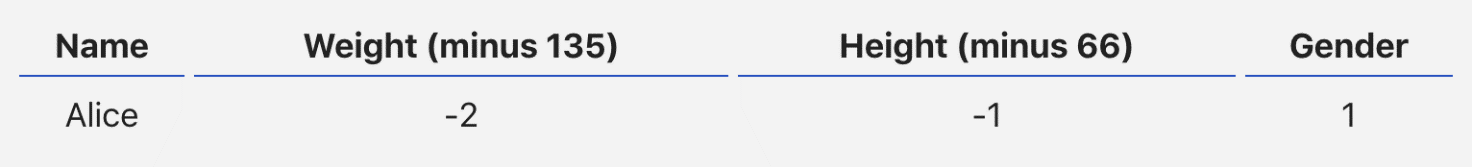
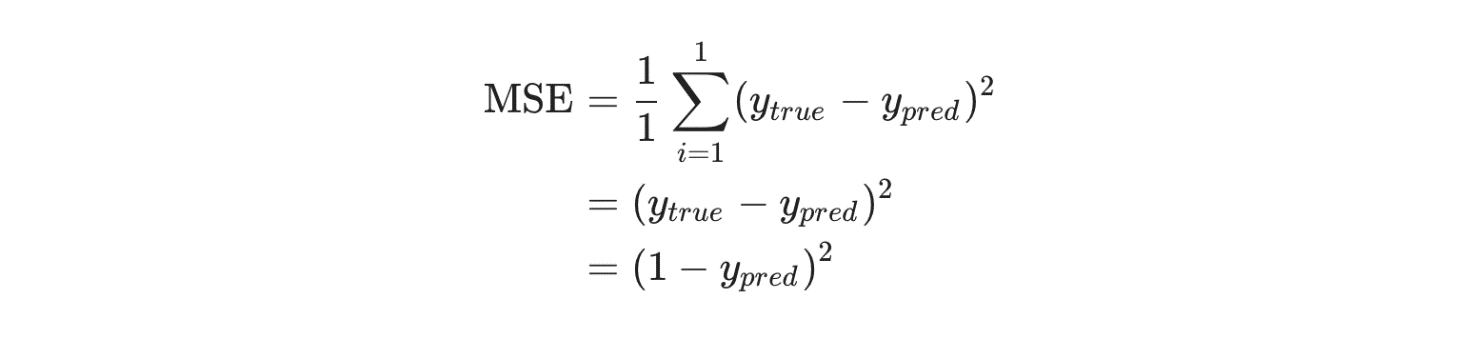
# ANNEXOS

## Exemple derivada parcial Backpropagation

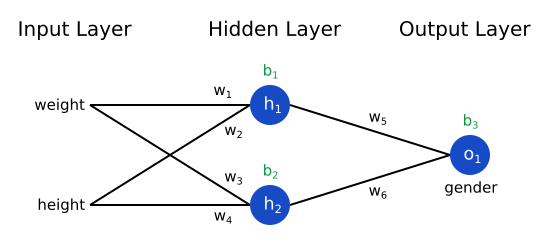
For simplicity, let’s pretend we only have Alice in our dataset:



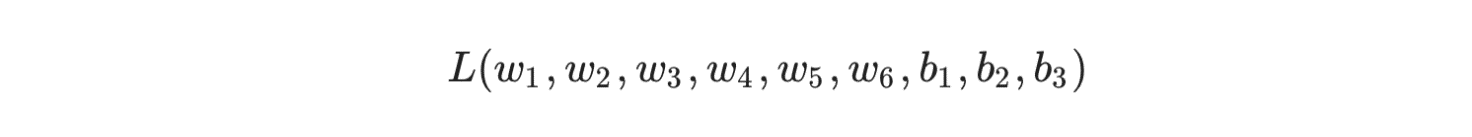
Then the mean squared error loss is just Alice’s squared error:



Another way to think about loss is as a function of weights and biases. Let’s label each weight and bias in our network:

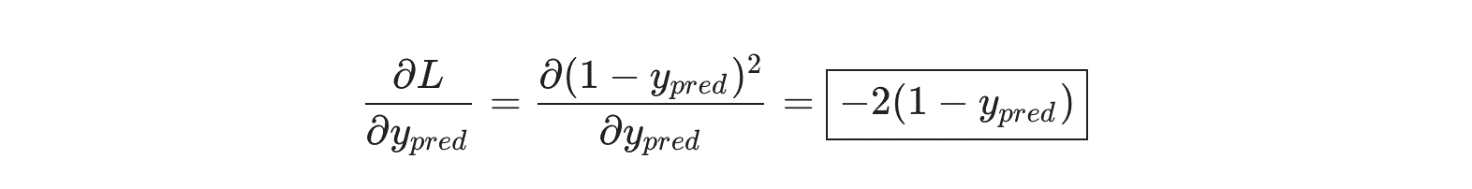


Then, we can write loss as a multivariable function:



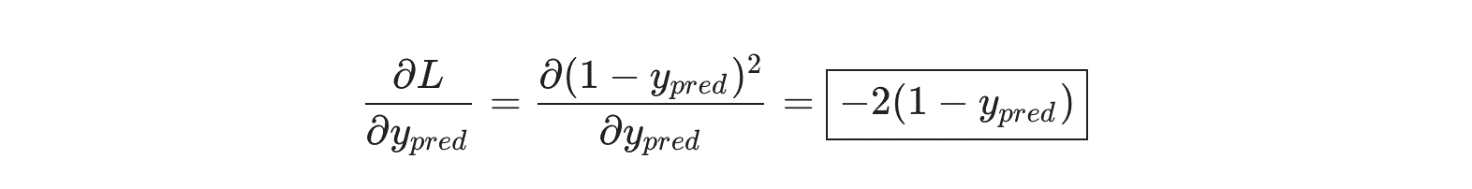
Imagine we wanted to tweak w1​. How would loss L change if we changed w1​? That’s a question the partial derivative can answer. How do we calculate it?

To start, let’s rewrite the partial derivative in terms of ∂y\_pred/∂w1​​ instead:

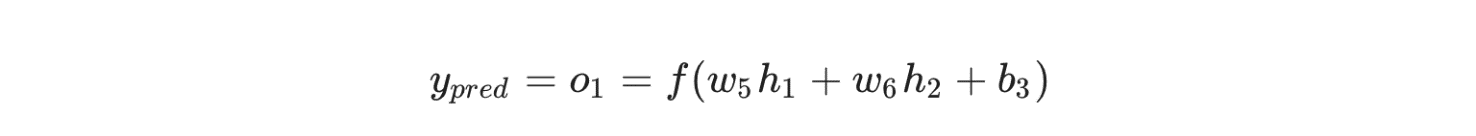


This works because of the Chain Rule.

We can calculate ∂L/∂y\_pred​ because we computed L= (1−y\_pred​)² above:

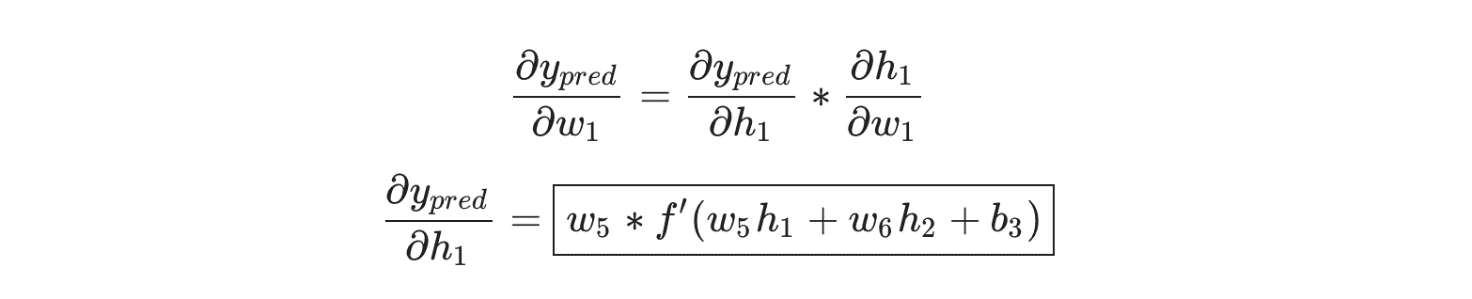


Now, let’s figure out what to do with ∂y\_pred/∂w1. Just like before, let h1​, h2​, o1​ be the outputs of the neurons they represent. Then

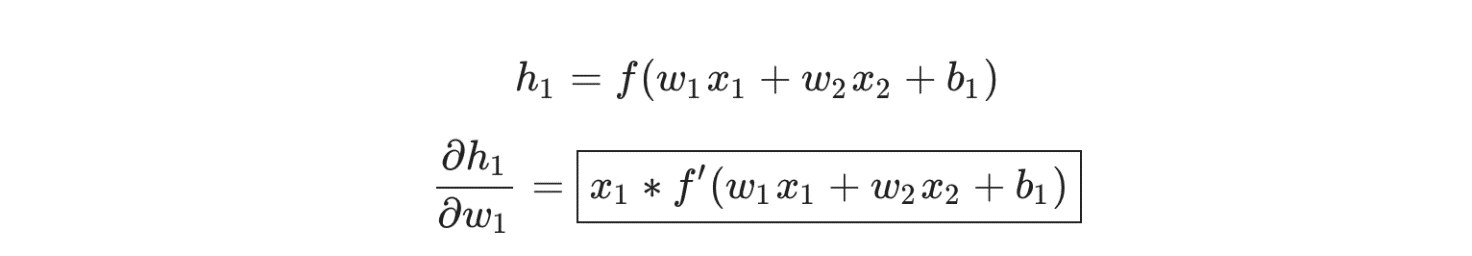


f is the sigmoid activation function, remember?

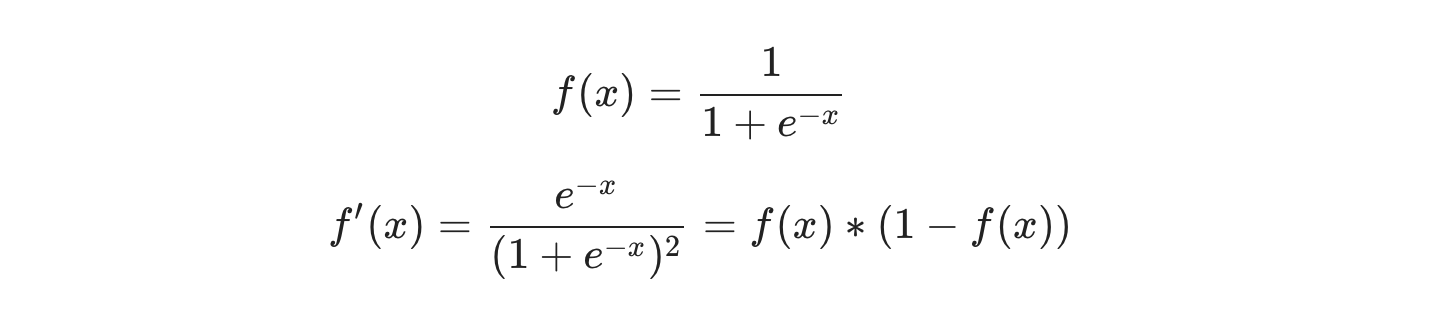
Since w1​ only affects h1​ (not h2​), we can write



We do the same thing for ∂h1​​/∂w1:

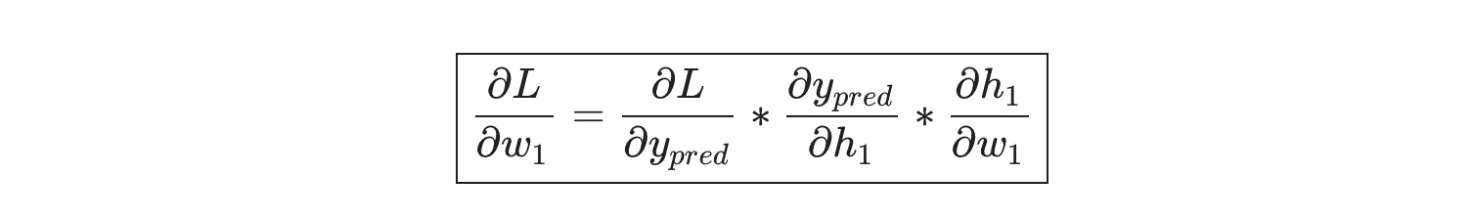


x1​ here is weight, and x2​ is height. This is the second time we’ve seen f′(x) (the derivate of the sigmoid function) now! Let’s derive it:



We’ll use this nice form for f′(x) later.

We’re done! We’ve managed to break down ∂L/∂w1​ into several parts we can calculate:



This system of calculating partial derivatives by working backwards is known as backpropagation, or “backprop”.

## ROC i AUC

An **ROC curve** (**receiver operating characteristic curve**) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

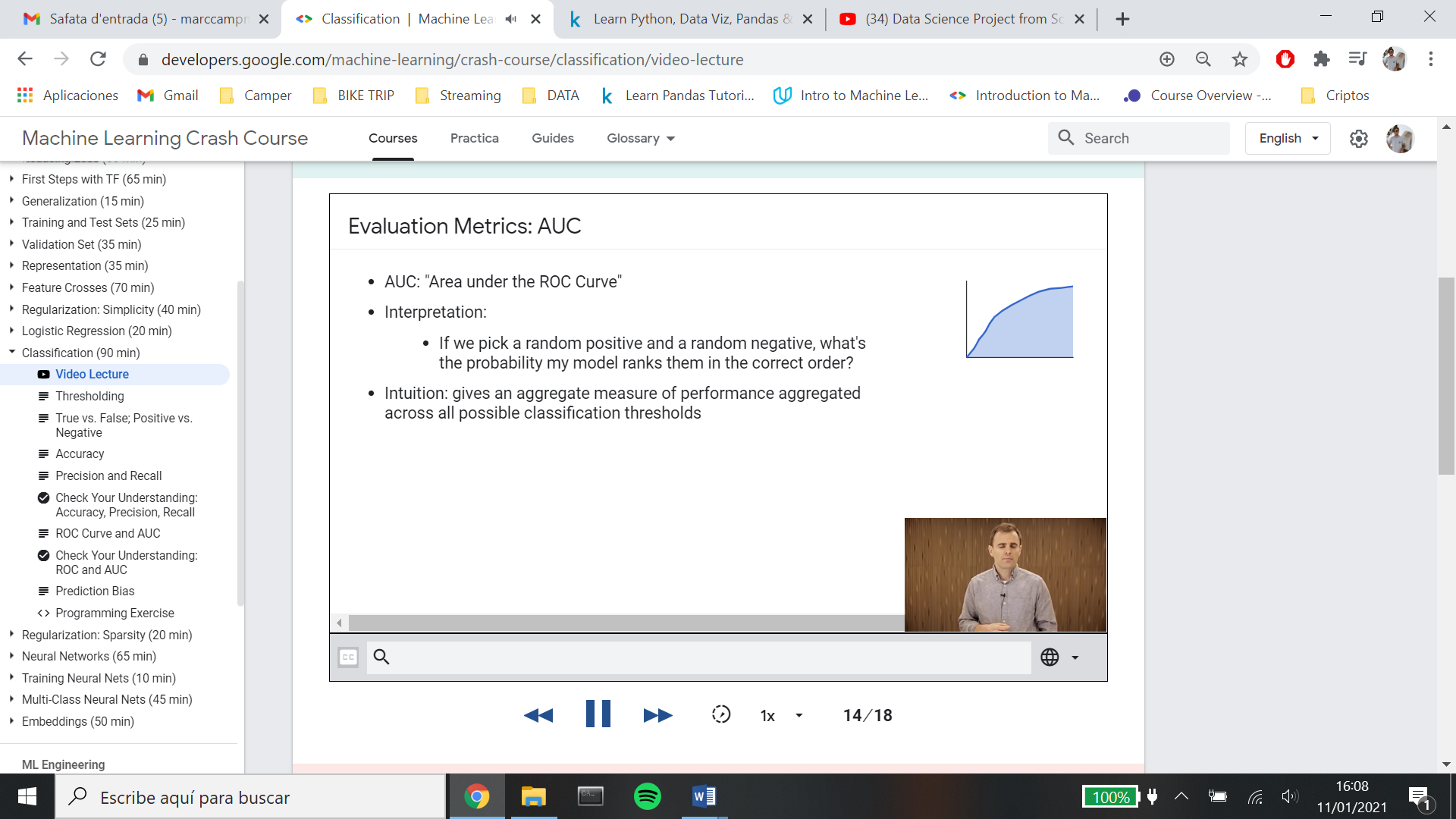
**True Positive Rate** (**TPR**) is a synonym for recall and is therefore defined as follows:

TPR=TP/ (TP+FN)

**False Positive Rate** (**FPR**) is defined as follows:

FPR=FP/ (FP+TN)

An ROC curve plots TPR vs. FPR at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. The following figure shows a typical ROC curve.



**AUC** stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve (think integral calculus) from (0,0) to (1,1).

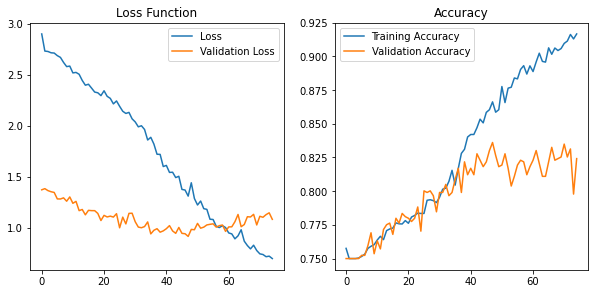
## Data augmentation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Loss | Accuracy | Precision | Recall | AUC |
| Train | 0.8381 | 0.8489 | 0.7291 | 0.6295 | 0.8780 |
| Val. | 0.8869 | 0.8445 | 0.7005 | 0.6603 | 0.8660 |
| Test | 0.8266 | 0.8486 | 0.7125 | 0.6610 | 0.8825 |



## Equilibri de pesos

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Loss | Accuracy | Precision | Recall | AUC |
| Train | 0.7328 | 0.9147 | 0.8712 | 0.7729 | 0.9654 |
| Val. | 1.0824 | 0.8242 | 0.6632 | 0.6029 | 0.8660 |
| Test | 1.0310 | 0.8476 | 0.7167 | 0.6457 | 0.8752 |



## Transfer Learning

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Loss | Accuracy | Precision | Recall | AUC |
| Train | 0.1543 | 0.9795 | 0.9695 | 0.9479 | 0.9974 |
| Val. | 0.8685 | 0.8529 | 0.7150 | 0.6842 | 0.9039 |
| Test | 0.7546 | 0.8650 | 0.7413 | 0.7065 | 0.9111 |

