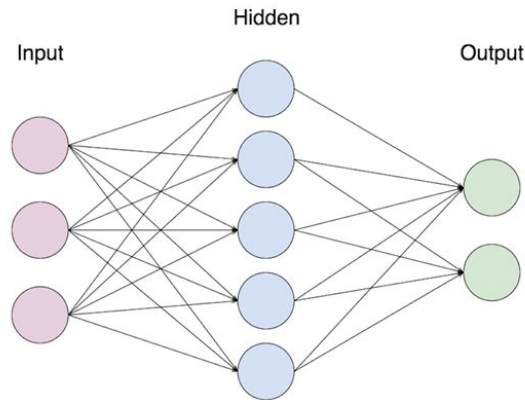


Neural Networks

Optimizers, Loss Functions and Learning Rate

Recap - Last Video

- Artificial Neural Networks
 - Structure
 - Elements in a Neural Network
- Layers in a Neural Network
 - Different types of layers
- Activation Functions in a Neural Network
 - How they affect a Neural Network
 - Different types of Activation Functions
 - When to use which activation function
- How to create a model in Python with Keras



Optimization, Loss and Learning Rate

- Loss and Loss Functions
 - What is it and what is it used for?
 - Different types of loss functions in Neural Networks
- Optimization
 - What is it and what is it used for?
 - Optimizers in Neural Networks
- Learning Rate
 - What is it?
 - How does it affect the Neural Network during training?
 - How to choose the right learning rate?

Loss and Loss Functions

- The objective during training of a Neural Network is to minimize the loss function by iteratively updating the weights in the network
- The loss is calculated by using the true labels from the input data and the predicted output from the Neural Network

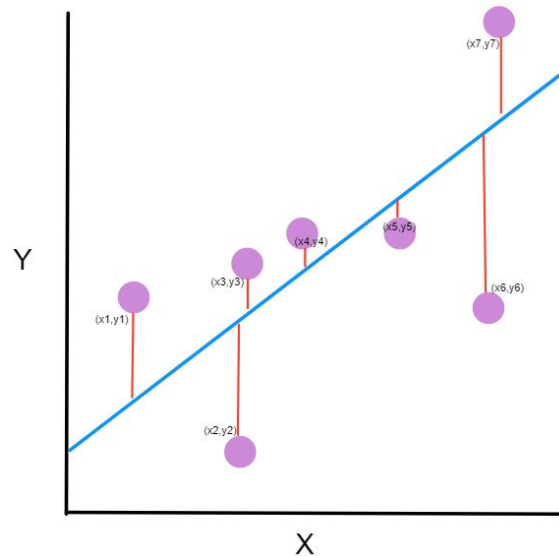
$$\text{error} = \text{model's prediction} - \text{true label}$$

- Mean Squared Error (MSE)

$$MSE = (\text{model's prediction} - \text{true label}) - (\text{model's prediction} - \text{true label})$$

Loss Functions

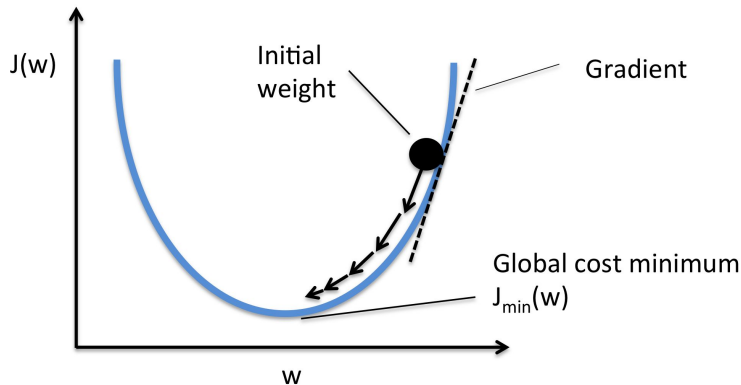
- Regression Loss Functions
 - Mean Squared Error
 - A lot more
- Probabilistic Loss Functions
 - Categorical Crossentropy
 - Sparse Categorical Crossentropy
 - Binary Crossentropy
 - A lot more



$$MSE = (model's\ prediction - true\ label)^2 + (model's\ prediction - true\ label)^2$$

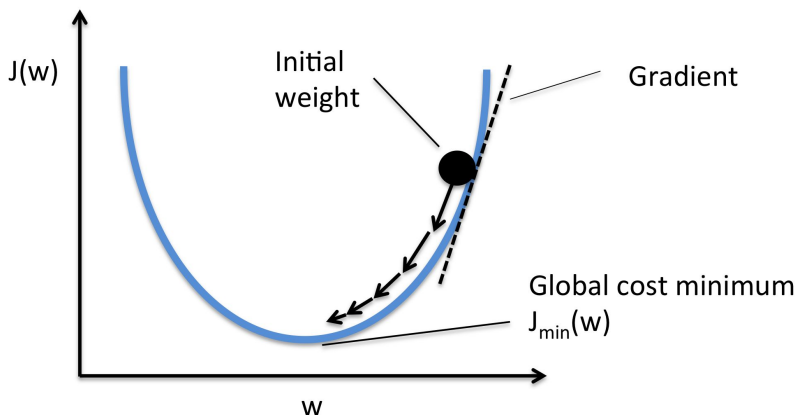
Optimizers in Neural Networks

- Specific implementation of the gradient descent algorithm
- Algorithm depends on the first order derivative of a loss function
- Calculates the way the weights should be altered so the loss function can reach a minima - a global minima is the goal and depends on the learning rate



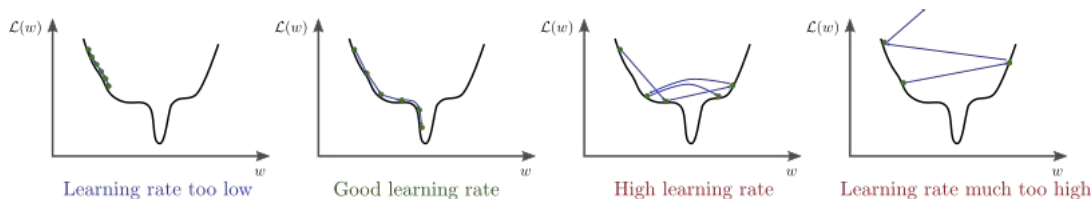
Optimizers in Neural Networks

- Stochastic Gradient Descent (SGD)
- Root Mean Square Prob (RMSprob)
- Adaptive Moment Estimation (Adam)
- Adaptive Gradient Algorithm (AdaGrad)



Learning Rate in Neural Networks

- Size of the steps we are taking to reach our minimized loss
- How fast do we want our model to learn?



*gradients * learning rate*

