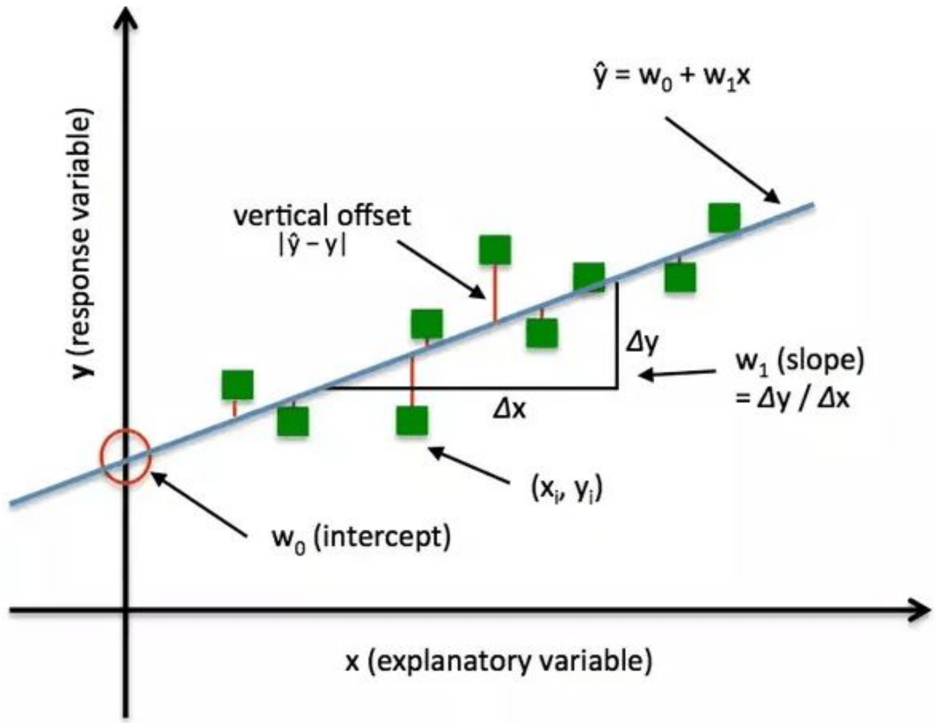
CSE4261: Neural Network and Deep Learning

Lecture: 21.05.2025

Sangeeta Biswas, Ph.D.
Associate Professor,
University of Rajshahi, Rajshahi-6205, Bangladesh

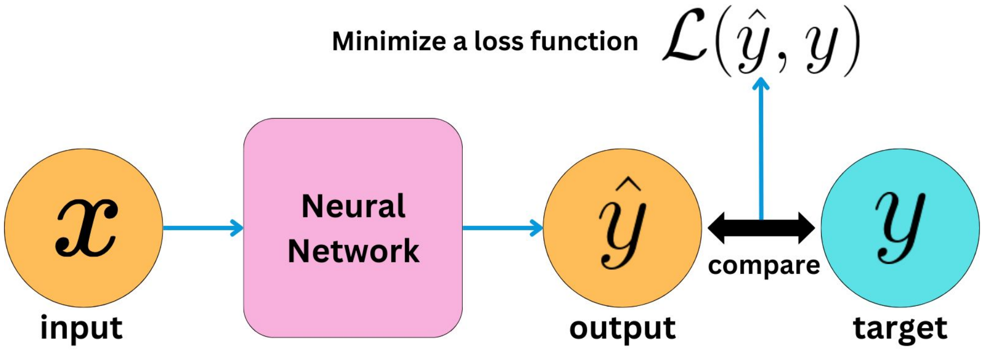
Loss

- Error is a measurement of the difference between the predicted output of the model and the actual output.
- Loss is the weighted error.
- Different types of loss:
  - Absolute loss: $|\hat{y} - y|$
  - Squared loss: $(\hat{y} - y)^2$
  - Log loss: $- y.\log(\hat{y})$



Loss Function

- A loss function is a function which estimates the total loss of all data samples.
- It is used to find optimum values of parameters of a neural network during training.
- It is a function of parameters.
- Different ways of writing:
    - $J(\omega)$; $L(\theta|x)$; $\ell(\omega)$; $\mathcal{L}(\hat{y} - y)$;



Different Loss Functions

Different loss functions penalizes wrong predictions differently

<u>Popular losses:</u>

- Regression:
  - Mean-Squared Error (MSE)
  - Mean Absolute Error (MAE)

- Classification:
  - Binary Cross-Entropy
  - Categorical Cross-Entropy
  - Focal Cross-Entropy
  - Hinge Loss

L1 & L2 Loss

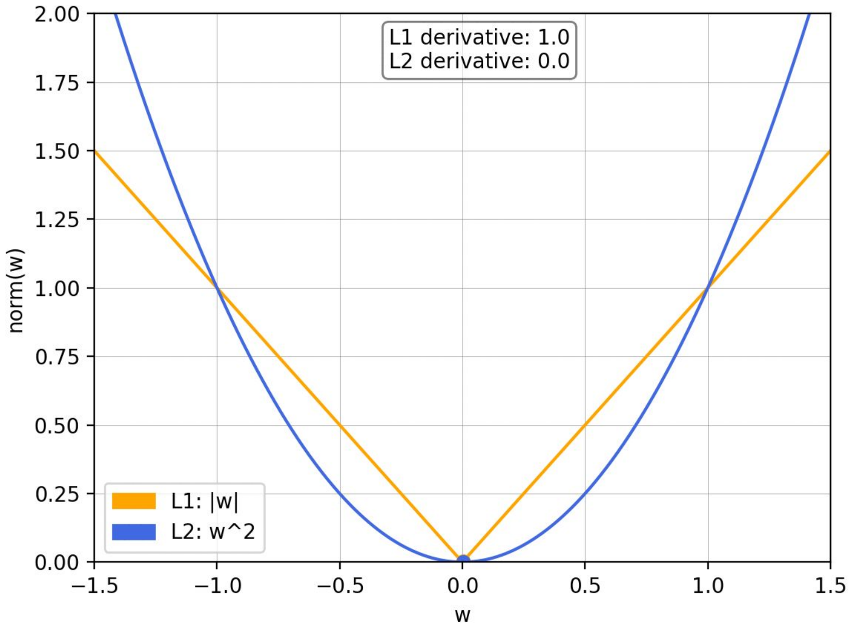
- Absolute Loss is also known as L1 loss
- Squared Error Loss is also known as L2 loss

$$L_1 = \sum_{i=1}^{n} |y_{gt} - y_{pred}|$$

$$L_2 = \sum_{i=1}^{n} (y_{gt} - y_{pred})^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_{gt} - y_{pred}|$$

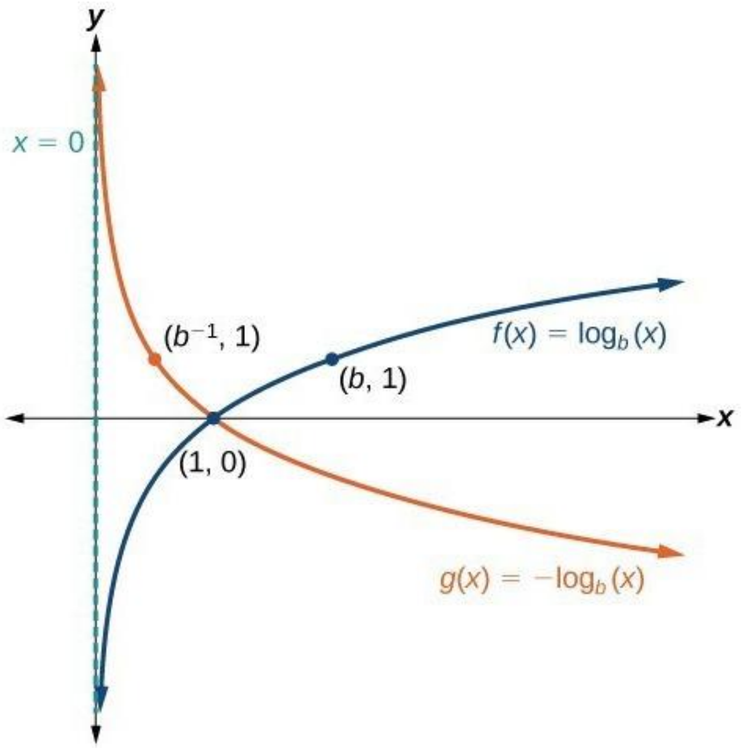
$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_{gt} - y_{pred})^2$$



Negative Log

$g(x) = -\log_b(x)$

- domain $(0,\infty)$
- range, $(-\infty,\infty)$
- vertical asymptote $x = 0$



Entropy

Entropy:

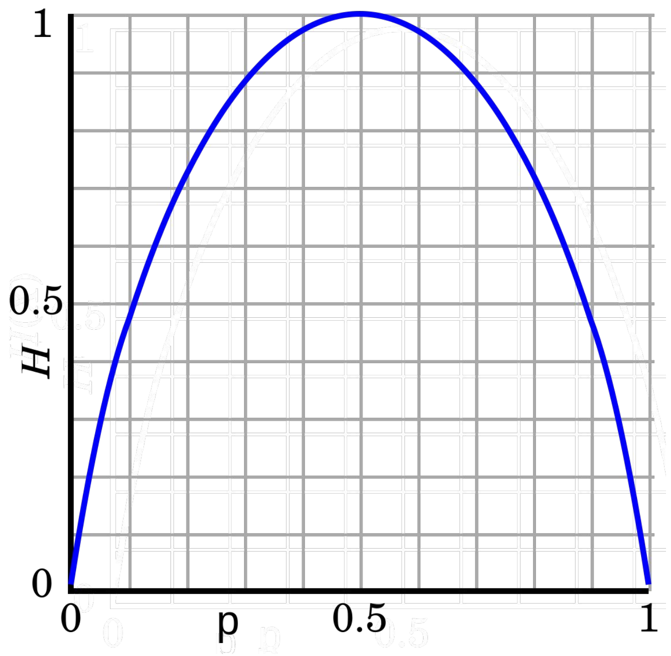
- $H(p) = -\Sigma\ p\_i \times log(p\_i)$
- the average amount of "surprise" or uncertainty associated with a random variable.

Entropy vs Cross-Entropy:

- Entropy measures the inherent uncertainty or randomness of a single distribution
- Cross-entropy measures the difference between two probability distributions.
    - $H(P, Q) = -P \times log(Q)$

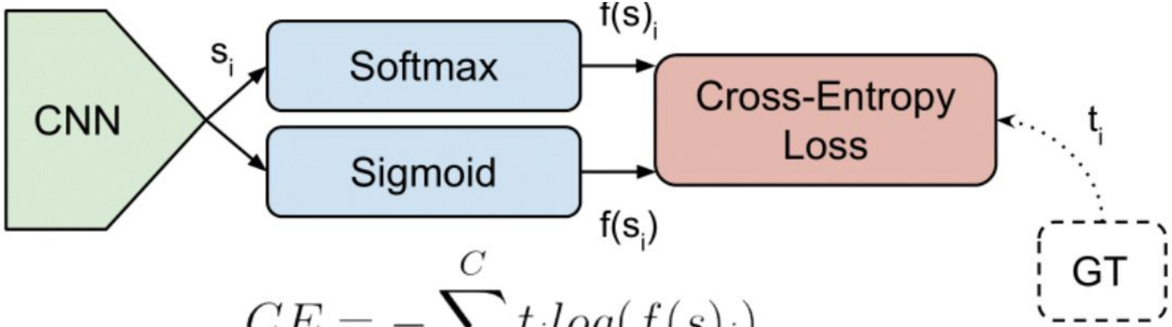
Binary Cross-Entropy:

- $H(p) = -p \times log(p) + (1-p) \times log(1-p)$



Cross-Entropy Loss

- It is also known as Cross-entropy log loss
- It is based on the probability of a model's output

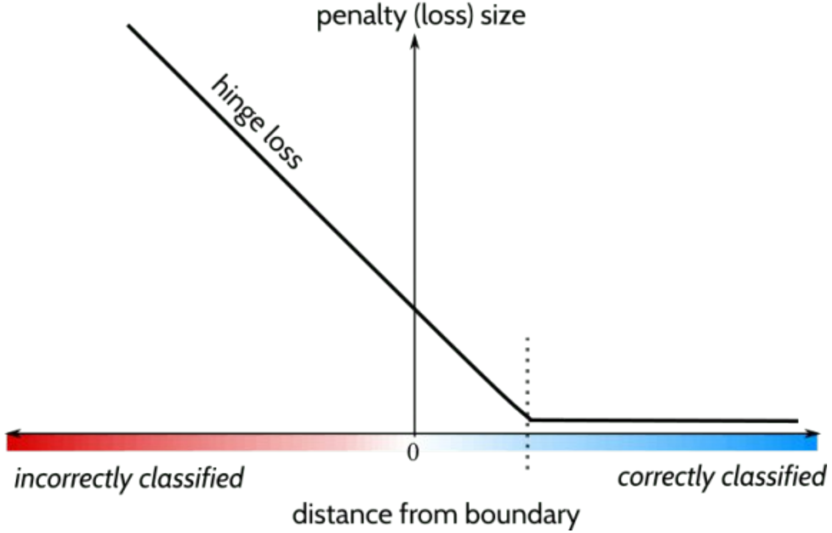


$$CE = -\sum_{i}^{C} t_i log(f(s)_i)$$

$$CE = -\sum_{i=1}^{C'=2} t_i log(f(s_i)) = -t_1 log(f(s_1)) - (1 - t_1) log(1 - f(s_1))$$

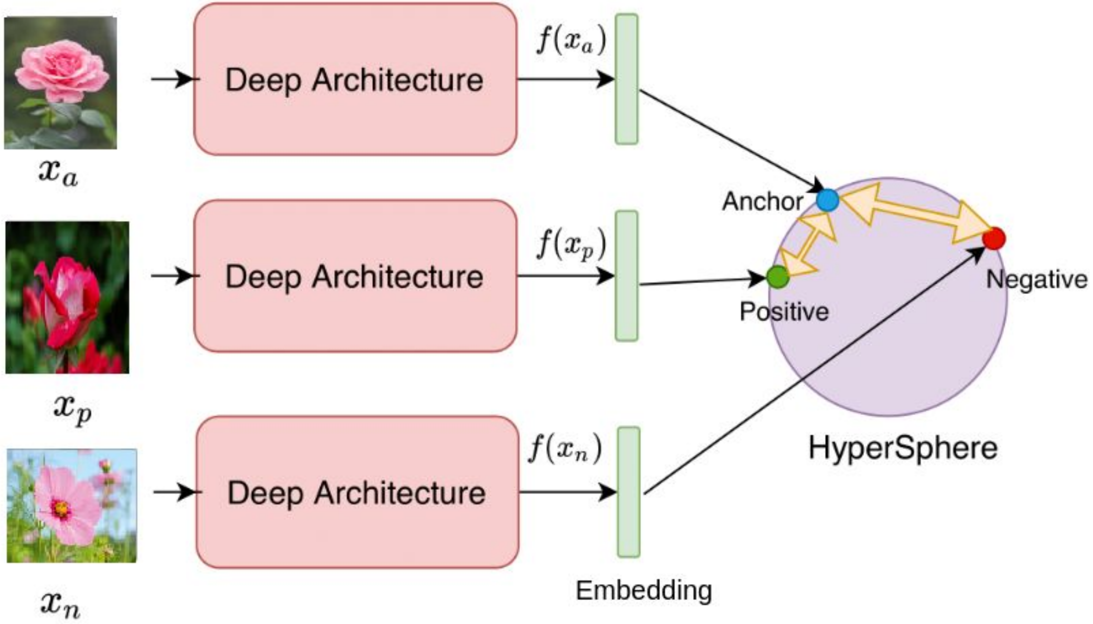
Hinge Loss

- It penalizes misclassified or correctly classified predictions which are too close to the decision boundary in a linear way.
  - Predictions that are far from the decision boundary get more punishments.
- It is generally used in Support Vector Machine based classifier.



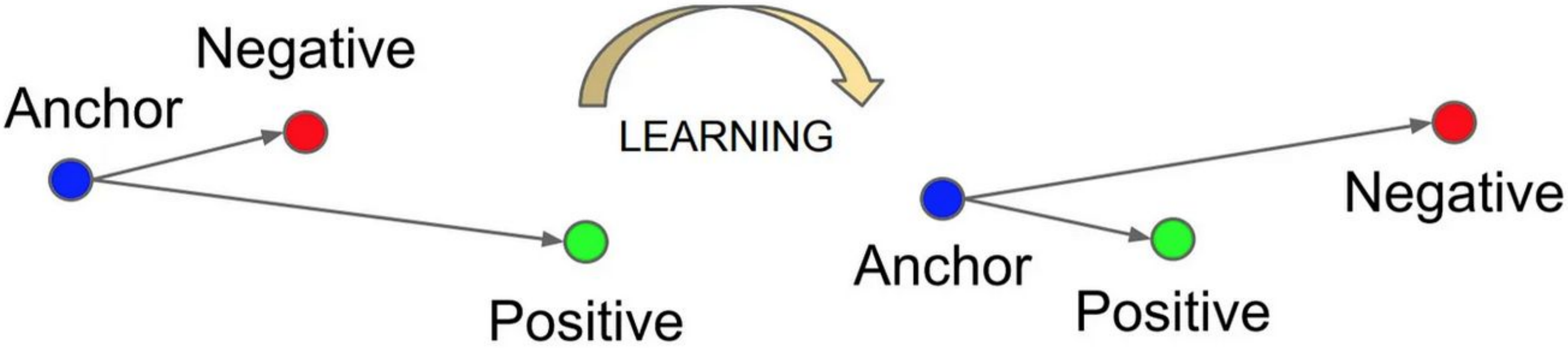
Triplet Loss

In triplet loss, a reference input (called anchor) is compared to a matching input (called positive sample) and a non-matching input (called negative sample)



Triplet Loss

- minimizes the distance between an anchor and a positive
- maximizes the distance between the Anchor and a negative of a different identity.



Triplet Loss

L(a, p, n) = max(d(a, p) - d(a,n) + α, 0)

where

- $d(a, p)$: distance between anchor and positive sample
- $d(a,n)$: distance between anchor and negative sample
- α: margin of error

- Lecture: https://www.youtube.com/watch?v=d2XB5-tuCWU

Training Neural Network

During training we try to find the parameter set for which loss is minimum.
Starting at random values, we update parameters by gradient descent algorithm

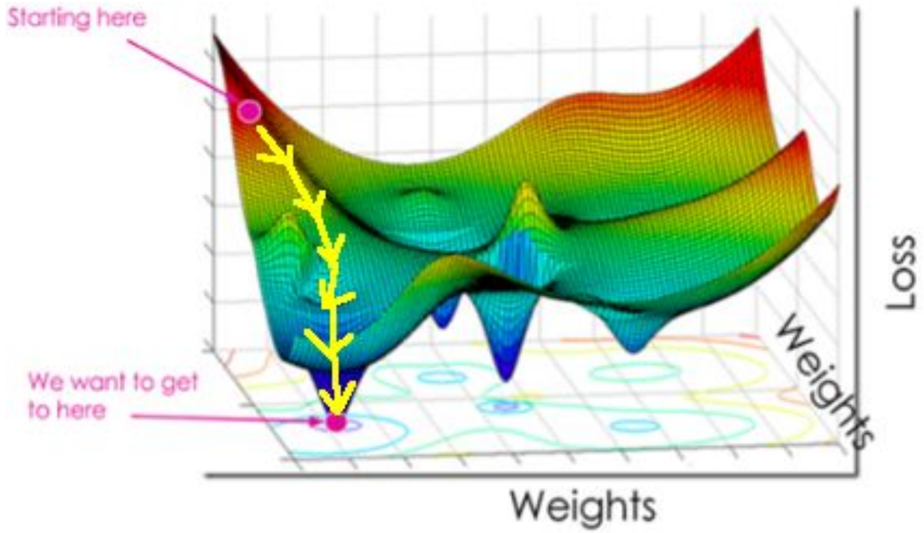


Image Source: Google Search Engine