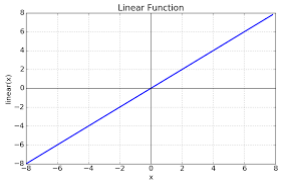
1) Regression

a) Activation Functions

- **Linear (No Activation Functions):**

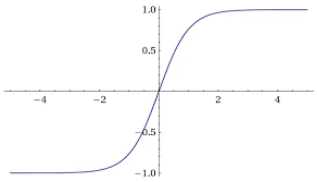
$$\text{Equation: } f(x) = x$$



  - **When Used**: Standard regression problems where the output is a real number.

- **Tanh**:

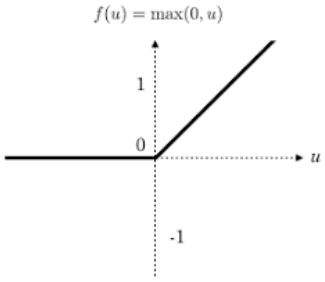
$$\text{Equation: } f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$$



  - **When Used**: Regression problems where the output needs to be scaled between -1 and 1.

- **ReLU**:

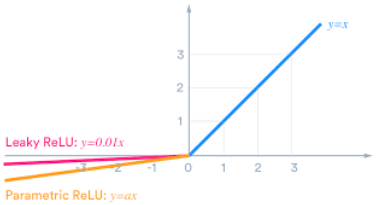
$$\text{Equation: } f(x) = \max(0, x)$$



  - **When Used**: Specific cases of regression problems where non-linearity is needed in the output.

- **Leaky ReLU**:

Equation: $f(x) = \max(0.01x, x)$



  - **When Used**: To avoid the "dying ReLU" problem in regression tasks.

b) Cost Functions

- **Mean Squared Error (MSE)**:

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

  - **When Used**: Standard regression problems to minimize the squared difference between predicted and actual values.

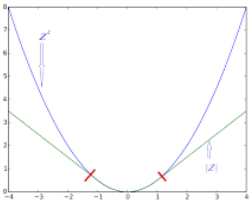
- **Mean Absolute Error (MAE)**:

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

  - **When Used**: Regression problems to minimize the absolute difference between predicted and actual values.

- **Huber Loss**:
  - **Equation**:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \le \delta \\ \delta(|a| - \frac{1}{2}\delta) & \text{for } |a| > \delta \end{cases}$$



  - **When Used**: Regression problems to be less sensitive to outliers than MSE.

c) Examples

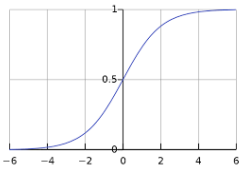
- **Predicting house prices**
- **Predicting stock prices**
- **Predicting normalized values that can take both positive and negative values**

2) Binary Classification

a. Activation Functions

- **Sigmoid**:

$$\text{Equation: } f(x) = \frac{1}{1+e^{-x}}$$



  - **When Used**: Binary classification problems where the output is a probability between 0 and 1.

b. Cost Functions

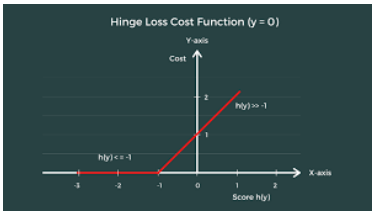
- **Binary Cross-Entropy** (Log Loss):

$$\text{Equation: } \text{BCE} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$$

  - **When Used**: Binary classification problems to measure the performance of a classification model whose output is a probability value.

- **Hinge Loss**:

$$\text{Equation: } \text{Hinge Loss} = \frac{1}{n}\sum_{i=1}^{n}\max(0, 1 - y_i \cdot \hat{y}_i)$$



  - **When Used**: Binary classification problems, often used with Support Vector Machines (SVMs).

c. Examples

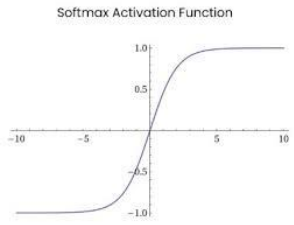
- **Predicting whether an email is spam or not**
- **Predicting whether a customer will buy a product (yes/no)**

3) Multi-Class Classification

A. Activation Functions

- **SoftMax**:

$$\text{Equation: } f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$



  - **When Used**: Multi-class classification problems where the output is a probability distribution over multiple class.

B. Cost Functions

- **Categorical Cross-Entropy**:

$$\text{Equation: } \mathrm{CCE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} y_{i,c}\log(\hat{y_{i,c}})$$

  - **When Used**: Multi-class classification problems to measure the performance of a classification model whose output is a probability value over multiple classes.

- **Sparse Categorical Cross-Entropy**:

$$\text{Equation: } \mathrm{SCCE} = -\frac{1}{n}\sum_{i=1}^{n}\log(\hat{y_{i,c}})$$

  - **When Used**: Multi-class classification problems when labels are integers rather than one-hot encoded vectors.

C. Examples

- **Classifying images of handwritten digits into one of 10 classes (0-9)**
- **Classifying types of animals in an image dataset (e.g., cat, dog, horse)**

4) Summary

a) Regression:

- **Activation Functions**: Linear, Tanh, ReLU, Leaky ReLU
  - Linear: $f(x) = x$
  - Tanh: $f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$
  - ReLU: $f(x) = \max(0, x)$
  - Leaky ReLU: $f(x) = \max(0.01x, x)$
- **Cost Functions**: Mean Squared Error (MSE), Mean Absolute Error (MAE), Huber Loss
  - MSE: $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
  - MAE: $\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
  - Huber Loss:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \le \delta \\ \delta(|a| - \frac{1}{2}\delta) & \text{for } |a| > \delta \end{cases}$$

b) Binary Classification:

- **Activation Function**: Sigmoid
  - Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$
- **Cost Functions**: Binary Cross-Entropy, Hinge Loss
  - Binary Cross-Entropy: $\mathrm{BCE} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$
  - Hinge Loss: $\text{Hinge Loss} = \frac{1}{n}\sum_{i=1}^{n}\max(0, 1 - y_i \cdot \hat{y}_i)$

c) Multi-Class Classification:

- **Activation Function**: Softmax
  - Softmax: $f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$
- **Cost Functions**: Categorical Cross-Entropy, Sparse Categorical Cross-Entropy
  - Categorical Cross-Entropy: $\mathrm{CCE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} y_{i,c}\log(\hat{y_{i,c}})$
  - Sparse Categorical Cross-Entropy: $\mathrm{SCCE} = -\frac{1}{n}\sum_{i=1}^{n}\log(\hat{y_{i,c}})$