

# Logistic Regression: Testing

To test your model, you would run a subset of your data, known as the validation set, on your model to get predictions. The predictions are the outputs of the sigmoid function. If the output is  $\geq 0.5$ , you would assign it to a positive class. Otherwise, you would assign it to a negative class.

$$\begin{aligned}
 & \bullet X_{val} \ Y_{val} \ \theta \\
 & \quad h(X_{val}, \theta) \\
 & \text{pred} = h(X_{val}, \theta) \geq 0.5 \quad \begin{bmatrix} 0.3 \\ 0.8 \\ 0.5 \\ \vdots \\ h_m \end{bmatrix} \geq 0.5 = \begin{bmatrix} 0.3 \geq 0.5 \\ 0.8 \geq 0.5 \\ 0.5 > 0.5 \\ \vdots \\ \text{pred}_m \geq 0.5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ \text{pred}_m \end{bmatrix}
 \end{aligned}$$

In the video, I briefly mentioned  $X$  validation. In reality, given your  $X$  data you would usually split it into three components.  $X_{train}$ ,  $X_{val}$ ,  $X_{test}$ . The distribution usually varies depending on the size of your data set. However, an 80, 10, 10 split usually works fine.

To compute accuracy, you solve the following equation:

$$\text{Accuracy} \longrightarrow \sum_{i=1}^m \frac{(\text{pred}^{(i)} == y_{val}^{(i)})}{m}$$

In other words, you go over all your training examples,  $m$  of them, and then for every prediction, if it was right you add a one. You then divide by  $m$ .