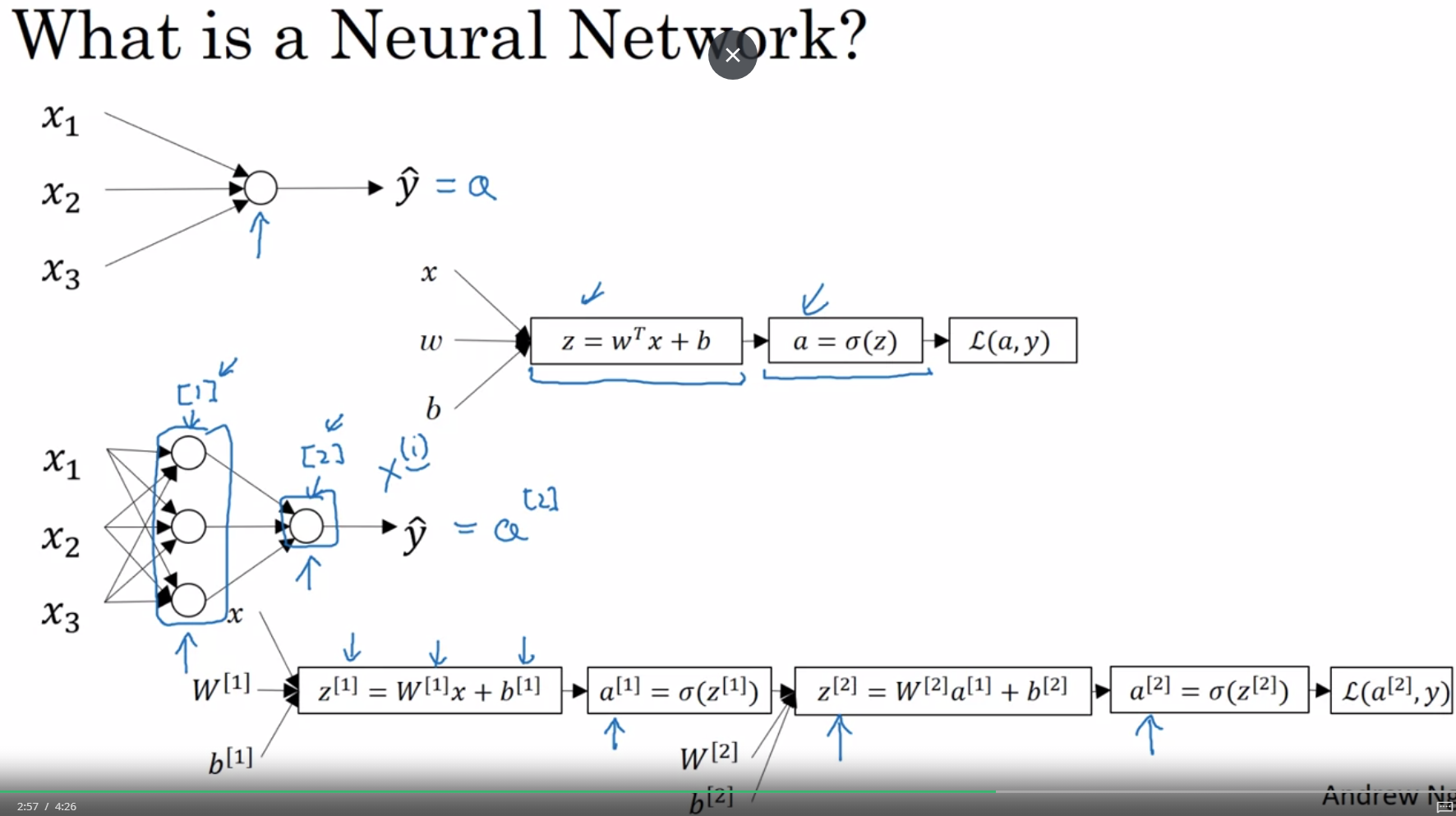
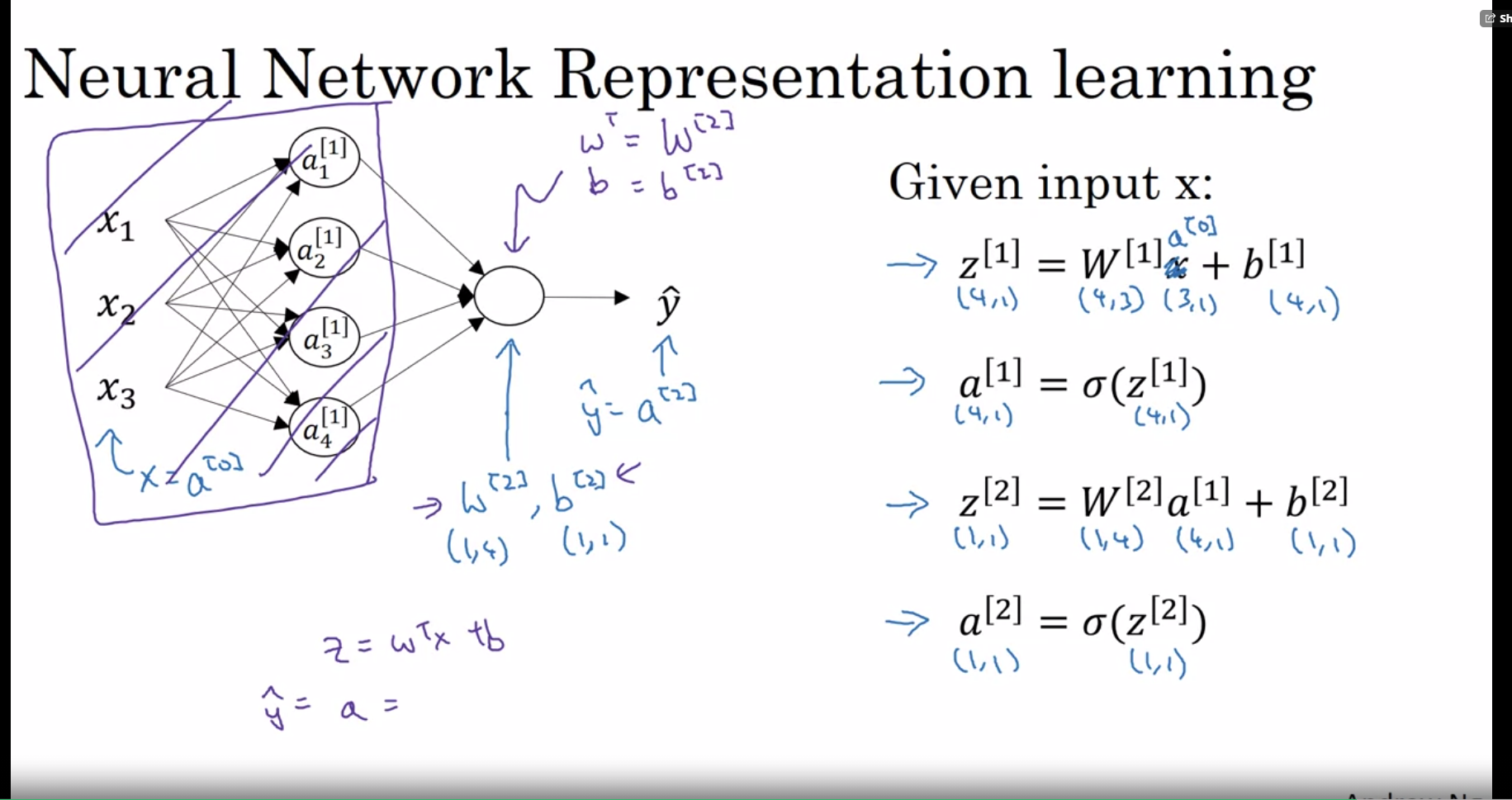
5/20/2021

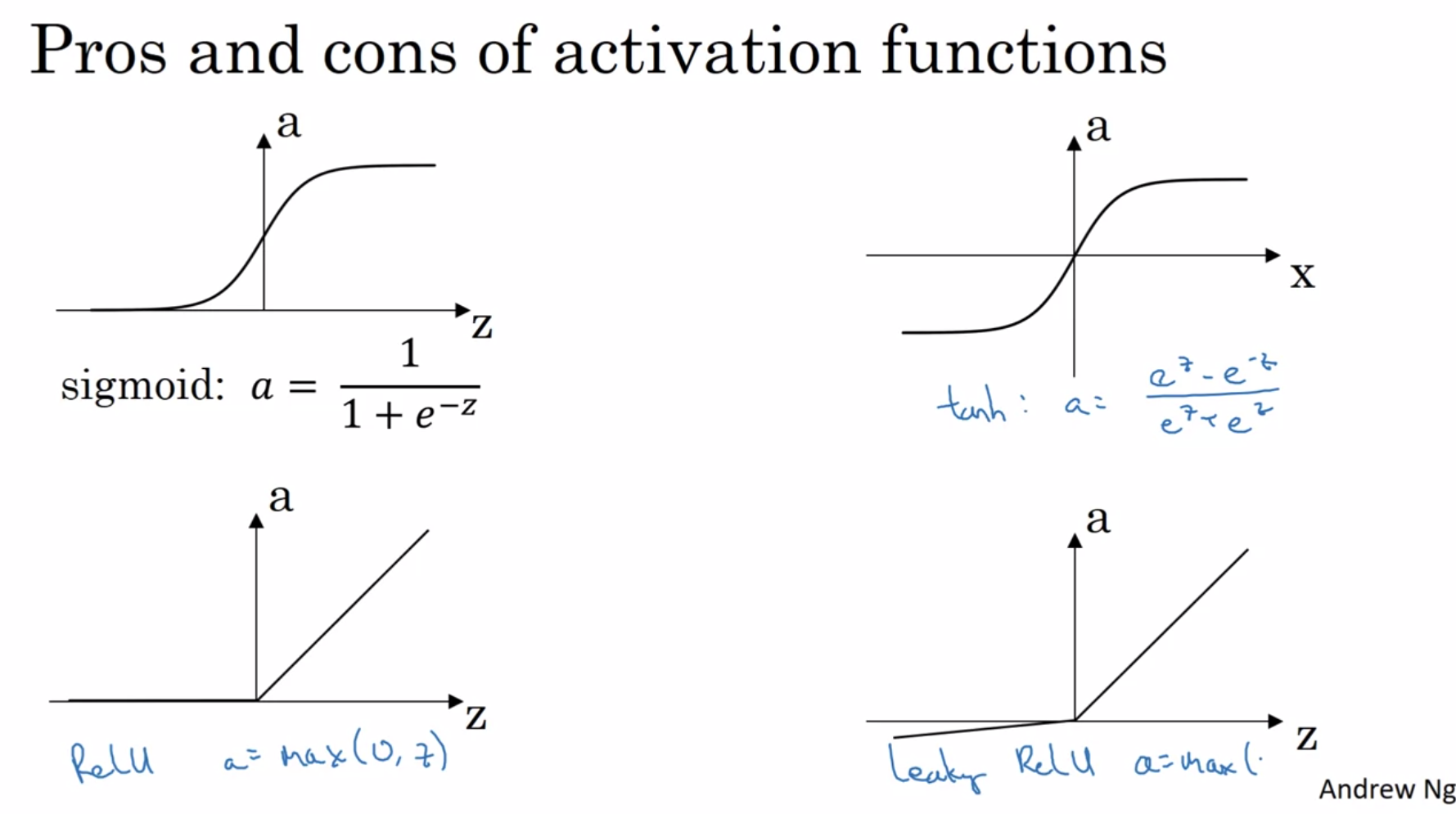
**1/ Introduction to Neural Networks – 1 Hidden Layer:**



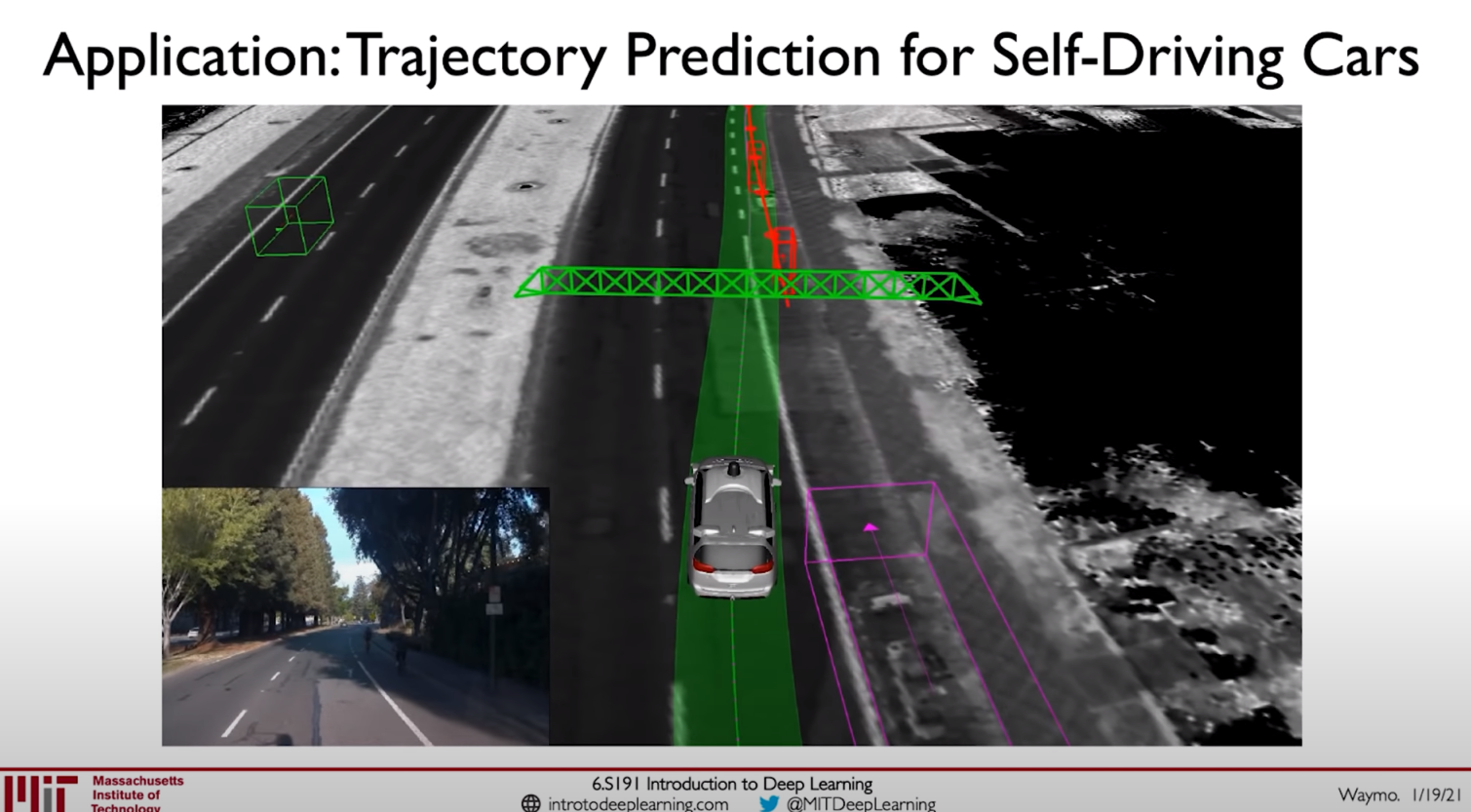


**2/ Activation Functions:**

* Sigmoid or tanh function when z is very large or small, the slope is near to 0 thus decrease the gradient decent process.
* ReLU, the z most of the time > 0 thus we got slope never be 0

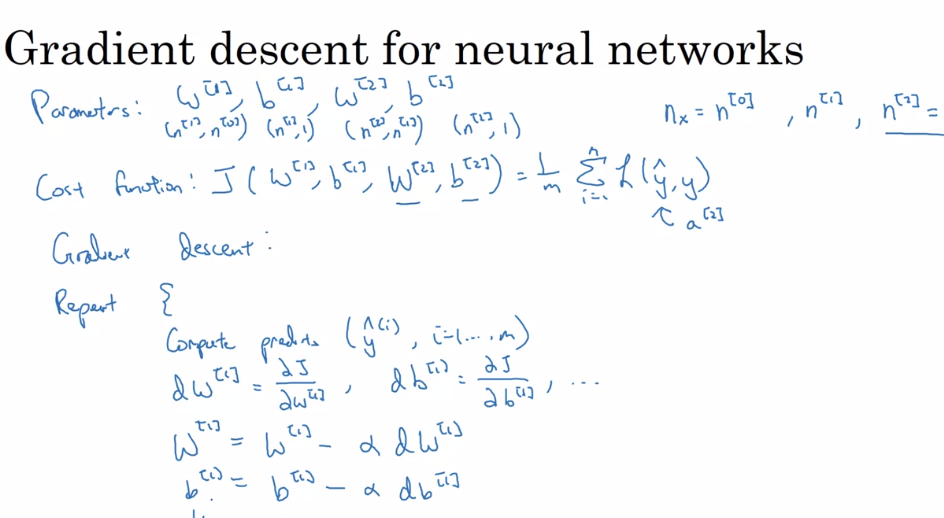


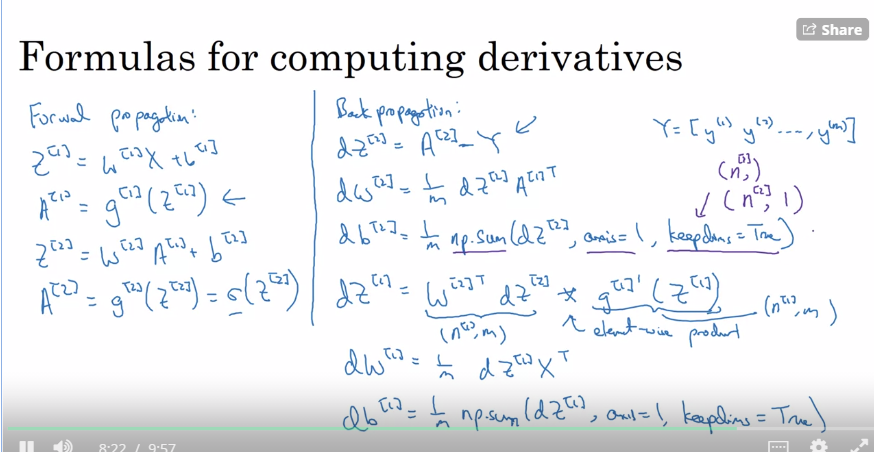
**3/ Why do we need non-linear function?**



* Application for Transformers

**4/ Gradient Descent for NN:**

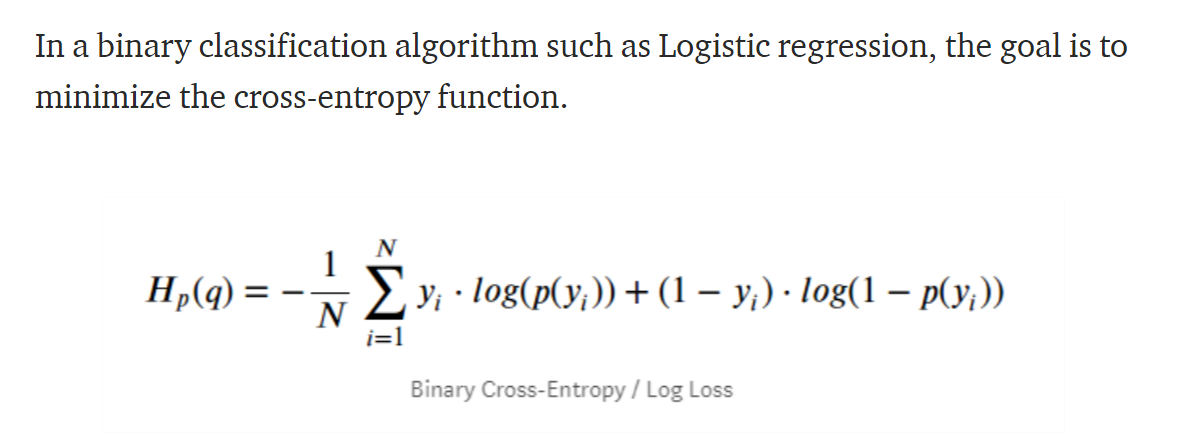




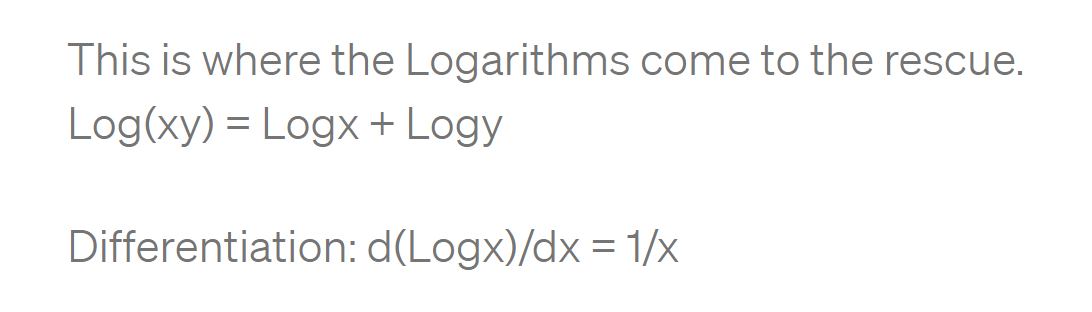
**5/ Why the loss function is a cross-entropy or log loss function in Logistic Regression?**

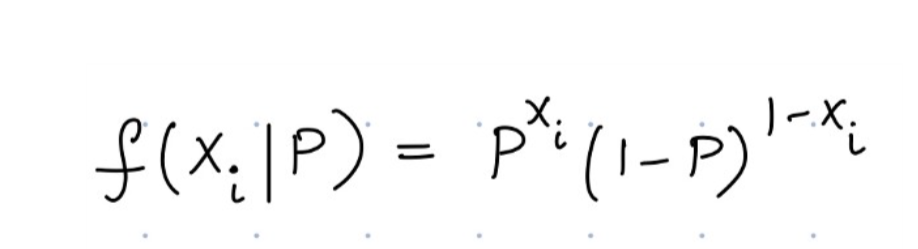
Resource: <https://towardsdatascience.com/log-loss-function-math-explained-5b83cd8d9c83>

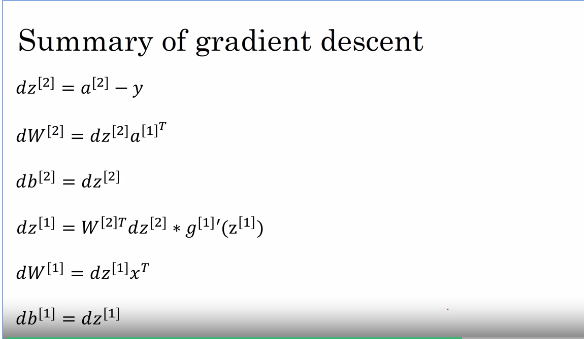
* A good binary classification algo should produce a high value of probability of predicting a class for an input sample. Which is the closest estimate to probability p of predicting the malignate class of total population



* Cross-entroy = measure of the difference between 2 probability distributions for a given random variable or set of events
* The idea is to find the max of likelihood function for a particular value of input
* To find the max of function means to differnentiate the function Dl/Dx = 0



* Why apply log to the likelihood function? It simplifies the expression into the sm of the log of probabilities.
* Bernoulli Distribution: 



**QUIZ:**

