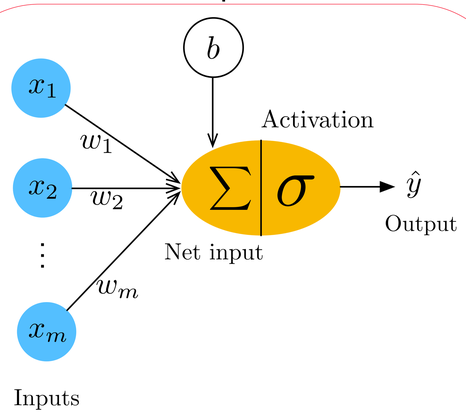
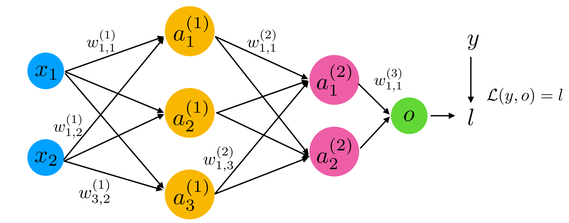
Perceptron



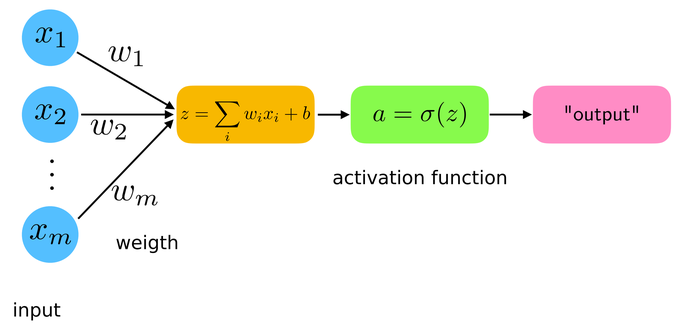
Fully-Connected Layers



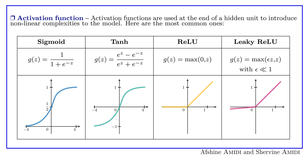
Neural network

Important terms

* backpropagation
* learning rate
* loss function
* gradient descent
* stochastic gradient descent
* parameter vs hyperparameter
* training/validation test splits
* updating weights
* activation function
* neural network architecture



activation functions



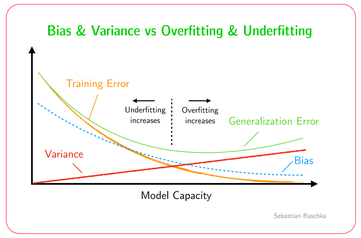
loss functions

method of evaluating how well your algorithm models the dataset.

* Mean Squared Error (MSE)
* cross-entropy loss
* Hinge loss....

updating weights

(a) Compute output (prediction)
(b) Calculate error
(c) Update W, b



"Tricks" for improving training

normalization

* Feature norm
* Batch norm
* InstanceNorm
* GroupNorm
* LayerNorm...

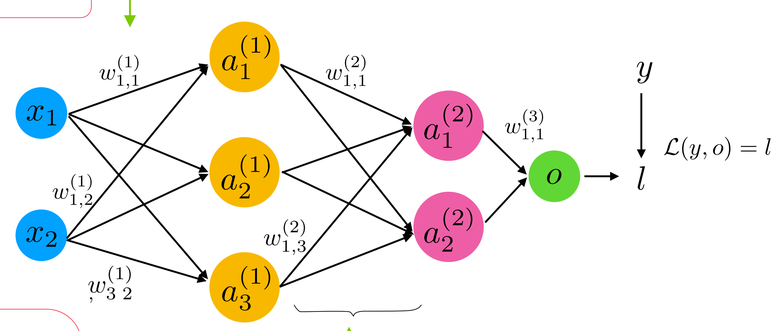
weight initialization

* Random
* Xavier
* He ...

Learning principles

* Learning rate
* Training epochs

* online-mode
* batch-mode
* mini-batch



Problems

* overfitting
* underfitting
* Exploding Gradients
* Vanishing gradient

underfitting: high bias low variance

overfitting: low bias high variance

Avoid overffiting

regularization

1. Avoiding overfitting with more data and data augmentation
2. Reducing network capacity & early stopping
3. Adding norm penalties to the loss: L1 &L2 regularization
4. Dropout