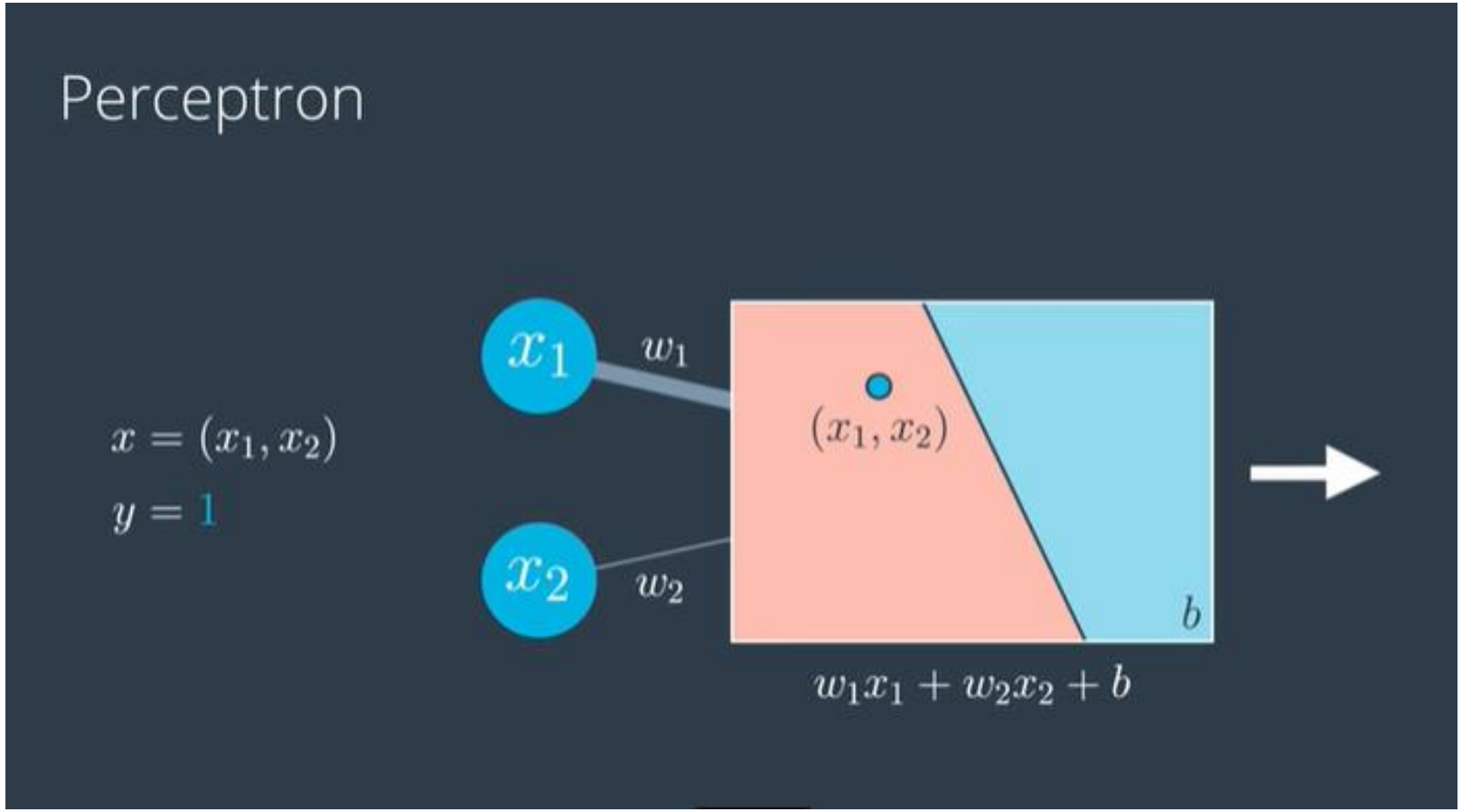
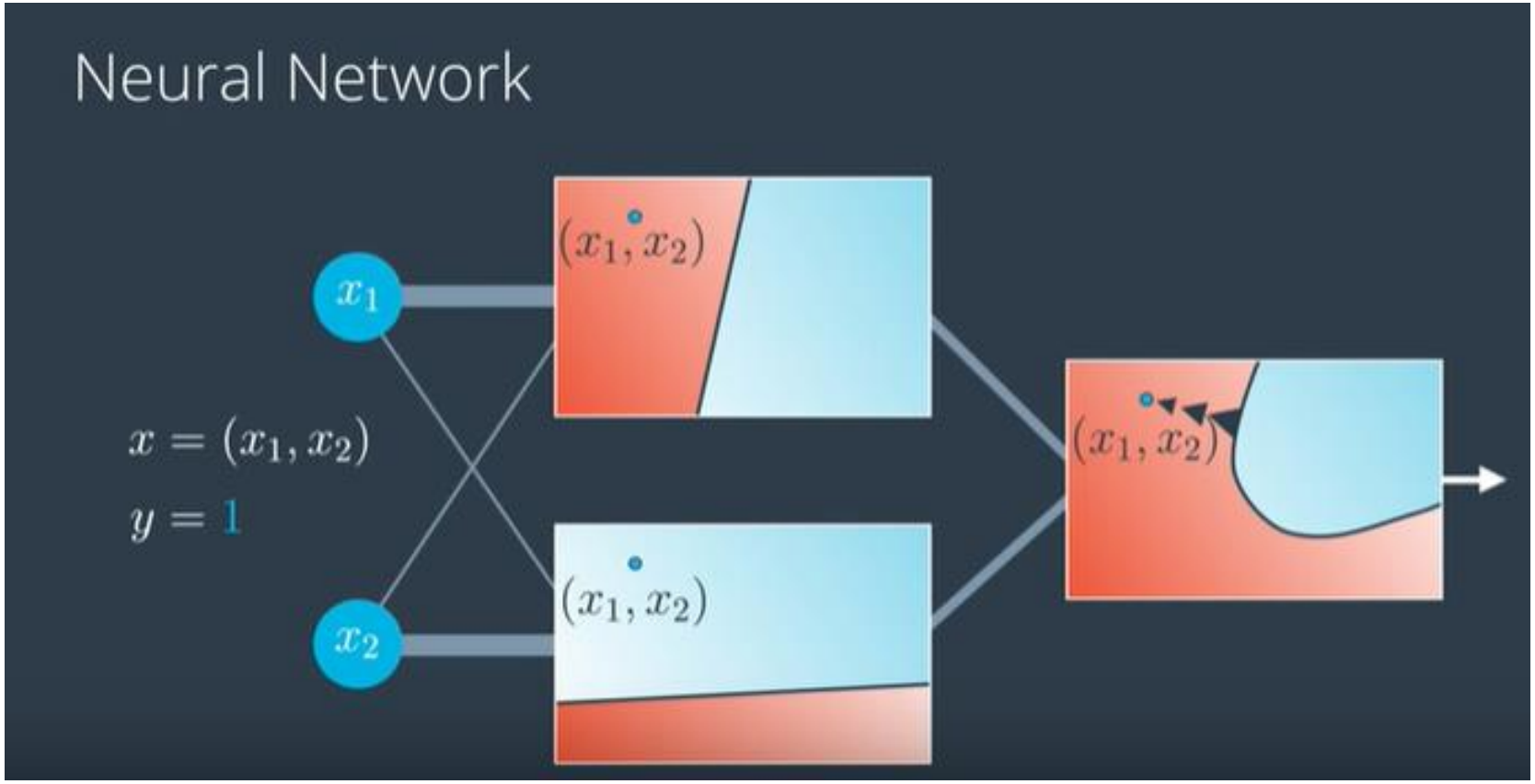
Feedforward

Feedforward is the process neural networks use to turn the input into an output.



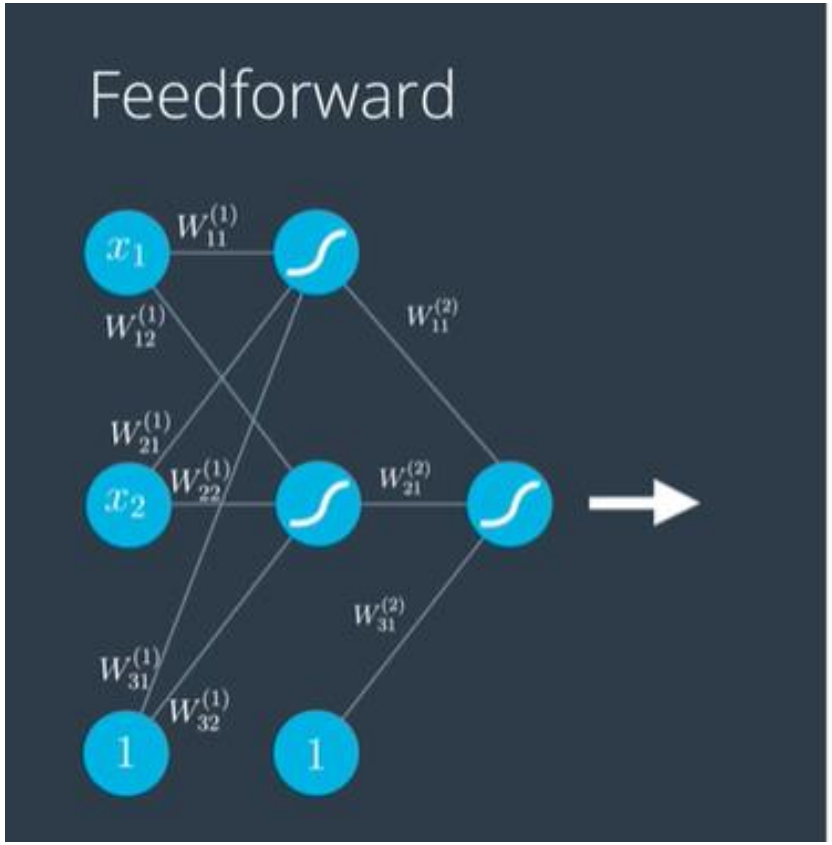
Feedforward

A more complex model



Feedforward

Other notation

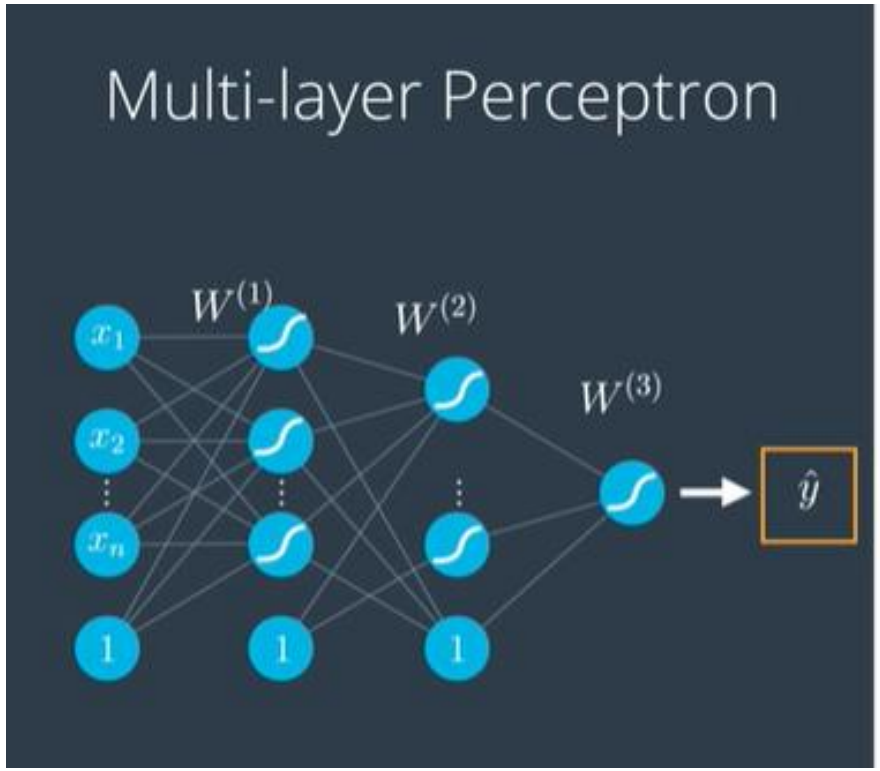


$$\hat{y} = \sigma \begin{pmatrix} W_{11}^{(2)} \\ W_{21}^{(2)} \\ W_{31}^{(2)} \end{pmatrix} \sigma \begin{pmatrix} W_{11}^{(1)} & W_{12}^{(1)} \\ W_{21}^{(1)} & W_{22}^{(1)} \\ W_{31}^{(1)} & W_{32}^{(1)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$$

$$\hat{y} = \sigma \circ W^{(2)} \circ \sigma \circ W^{(1)}(x)$$

Feedforward

Other notation



PREDICTION

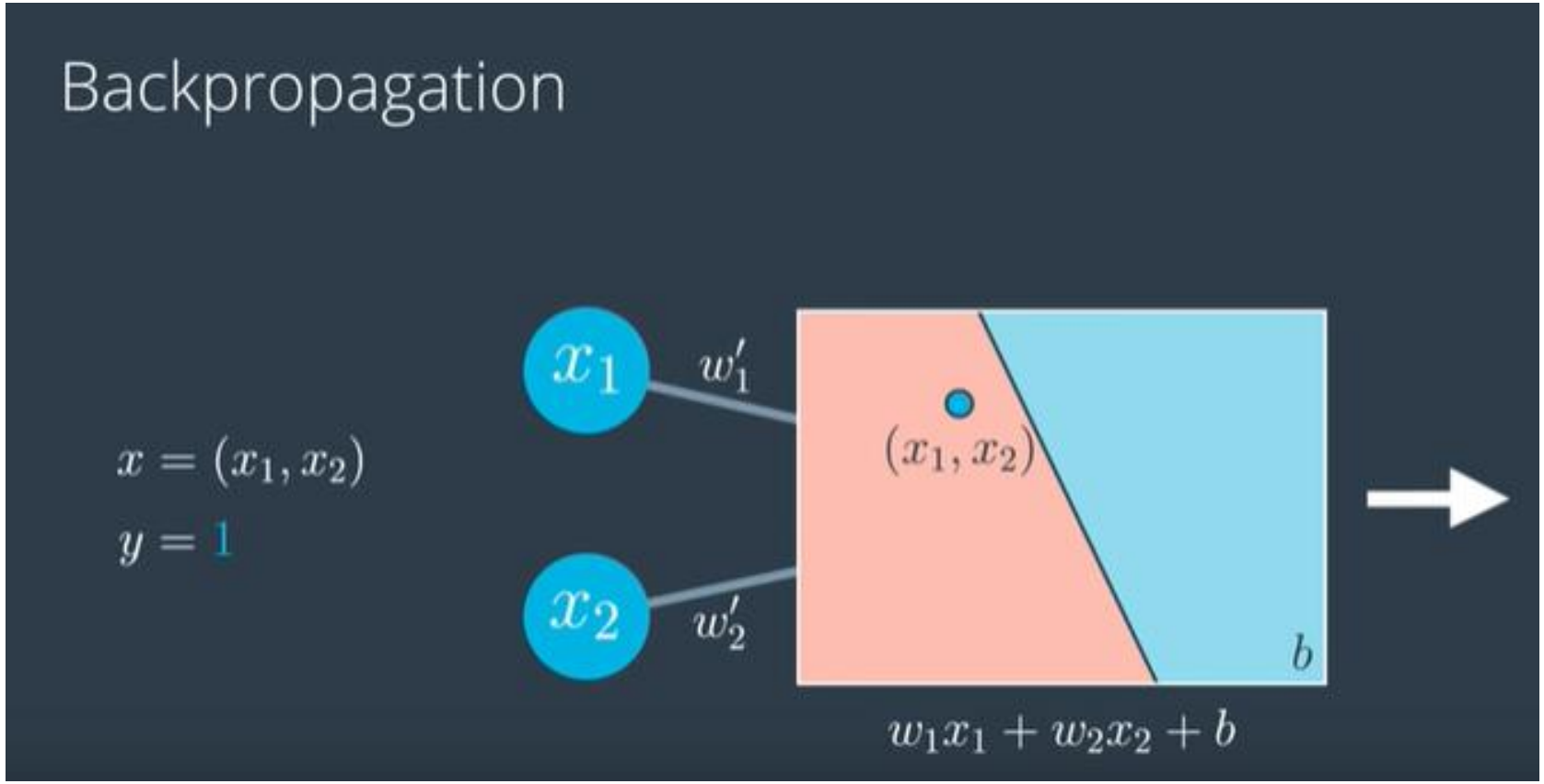
$$\hat{y} = \sigma \circ W^{(3)} \circ \sigma \circ W^{(2)} \circ \sigma \circ W^{(1)}(x)$$

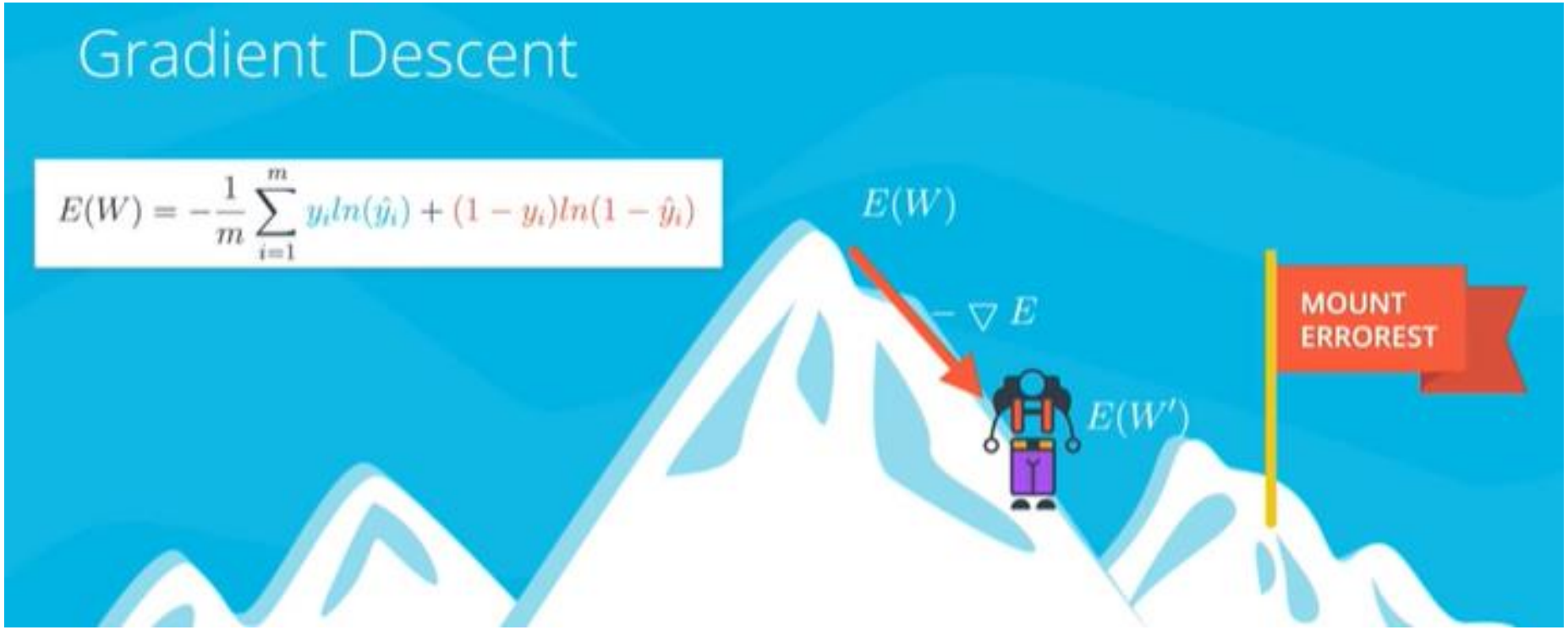
Backpropagation

In a nutshell, backpropagation will consist of:

- Doing a feedforward operation.
- Comparing the output of the model with the desired output.
- Calculating the error.
- Running the feedforward operation backwards (backpropagation) to spread the error to each of the weights.
- Use this to update the weights, and get a better model.
- Continue this until we have a model that is good.

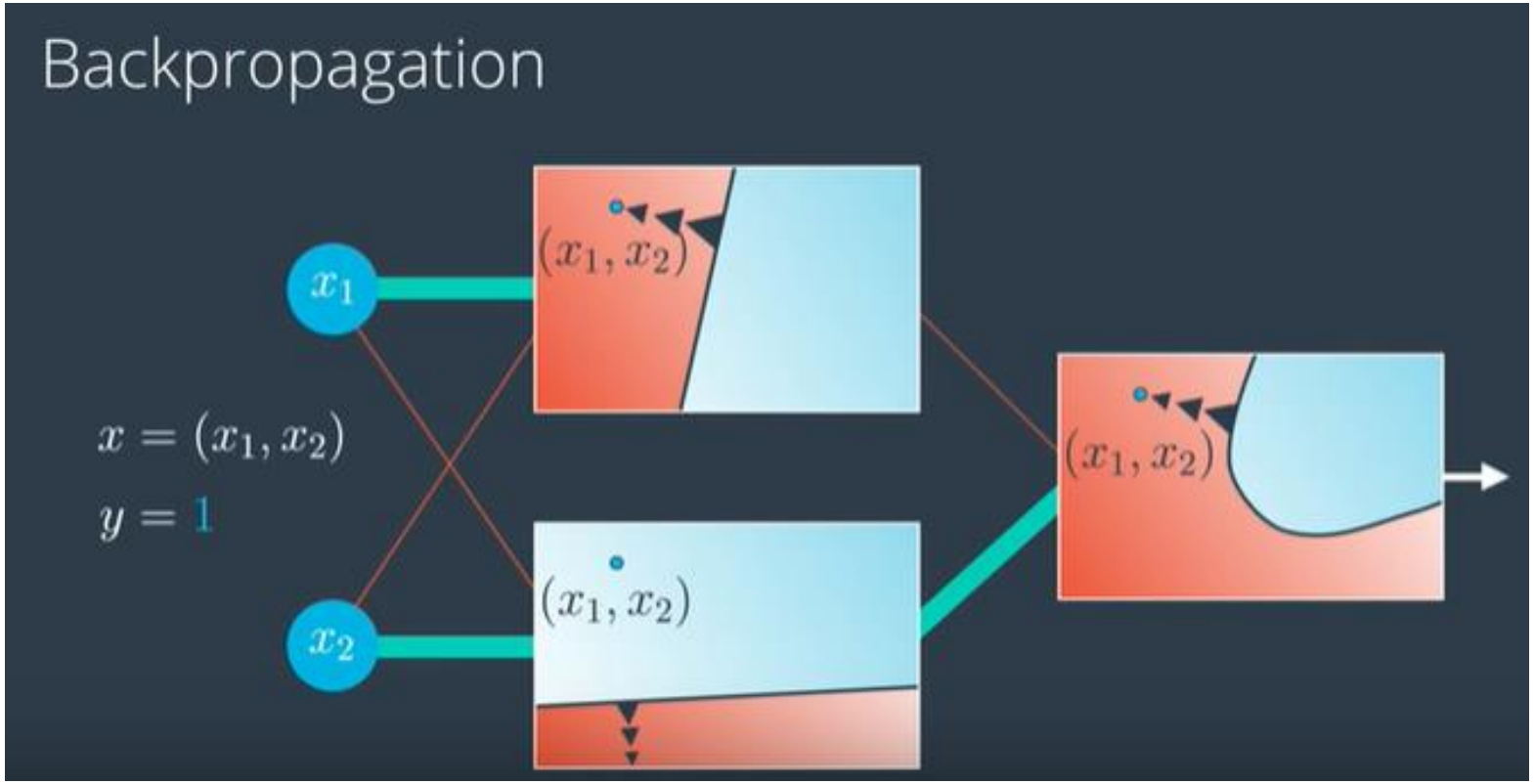
Backpropagation

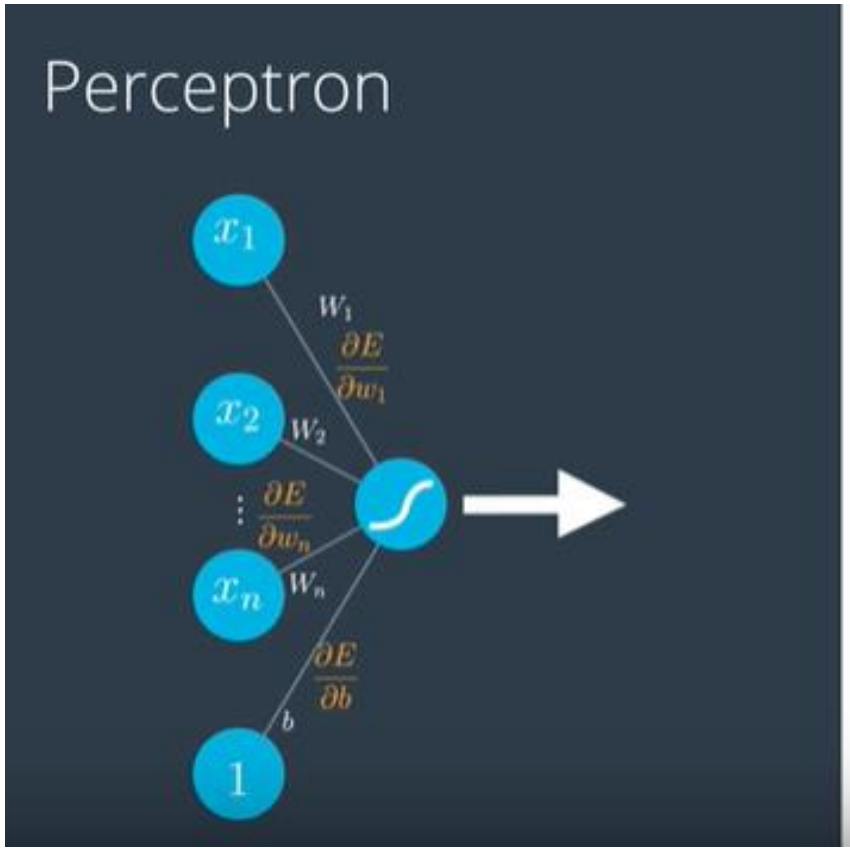




* Log-Loss vs Mean Squared Error

Backpropagation





PREDICTION

$$\hat{y} = \sigma(Wx + b)$$

ERROR FUNCTION

$$E(W) = -\frac{1}{m}\sum_{i=1}^{m} y_i ln(\hat{y}_i) + (1 - y_i) ln(1 - \hat{y}_i)$$

GRADIENT OF THE ERROR FUNCTION

$$\nabla E = \left(\frac{\partial E}{\partial w_1}, \ldots, \frac{\partial E}{\partial w_n}, \frac{\partial E}{\partial b}\right)$$

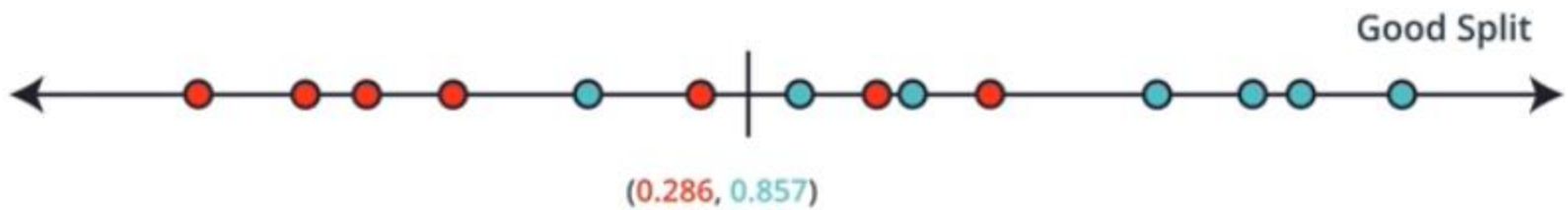
Evaluation Metrics - ROC

Evaluation Metrics - ROC

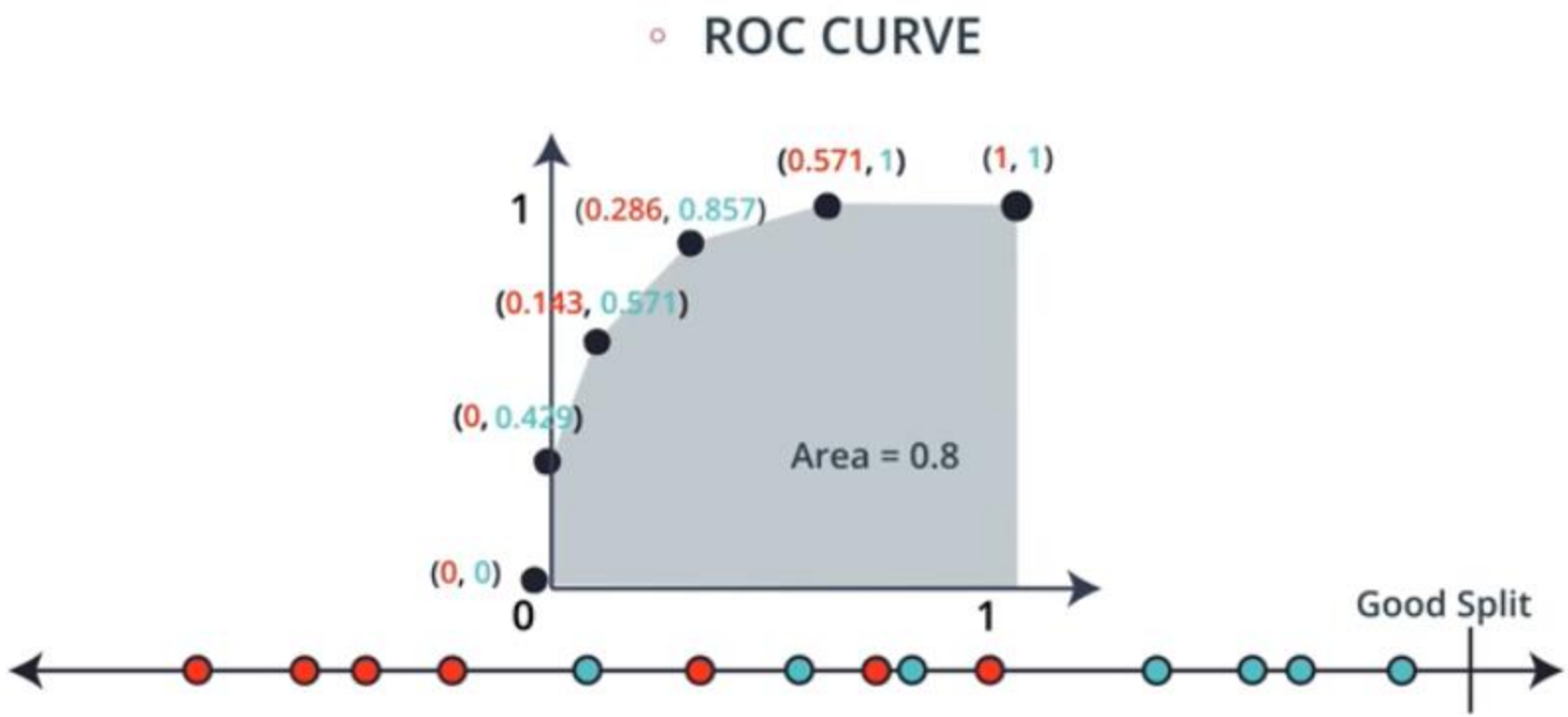
- ROC CURVE

$$\text{True Positive Rate} = \frac{\text{TRUE POSITIVES}}{\text{ALL POSITIVES}} = \frac{6}{7} =$$

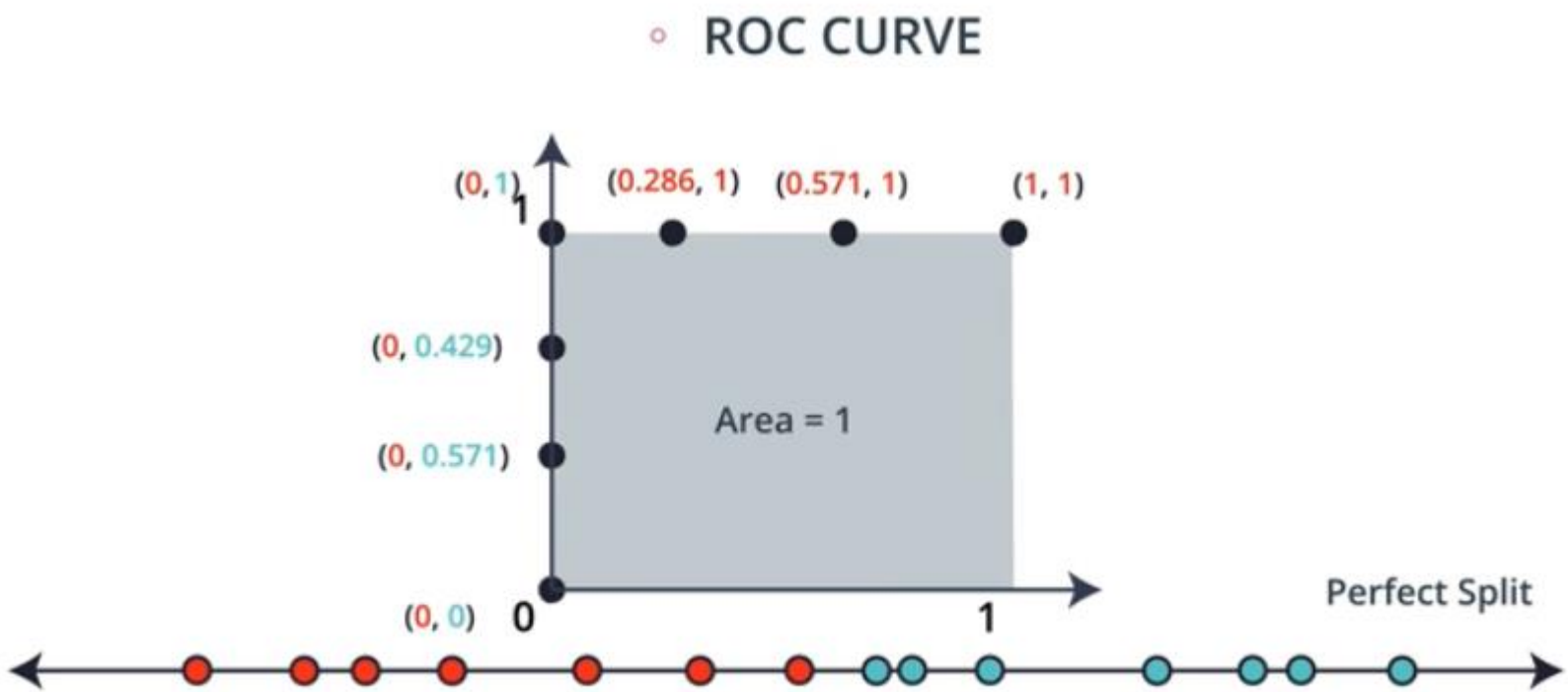
$$\text{False Positive Rate} = \frac{\text{FALSE POSITIVES}}{\text{ALL NEGATIVES}} = \frac{2}{7} =$$



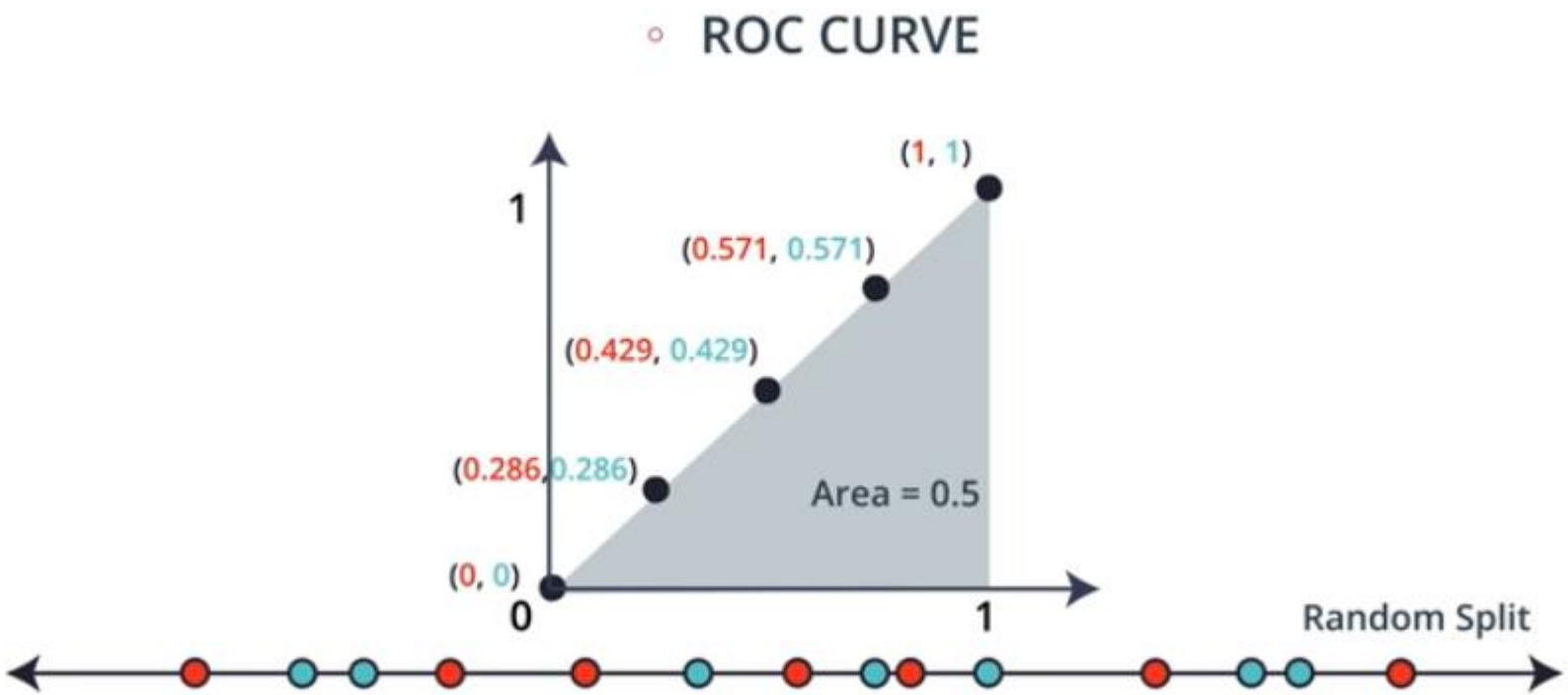
Evaluation Metrics - ROC



Evaluation Metrics - ROC



Evaluation Metrics – ROC



Investigar para el jueves: Diferencias, pros & cons de métricas: Accuracy, Precision, Recall, ROC, F1-Score (https://en.wikipedia.org/wiki/F1_score)

PCA – Key Ideas

Information loss:

- Maximal variance minimizes information loss

Cuándo usar?

- Detectar features latentes/ocultas significativas (patrones)
- Reducción de Dimensionalidad
  - Visualizar data de Alta Dimensionalidad
  - Reducir Ruido
  - Mejorar la performance de algoritmos seleccionando n-componentes

Aplicación:
https://scikit-learn.org/stable/auto_examples/applications/plot_face_recognition.html

PCA – Key Ideas