UE21CS343BB2
Topics in Deep Learning

Dr. Shylaja S S
Director of Cloud Computing & Big Data (CCBD), Centre for Data Sciences & Applied Machine Learning (CDSAML)
Department of Computer Science and Engineering
shylaja.sharath@pes.edu

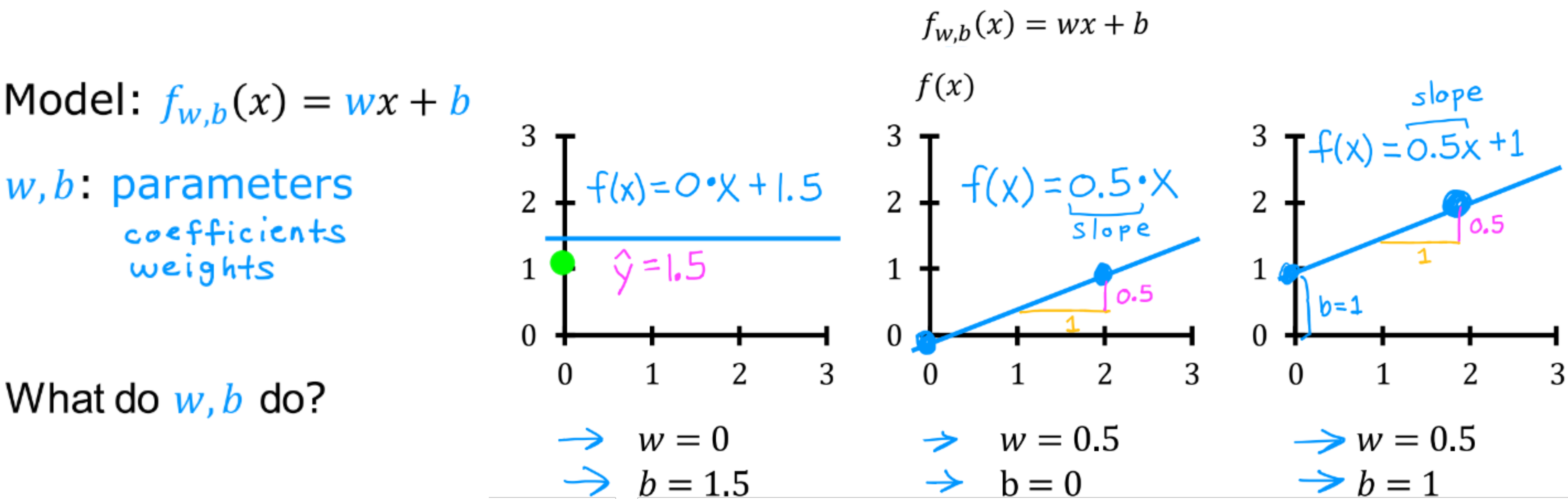
Ack: Aryan Sharma, Teaching Assistant

UE21CS343BB2: Topics in Deep Learning

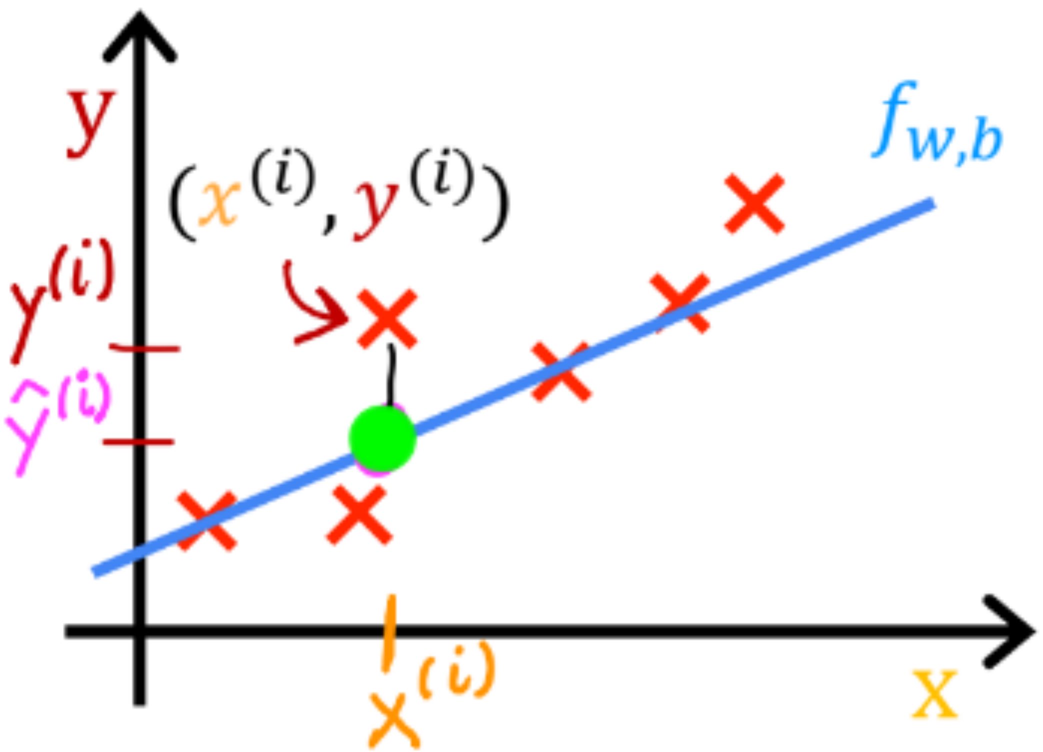
Loss Functions

What is a loss function?

- To measure how well the neural network is performing on a specific task, we need loss functions.
- The goal of Loss Function is to measure the error that the model made.
- This function has to be minimized in order to get the most accurate outputs from our model. Let’s look at how we will do this.



What is a loss function?



$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

$$f_{w,b}(x^{(i)}) = wx^{(i)} + b$$

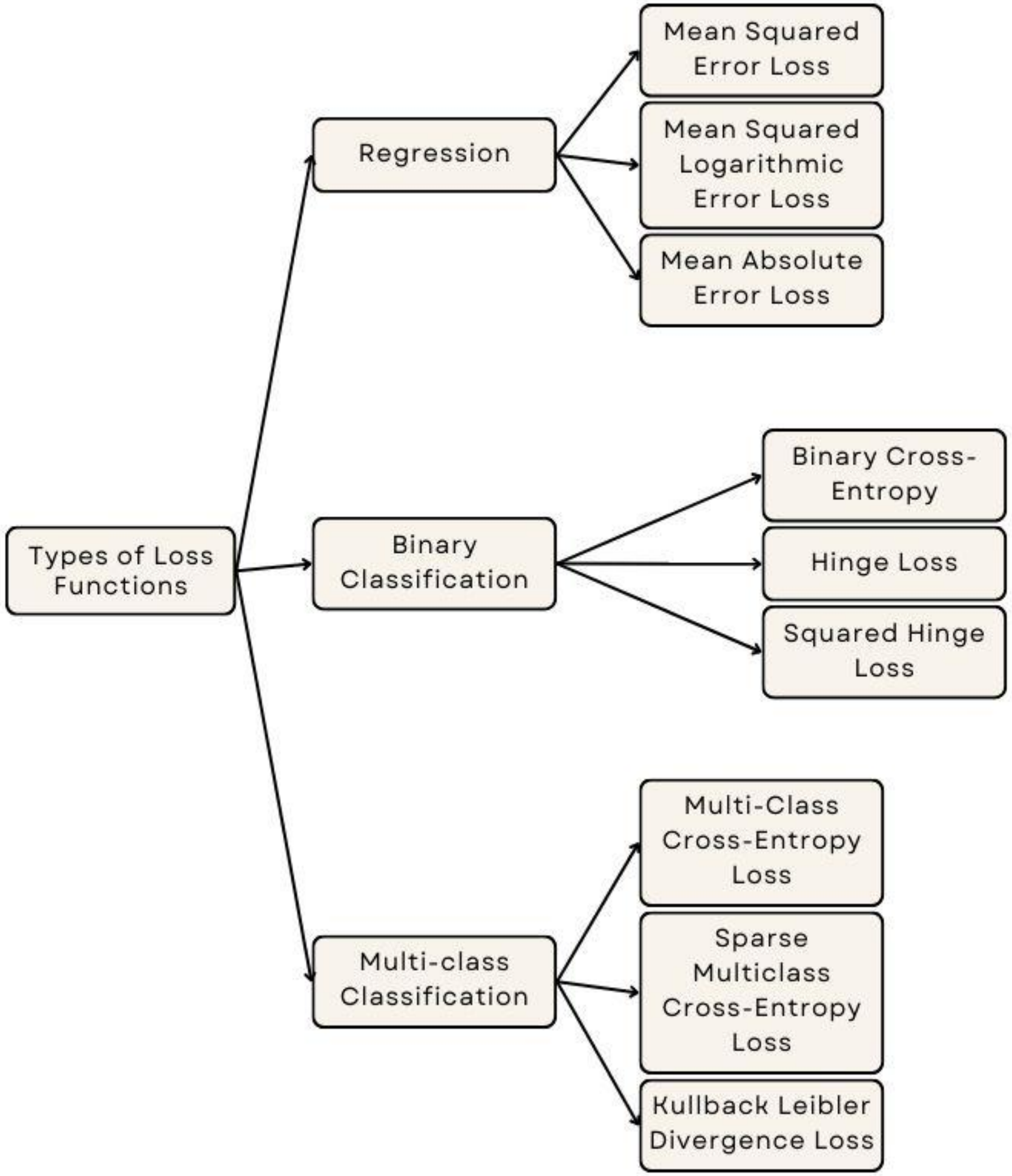
Let J(w,b) be a loss function called Mean Squared Error Loss Function.

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})^2$$

Find w, b:
$\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.

Thus, our goal is to choose the right weights/coefficients(w,b) to minimize J(w,b).

Types of Loss Functions



- The loss functions are selected based on the type of problem.
- Typically, this involves the difference between the actual value and approximated(predicted) value.
- Cross-entropy loss is often simply referred to as "cross-entropy," "logarithmic loss," "logistic loss," or "log loss" for short.

Types of Problems

We will consider 4 main cases to pair up problems with activation and loss functions.

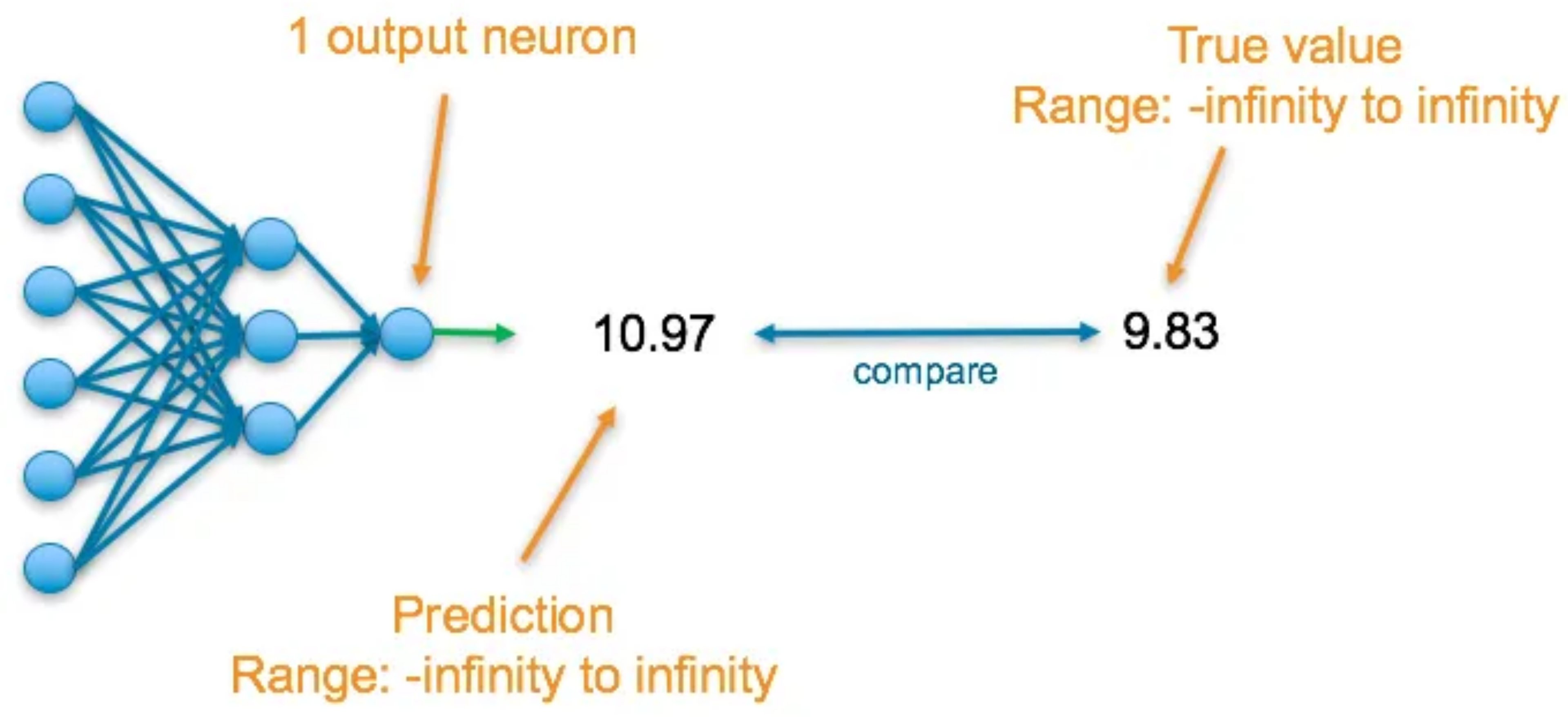
1. **Regression**: Predicting a **Numerical Value**

1. **Categorical**: Predicting a **Binary Outcome**

1. **Categorical**: Predicting a **Single Label** from **Multiple Classes**

1. **Categorical**: Predicting **Multiple Labels** from **Multiple Classes**

Case 1 : Regression - Predicting a Numerical Value



The final layer of the neural network will have <u>one neuron</u> and the value it returns is a <u>continuous numerical value</u>.
To evaluate the accuracy of the prediction, it is compared with the true value which is also a continuous number.

Regression Problems

There are three metrics which are generally used for evaluation of Regression problems (like Linear Regression, Decision Tree Regression, Random Forest Regression etc.):

1. **Mean Absolute Error (MAE):**

   This measures the absolute average distance between the real data and the predicted data, but it fails to punish large errors in prediction.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}|$$

Where, $\hat{y}$ - is predicted value of y and y_i - is actual value of y

2. Mean Square Error (MSE):

This measures the squared average distance between the real data and the predicted data. Here, larger errors are well noted (better than MAE).

But the disadvantage is that it also squares up the units of data as well. So, evaluation with different units is not at all justified.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

Where, $\hat{y}$ - is predicted value of y and y_i - is actual value of y

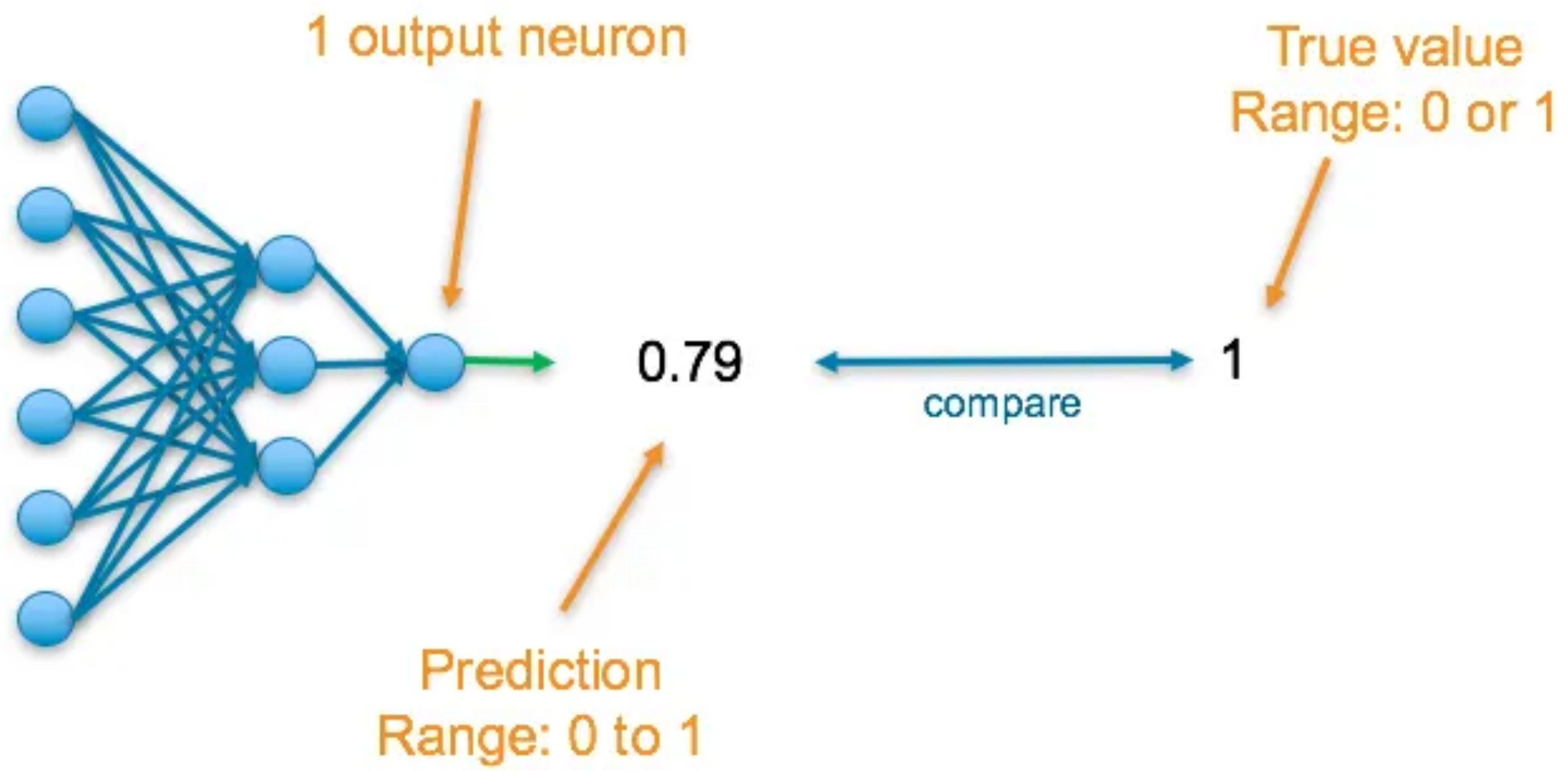
3. Root Mean Squared Error (RMSE):

This is actually the square root of MSE. Also, this metrics solves the problem of squaring the units.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2} = \sqrt{MSE}$$

Where, $\hat{y}$ - is predicted value of y and y_i - is actual value of y

Case 2 : Predicting a Binary Outcome



The final layer of the neural network will have <u>one neuron</u> and will return a <u>value between 0 and 1</u>, which can be inferred as a probability.
To evaluate the accuracy of the prediction, it is compared with the true value which is also a continuous number.

Case 2 : Predicting a Binary Outcome

Loss Function for Binary Classification Problems

- **Binary Cross Entropy**: Cross entropy quantifies the difference between two probability distribution.

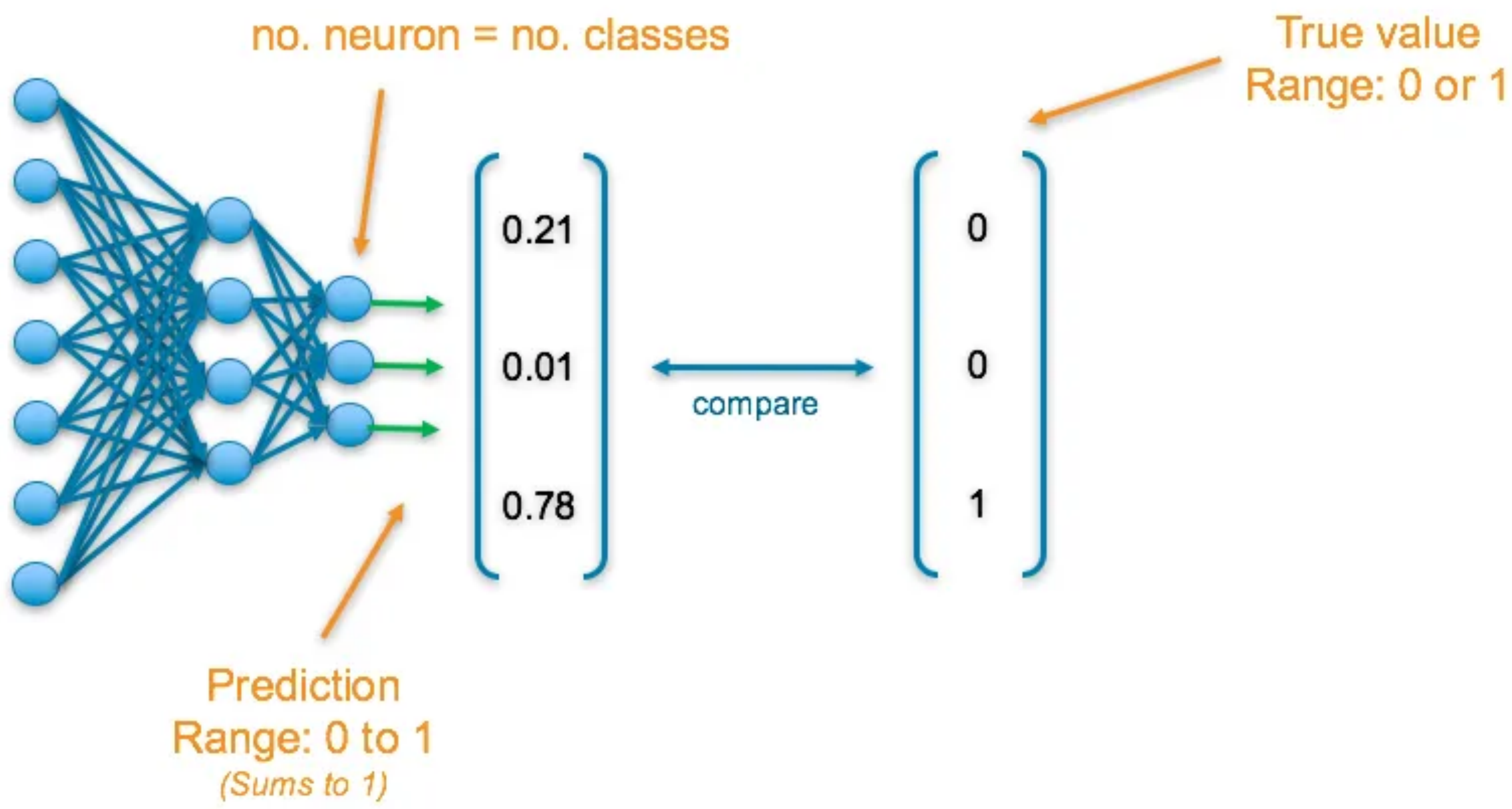
  The model predicts a model distribution of {p, 1-p} as we have a binary distribution.

  We use binary cross-entropy to compare this with the true distribution {y, 1-y}.

$$Binary\ Cross\ Entropy = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y}))$$

Where, $\hat{y}$ - is predicted value of y and y_i - is actual value of y

Case 3 : Predicting a Single Label from Multiple Classes



The final layer of the neural network will have <u>one neuron for each of the classes</u> and they will return a value between 0 and 1, which will be the <u>probability of it being the class</u>.

The output is then a probability distribution and will sum to 1.

Case 3 : Predicting a Single Label from Multiple Classes

PES UNIVERSITY

Loss Function for Categorical Problems

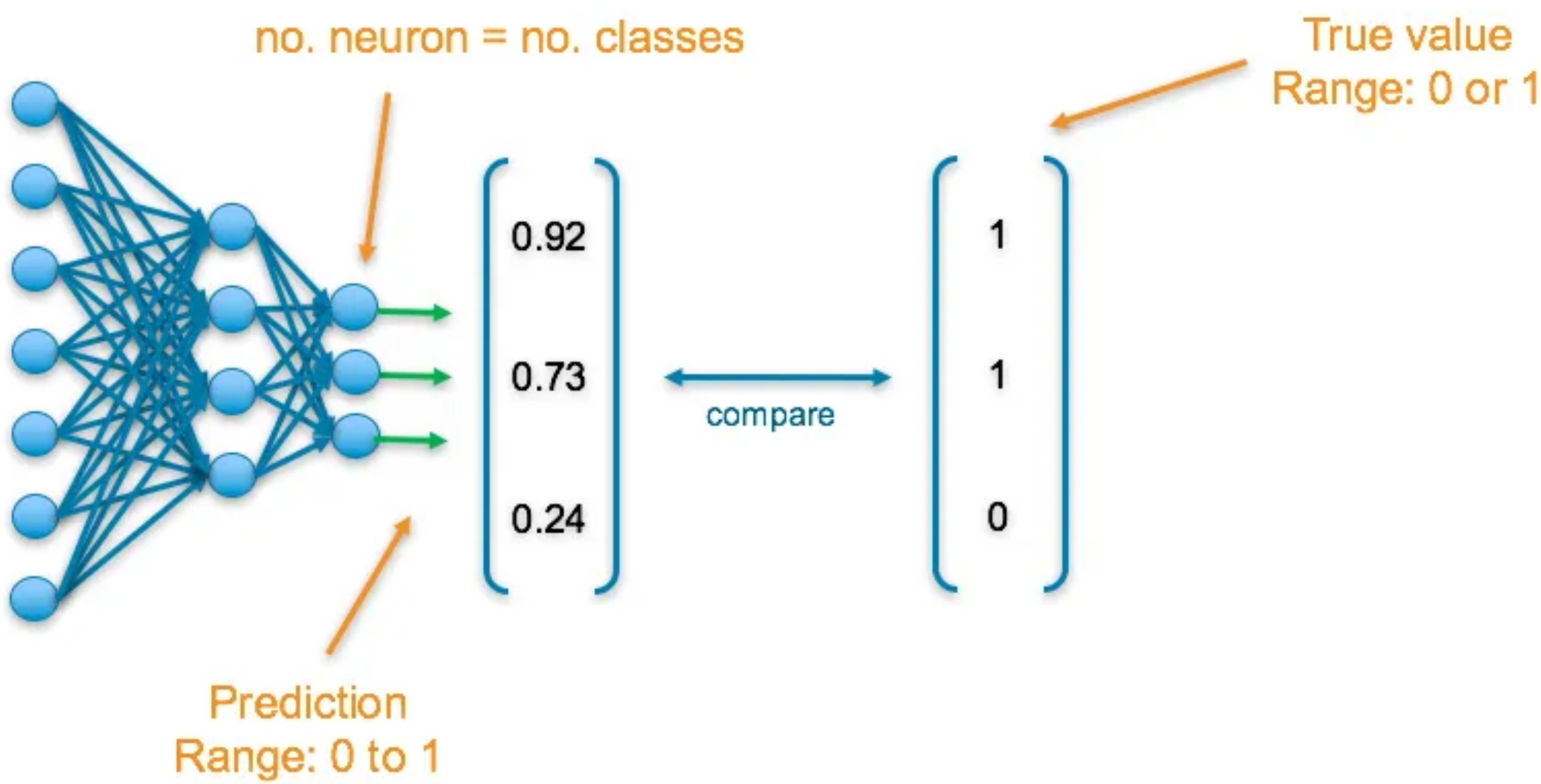
- **Cross Entropy**: Cross entropy quantifies the difference between two probability distribution. Our model predicts a model distribution of {p1, p2, p3} (where p1+p2+p3 = 1).

  We use cross-entropy to compare this with the true distribution {y1, y2, y3}.

$$Cross\ Entropy = -\sum_{i}^{n} y_i \log(p_i)$$

Where, y_i − is actual/true label, p_i − probability for the i^{th} class and n − is the number of classes

Case 4 : Predicting Multiple Labels from Multiple Classes



The final layer of the neural network will have one neuron for each of the classes and they will return a value between 0 and 1, which will be the probability of it being the class.
The output is then a probability distribution and will sum to 1.

Case 4 : Predicting Multiple Labels from Multiple Classes

Loss Function for Categorical Problems

- **Binary Cross Entropy**: Cross entropy quantifies the difference between two probability distribution. Our model predicts a model distribution of {p, 1-p} (binary distribution) for each of the classes.

  We use binary cross-entropy to compare these with the true distributions {y, 1-y} for each class and sum up their results.

$$Binary\ Cross\ Entropy = -\frac{1}{n}\sum_{i}^{n}(y_i \log(p_i) + (1-y_i)\log(1-p_i))$$

Where, y_i − is actual/true label, p_i − probability for the i^{th} class and n − is the number of classes

What's the right loss function?

Problem Type	Output Type	Final Activation Function	Loss Function
Regression	Numerical value	Linear	Mean Squared Error (MSE)
Classification	Binary outcome	Sigmoid	Binary Cross Entropy
Classification	Single label, multiple classes	Softmax	Cross Entropy
Classification	Multiple labels, multiple classes	Sigmoid	Binary Cross Entropy

Although picking a loss function is not given much importance and overlooked, one must understand that there is no one-size-fits-all and choosing a loss function is as important as choosing the right machine learning model for the problem in hand.

The <u>choice of a loss function</u> is tightly coupled with the <u>choice of output unit</u> in a model.

Factors to consider when selecting a loss function

Factor	Description
Type of Learning Problem	Classification vs Regression; Binary vs Multiclass Classification.
Model Sensitivity to Outliers	Some loss functions are more sensitive to outliers (e.g., MSE), while others are more robust (e.g., MAE).
Desired Model Behavior	Influences how the model behaves, e.g., hinge loss in SVMs focuses on maximizing the margin.
Computational Efficiency	Some loss functions are more computationally intensive, impacting the choice based on available resources.
Convergence Properties	The smoothness and convexity of a loss function can affect the ease and speed of training.
The scale of the Task	For large-scale tasks, a loss function that scales well and can be efficiently optimized is crucial.

Acknowledgements & References

- https://towardsdatascience.com/deep-learning-which-loss-and-activation-functions-should-i-use-ac02f1c56aa8
- https://www.deeplearningbook.org/
- https://www.analyticsvidhya.com/blog/2022/06/understanding-loss-function-in-deep-learning/
- https://machinelearningknowledge.ai/cost-functions-in-machine-learning/
- https://deeplearning.ai
- https://www.datacamp.com/tutorial/loss-function-in-machine-learning

PES UNIVERSITY

UE21CS343BB2

Topics in Deep Learning

Dr. Shylaja S S
Director of Cloud Computing & Big Data (CCBD), Centre for Data Sciences & Applied Machine Learning (CDSAML)
Department of Computer Science and Engineering
shylaja.sharath@pes.edu

Ack: Aryan Sharma, Teaching Assistant