Cost Function and Lost Function

The loss function is to capture the difference between the actual and predicted values for a single record whereas cost functions aggregate the difference for the entire training dataset.

The Most commonly used loss functions are Mean-squared error and Hinge loss.

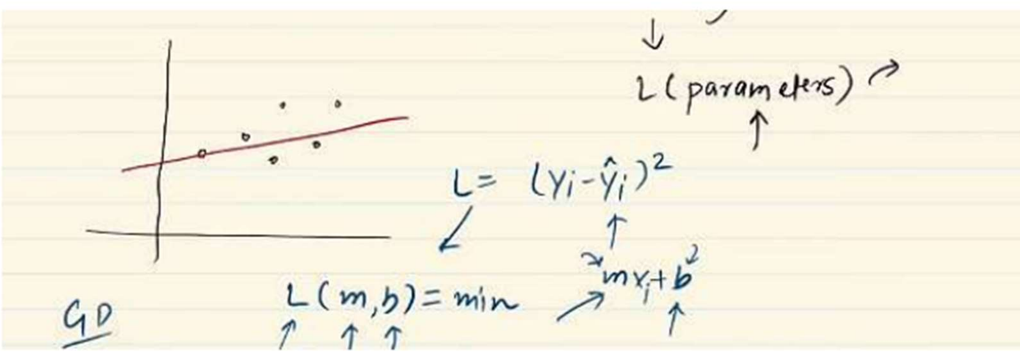
Note: If distance is high means model is poor

- Cumulative error – cost function
- Individual error – lost function

Loss function is a method of evaluating how well your algorithm is modelling your dataset.

- *High means Poor Model*
- *Low means great Model*

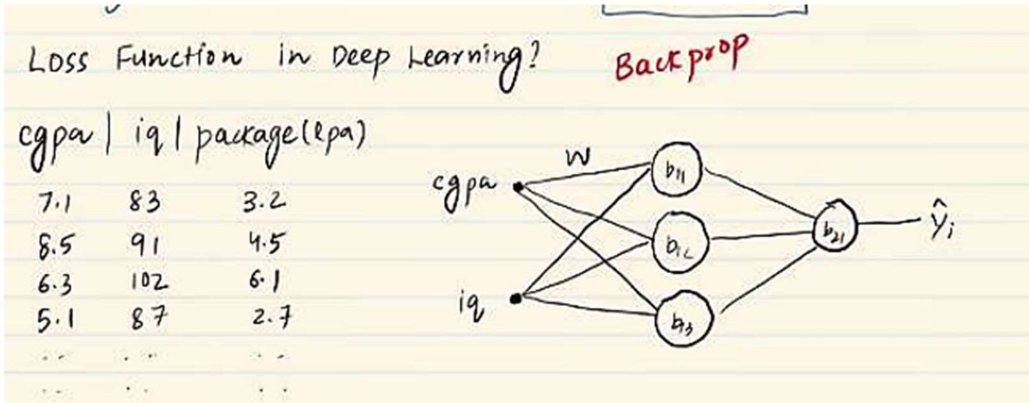
In this graph every row has its own error, so sum of all three errors gives – Mean Absolute Error (MSE)



Note: We don't usually use mean absolute error because it could be negative, it can be solve by:

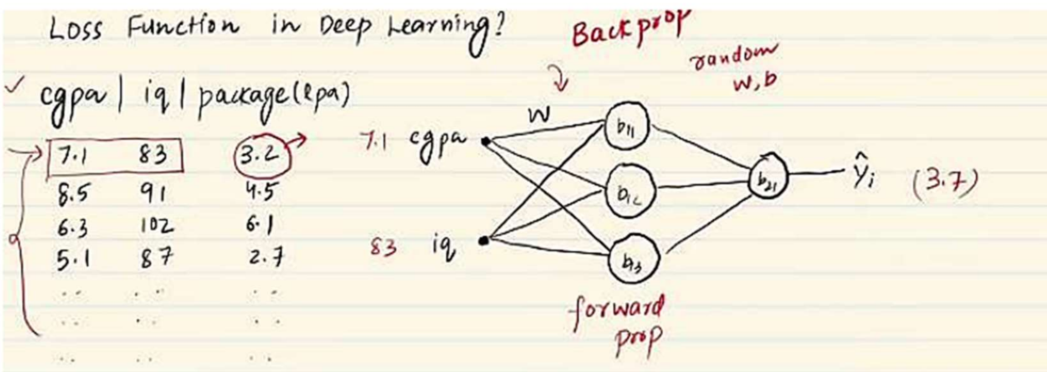
- Mean square error – but it doubles the actual value
- Root mean square error – it remains close to actual value – majorly use this

---

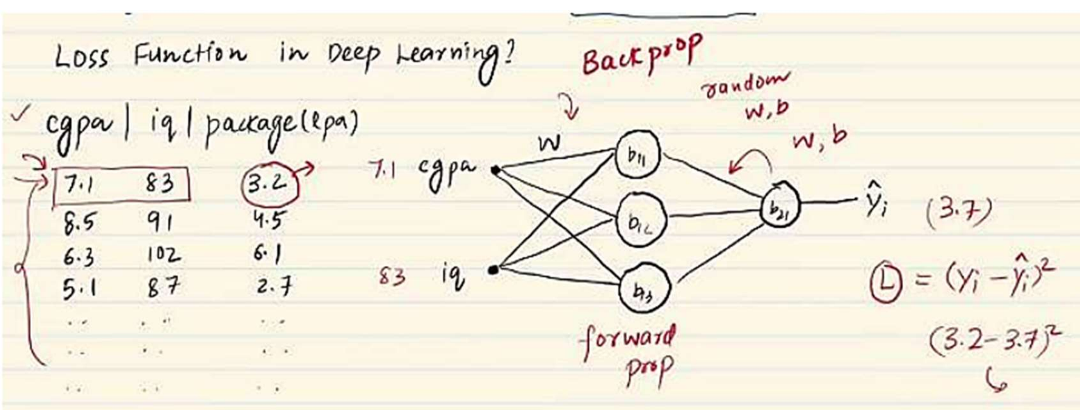


The data is continues which means it is regression

For First Row – cgpa 7.1, iq 83, ans 3.2 (but model predicted 3.7 using equation y = mx + c) – Means error generated

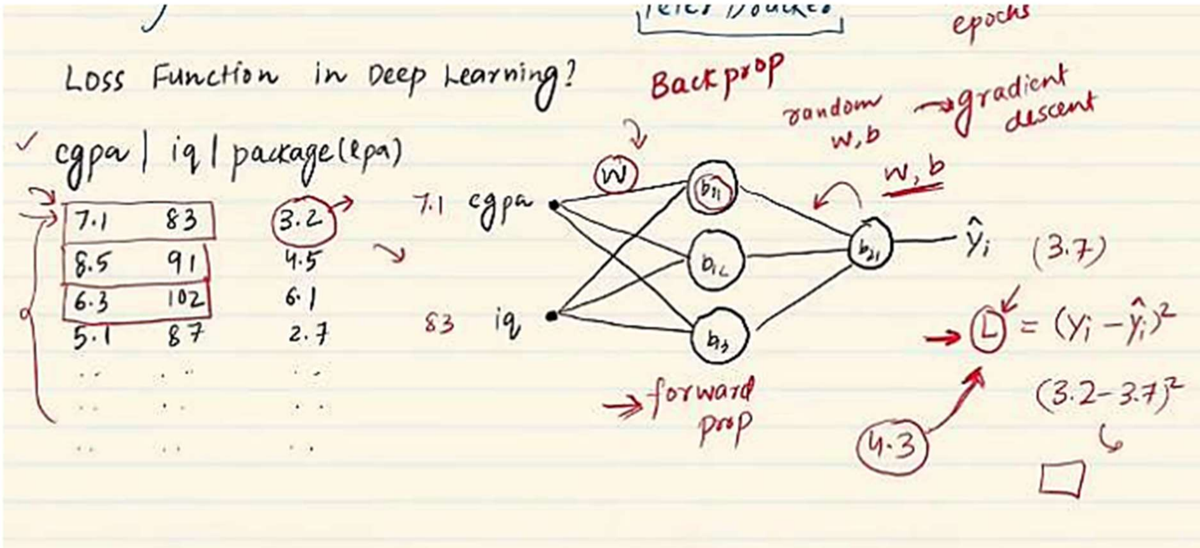


Error of First Row



Note: Like this error all rows generated at every iteration (not collectively – means it is loss function)

Error of all three rows



Key Points: Loss Function in Deep Learning

For regression problem there are six types of error:

- Mean square error – MSE
- Mean absolute error – MAE
- Root mean square error – RMSE
- Root Mean Square Logarithmic Error
- R2 error

For classification problem there are six types of error:

- For binary classification or multilabel classification – binary cross entropy
- For multiclass classification – categorical cross entropy
- For binary classification tasks, particularly in support vector machines (SVMs) – Hinge loss

For auto encoders: calculate Divergence loss

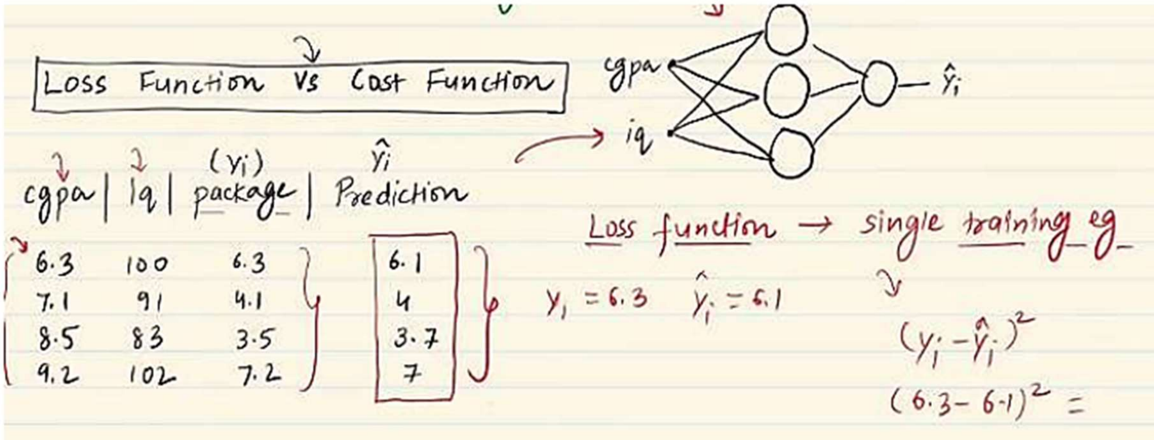
For embedding systems: calculate Triplet loss

For Object detection: calculate focal loss (e.g. yolo)

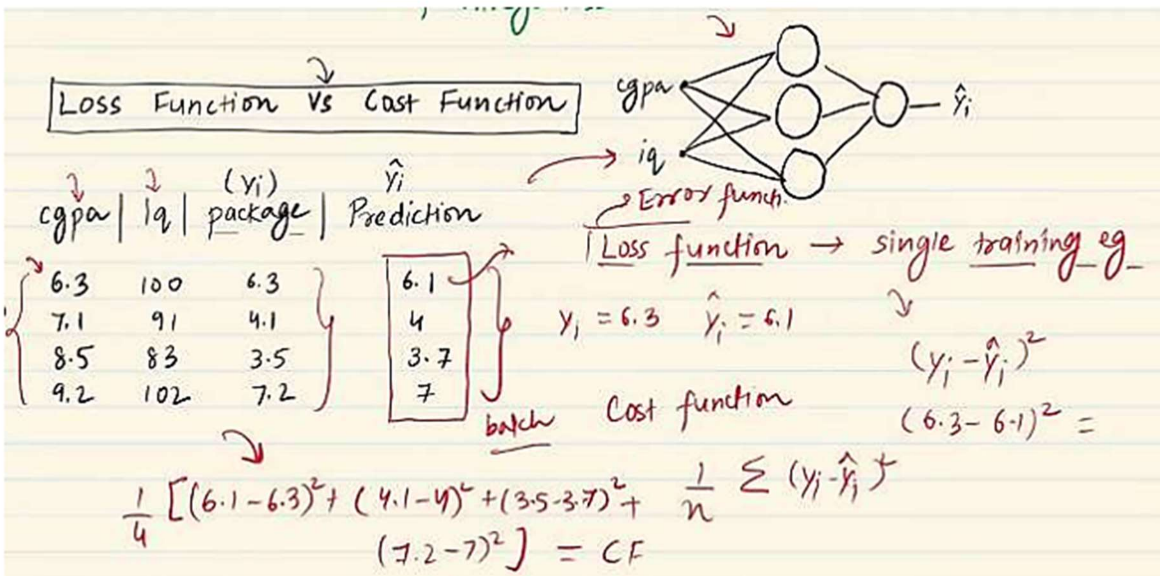
For Generative adversarial network (GAN) calculate:

- Discrimination loss
- Min-Max loss
- Gan loss

Loss Function VS Cost Function



Note: Above image shows Individual loss means loss function

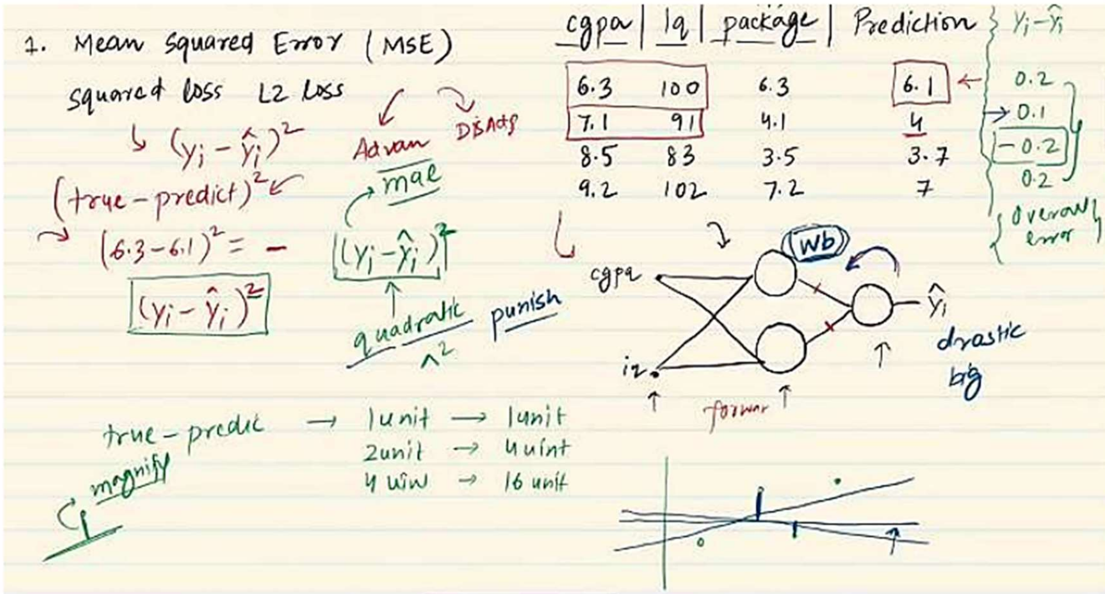


Note: Above image shows combined loss means cost function

Summary of the Difference:

- **Loss Function**: Measures error for a single data point. It's a function of the true label y and the predicted label $\hat{y}$.
  $L(y, \hat{y})$
- **Cost Function**: Aggregates the loss over the entire dataset (or a mini-batch). It's a function of the model parameters θ (or weights) and is often an average (or sum) of the loss function over all data points.
  $J(\theta) = \frac{1}{m}\sum_{i=1}^{m} L(y^{(i)}, \hat{y}^{(i)})$

Mean Squared Error (MSE) Squared Loss L2 Error



Advantages:

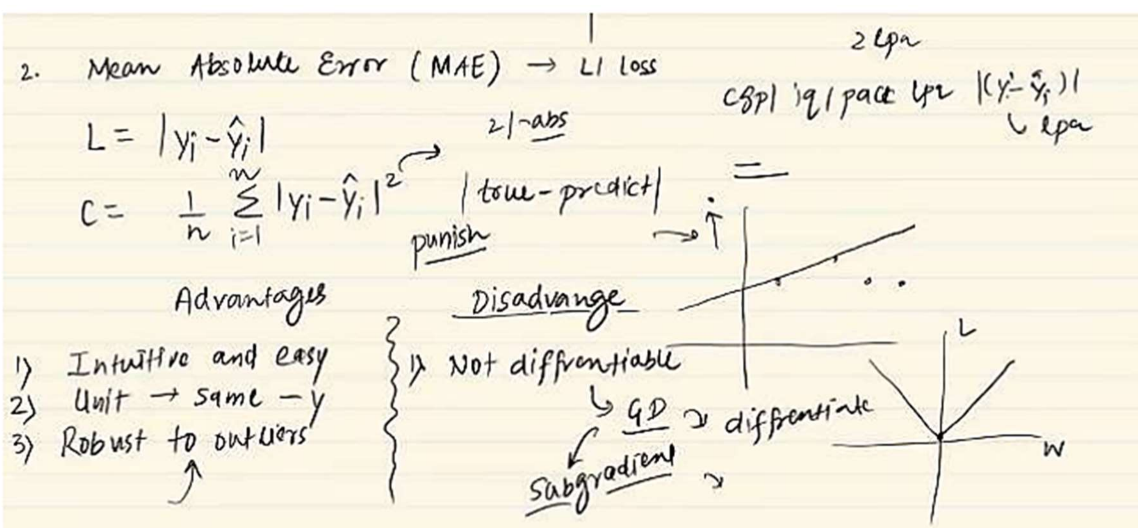
- Easy to Interpret
- Differentiable
- 1 local minima maxima

Disadvantages:

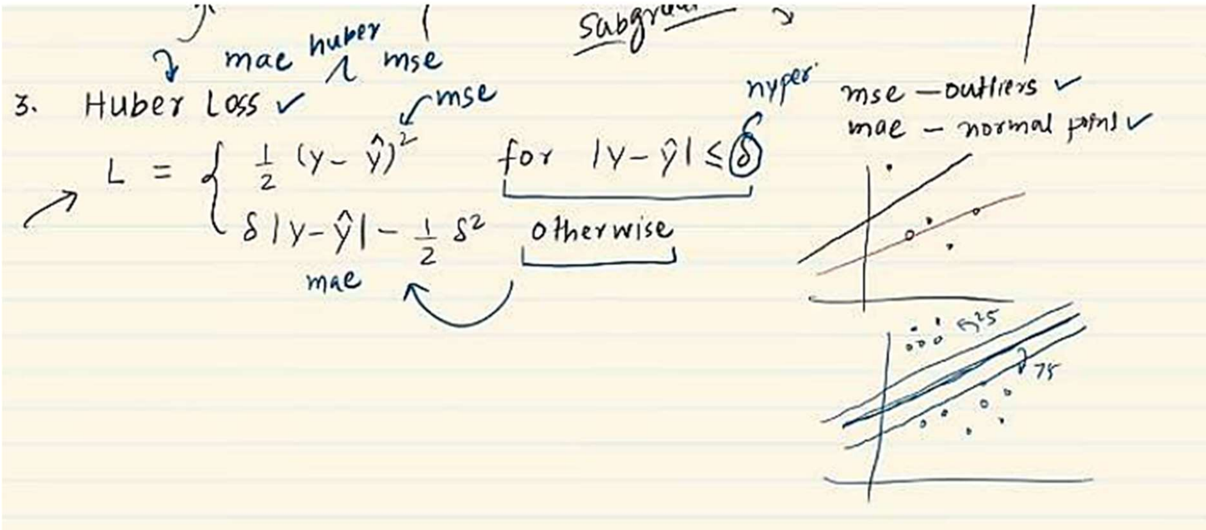
- Error unit Squared means – In L2 error it shows square of error
- Robust to Outliers (means it learn to outlier) – MSE is more sensitive to outliers, penalizing large deviations.

Overcome: Use Mean absolute error

Mean Absolute Error (MAE)



Huber Loss



Note: Hyperplane acts as a separator (in graph middle line is hyper plane). So in huber loss we set a limit that acts as a separator

E.g. greater than 50% use another function and for lesser than 50% use another function

---

Note: MSE behave in quadratic nature while MAE behave in linear nature

Quadratic

$$L_\delta = \begin{cases} \frac{1}{2}\left(y - f(x)\right)^2, & if\ |y - f(x)| \le \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2, & otherwise \end{cases}$$

Linear