

## NN loss

1/ when we have a distance based loss:

it won't give good training with multimodal distribution

example: MSE, RMSE, Binary Cross Entropy loss and Categorical Cross Entropy loss

⇒ an MLP with MSE won't be good at training if we have multimodal distribution.

### solution 1

we build a classifier to detect outliers

we build 2 training models:

1 - on normal data

1 - on outliers

### solution 2

we build a regression model for each mode

### solution 3

tree based models work better in these cases

## 1/ Mean absolute error

it checks the size of errors in a set of predicted values without caring about their positive or negative directions.

when is it used?

for regression problems when the distribution of the target variable has outliers.

## 2/ Mean squared error $\geq 0$

the squaring implies that larger mistakes have even larger errors than smaller ones

when is it used?

in major cases of regression problems.

## 3/ negative log likelihood (Boltzmann Softmax): NLL

the model is punished for making predictions with smaller probabilities and encouraged for making the prediction with higher probabilities.

⇒ it doesn't only care about ~~making~~ the prediction being correct but also about the model being certain about prediction

h/