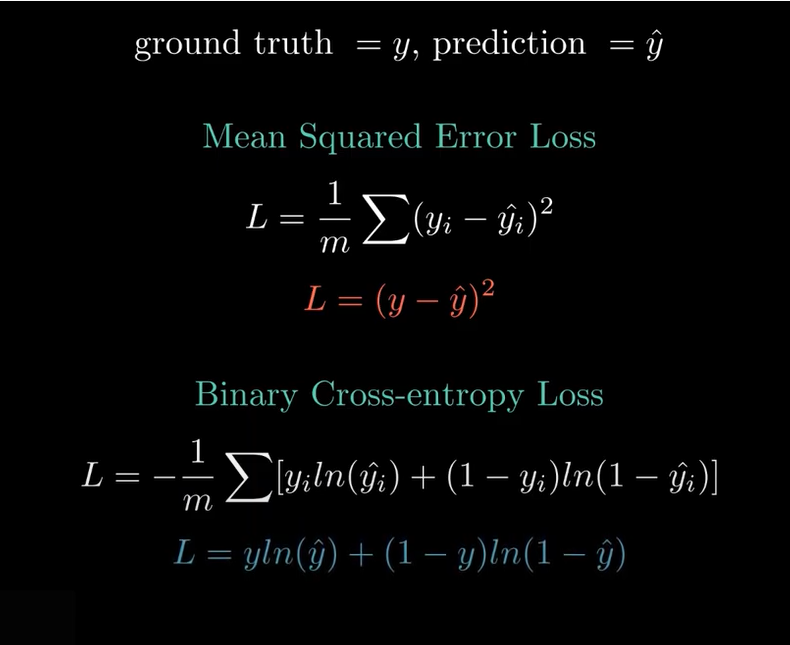


Softmax and Cross Entropy

<https://www.youtube.com/watch?v=ArxdklCAKs0>







**https://www.youtube.com/watch?v=gIx974WtVb4**