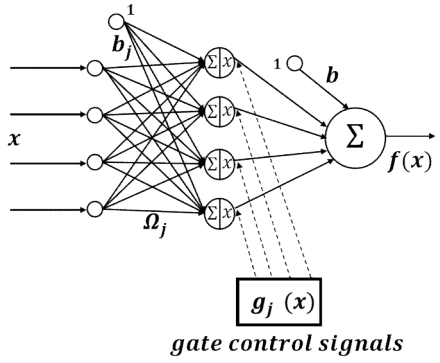
More about NN Structure

Gated Linear Network

Nonlinearities in a network are key important.



Case 1: $g_j(x) \equiv [1\ 1\ 1\ 1]$

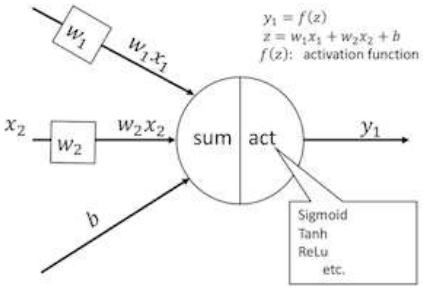
Linear mapping

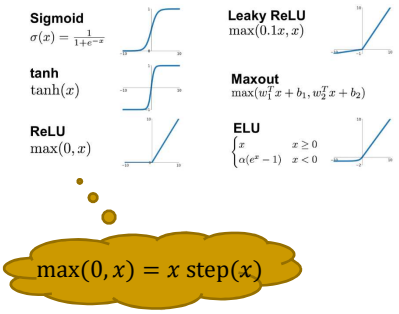
$f(x) = Wx + b$

Case 2: $g_j(x) \equiv [1\ 0\ 1\ 1], [1\ 1\ 0\ 1], [0\ 0\ 0\ 1], [1\ 1\ 1\ 0]$ binary sequences

Piecewise Linear mapping

Sigmoid and ReLU Activation Function



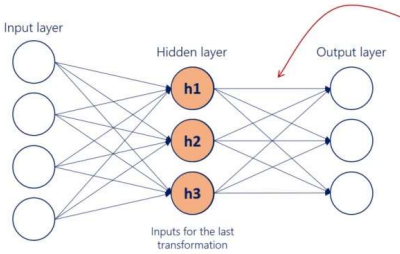


Softmax Activation Function

Softmax

a = xw+b
Linear combination

y = softmax (**a**)
Activation



$\mathbf{a}_h$ = hw+b

a = [-0.21,0.47,1.72]

softmax (**a**) = $\frac{e^{a_i}}{\sum_j e^{a_j}}$

$\sum_j e^{a_j} = e^{-0.21} + e^{0.47} + e^{1.72} = 8$

Applying Neural Networks

Example Problem

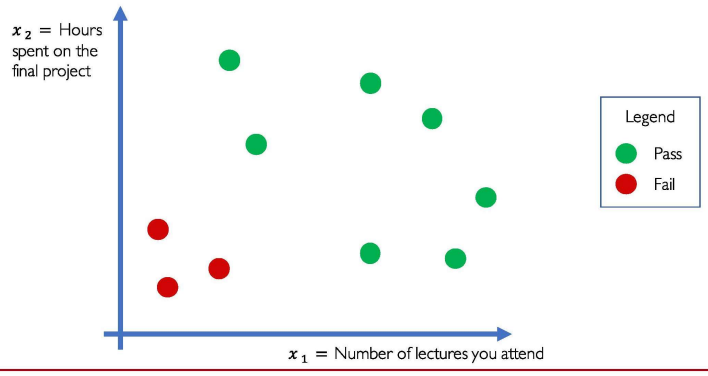
Will I pass this class?

Let's start with a simple two feature model

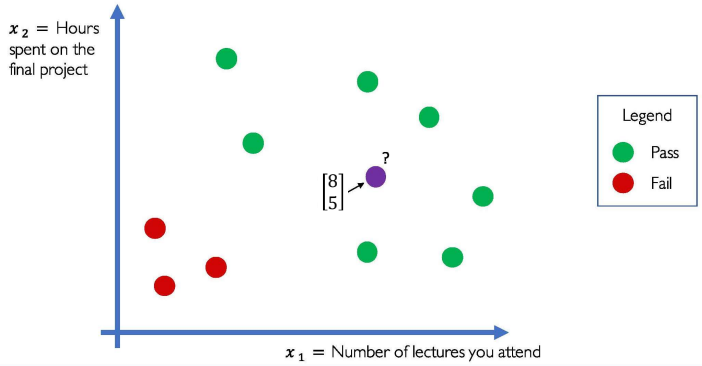
x_1 = Number of lectures you attend

x_2 = Hours spent on the final project

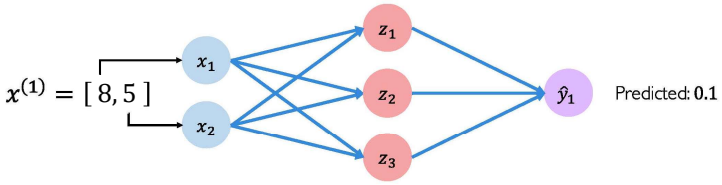
Example Problem: Will I pass this class?



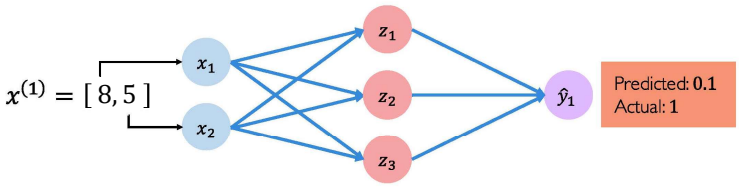
Example Problem: Will I pass this class?



Example Problem: Will I pass this class?

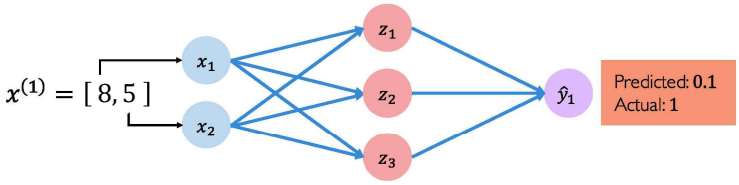


Example Problem: Will I pass this class?



Quantifying Loss

*The **loss** of our network measures the cost incurred from incorrect predictions*

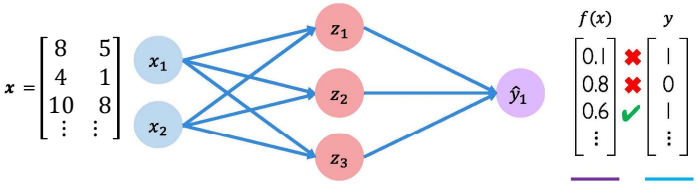


$$\mathcal{L}\left(f\left(x^{(i)}; \boldsymbol{\theta}\right), y^{(i)}\right)$$

Predicted Actual

Empirical Loss

*The **empirical loss** measures the total loss over our entire dataset*



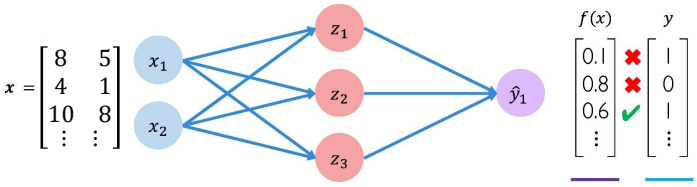
Also known as:
- Objective function
- Cost function
- Empirical Risk

$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}\left(f\left(x^{(i)};\theta\right), y^{(i)}\right)$$

Predicted Actual

Binary Cross Entropy Loss

***Cross entropy loss** can be used with models that output a probability between 0 and 1*

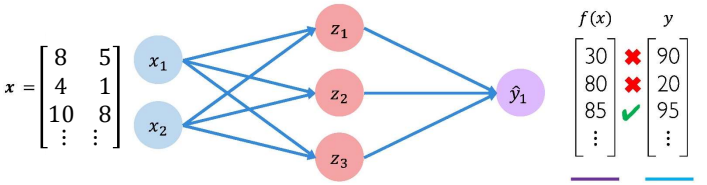


$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)} \log\left(f\left(x^{(i)};\theta\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - f\left(x^{(i)};\theta\right)\right)$$

Actual Predicted Actual Predicted

Mean Squared Error Loss

Mean squared error loss *can be used with regression models that output continuous real numbers*



$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - f(x^{(i)};\theta)\right)^2$$

Actual Predicted

Final Grades (percentage)