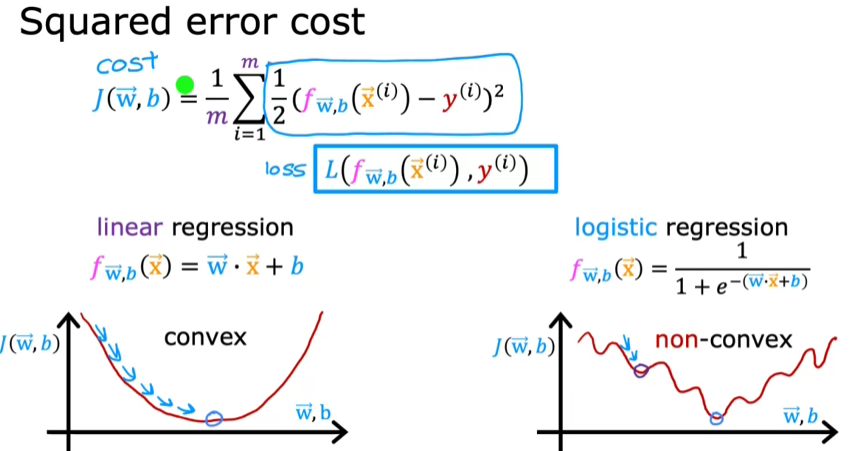
**COST FUNCTION WITH LOGISTIC REGRESSION**

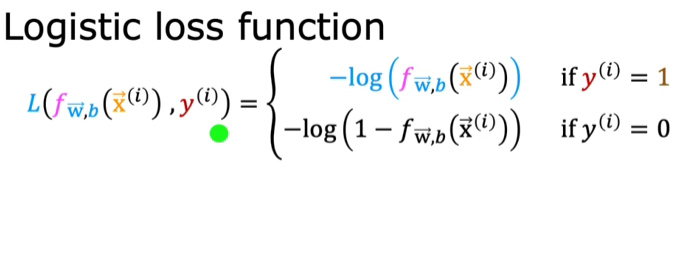
**Understanding Cost Functions**

* **The cost function measures how well a set of parameters fits the training data, guiding the selection of better parameters.**
* **The squared error cost function leads to a non-convex cost function for logistic regression, resulting in multiple local minima that complicate optimization.**

****

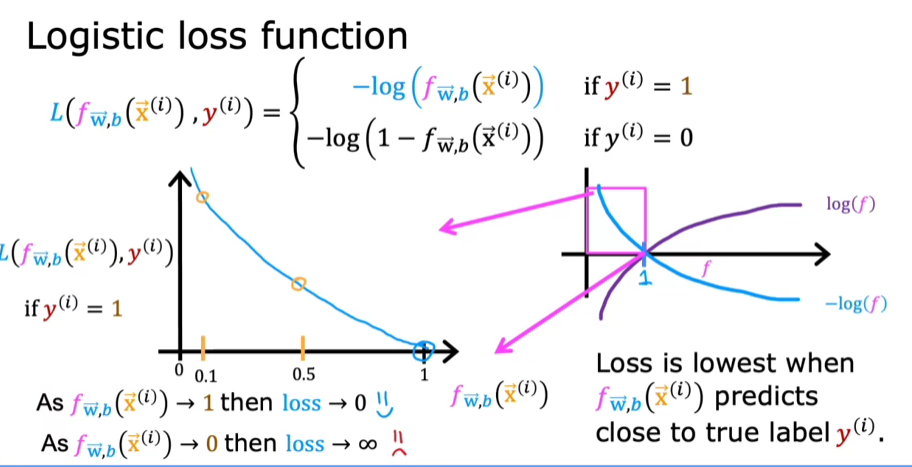
***Introducing the New Loss Function***

* **A new loss function is defined for logistic regression, which is based on the true label and the predicted probability.**
* **The loss function is structured to be convex, allowing gradient descent to reliably converge to the global minimum.**

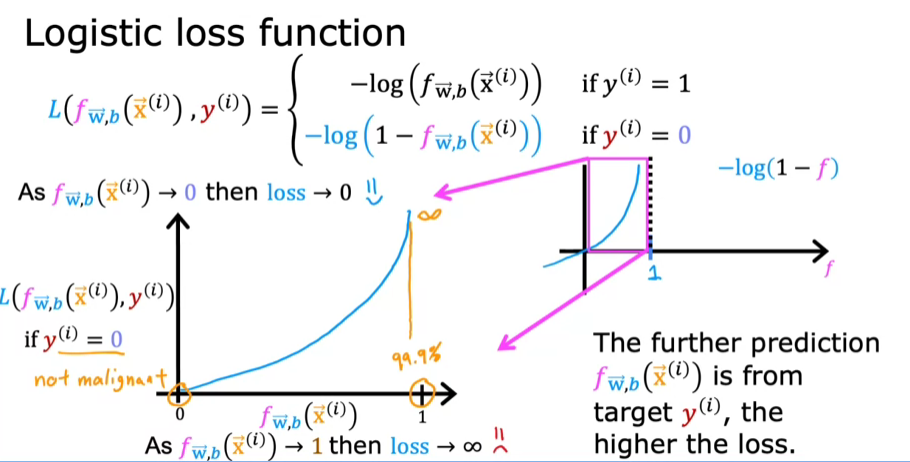
****

**Analyzing Loss for Different Labels**

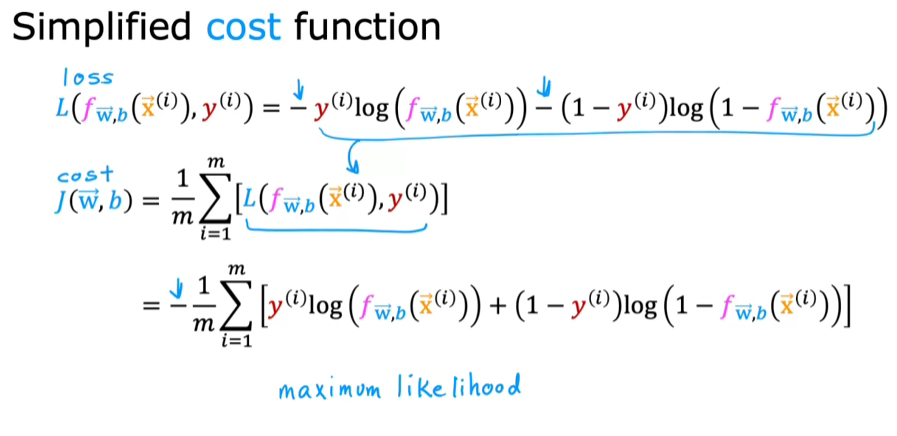
* **When the true label (y = 1), the loss incentivizes accurate predictions close to 1, minimizing loss when predictions align with the true label.**

****

* **Conversely, when (y = 0), the loss increases significantly as predictions deviate from 0, penalizing incorrect high predictions.**

****

**SIMPLIFIED COST FUNCTION FOR LOGISTIC REGRESSION**

****

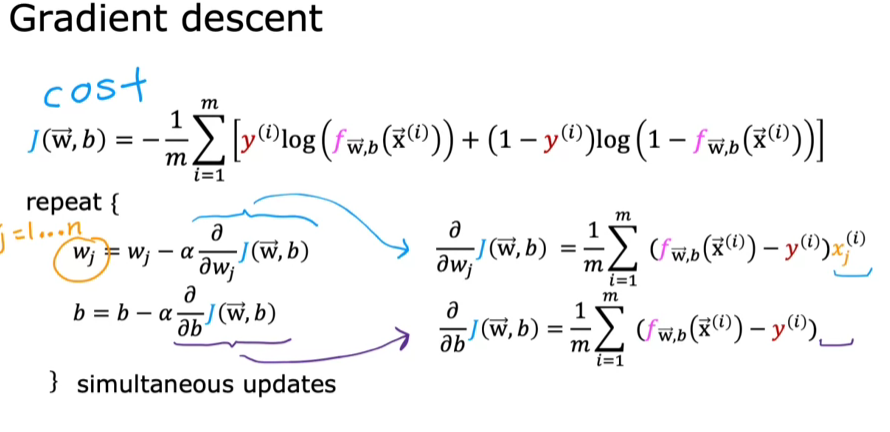
**Statistical Justification**

* **The chosen cost function is based on maximum likelihood estimation, a statistical principle that helps find model parameters efficiently.**
* **This cost function has the advantageous property of being convex, which is beneficial for optimization.**

**GRADIENT DESCENT IMPLEMENTATION**

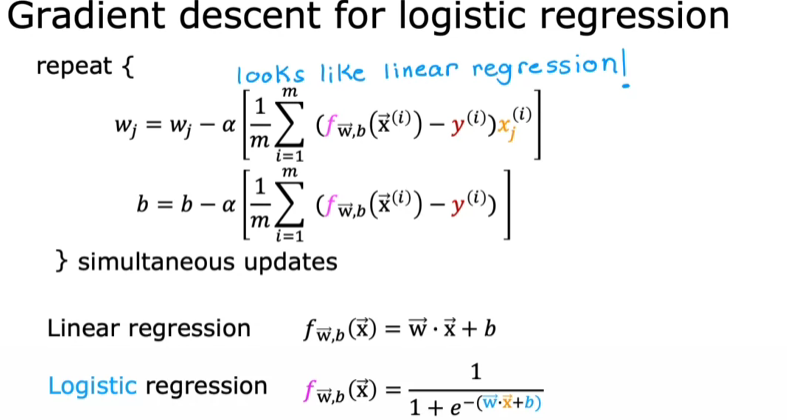
**Gradient descent algorithm**

* **The gradient descent algorithm updates parameters by calculating the derivative of the cost function with respect to (w) and (b).**
* **The updates are performed simultaneously for all parameters to ensure efficient convergence.**

****

**Difference between linear and logistic regression**

* **Although the gradient descent equations for both algorithms appear similar, they differ in the definition of the function (f(x)); logistic regression uses the sigmoid function, while linear regression uses a linear equation.**
* **Feature scaling can be applied to both algorithms to help gradient descent converge faster, enhancing the model's performance.**

****