LOSS FUNCTION

CHOOSING LOSS FUNCTION

Target "true" output ----------> y

Divergence

ŷ

LOSS (, ...)

Continous values

$$L_1(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

Mean Absolute Error

$$L_2(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Mean Squared Error

Sensitive to outliers

Classification

CAT 0.9

DOG 0.1

Probability distribution

KL Cross-Entropy Loss

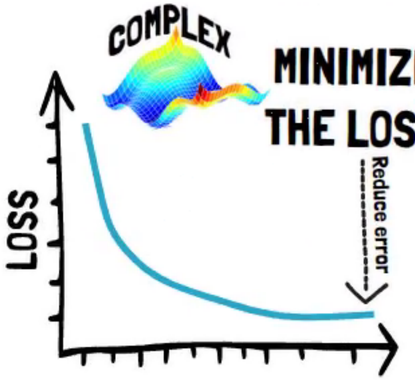
Difference between two probability distributions

1

0

True one-hot encoded labels

TRAINING



GRADIENT DESCENT

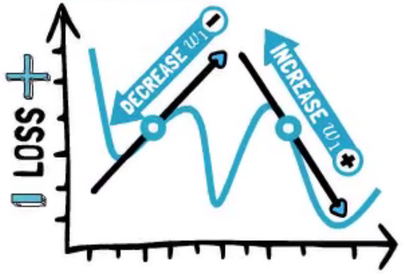
DERIVATIVE

Influence of weight on loss

$$\frac{\partial Loss}{\partial w_1}$$ = How much a minute increment affects the Loss?

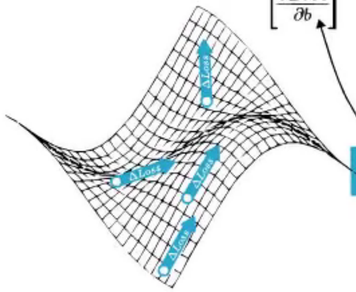
$w_1 + 0.001$

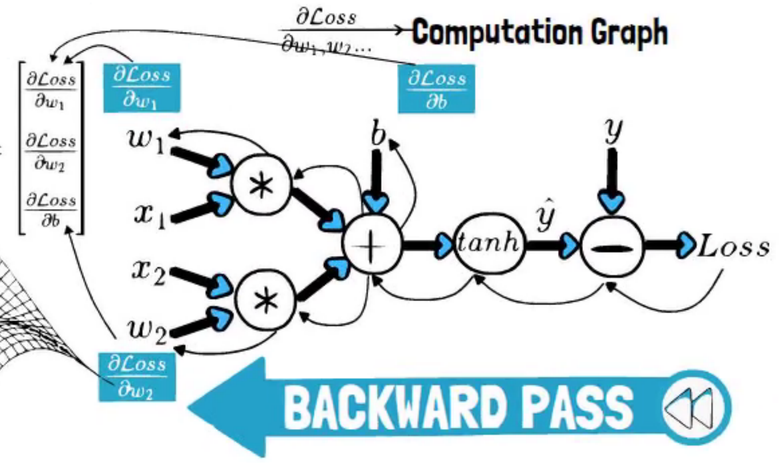
Update weights opposite to derivative



GRADIENT

$$\Delta Loss = \begin{bmatrix} \frac{\partial Loss}{\partial w_1} \\ \frac{\partial Loss}{\partial w_2} \\ \frac{\partial Loss}{\partial b} \end{bmatrix}$$





BACKWARD PASS