VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

Deep Learning

Neural Network Model Architecture

Jan Platoš, Radek Svoboda

March 24, 2024

Department of Computer Science
Faculty of Electrical Engineering and Computer Science
VŠB - Technical University of Ostrava

Neural Network Model Architecture

Neural Network Model Architecture

- Topology
  - Number of layers.
  - Number of Neurons in each layer.
  - Type of layers.
- Number of parameters to be optimized.
- Batch size.
- Regularization - L1, L2.
- Dropout.
- Batch Normalization.
- Underfitting / Overfitting

Batch Size

(Stochastic) Gradient Descent - is the process of getting internal parameters of the model (neural network) to fit the data using the calculation of an error gradient or slope of error and ”descent” refers to the moving down along that slope towards some minimum level of error.

Batch - is a set of samples/rows of the data that are processed before the internal state (weights) are updated.

Batch size - is the number of samples/rows in a batch. It heavily affects the efficiency of the algorithm and creates a variants: Batch Gradient Descent, Stochastic Gradient Descent, Minibatch Gradient Descent.

Regularization L1 and L2

Regularization - modifies the target cost function with penalties to improve the "shape" of the parameters.

L1 regularization also know as Lasso regularization (Least Absolute Shrinkage and Selection Operator). Adds an "absolute value of magnitude" as a penalty.

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

L2 regularization also know as Ridge regularization. Adds an "squared magnitude" as a penalty.

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Dropout

Dropout - is an approach to regularization in neural networks which helps reducing interdependent learning amongst the neurons.

Training Phase - for each hidden layer, for each training sample, for each iteration, ignore (zero out) a random fraction, p, of nodes (and corresponding activations).

Testing Phase - use all activations, but reduce them by a factor p (to account for the missing activations during training).

Batch Normalization

- The output of the each activation functions and its distribution shifts during the training - it is called Internal Covariance Shift.
- Natural solution is to normalize the outputs - it is called standardization (transform it to standard Gaussian distribution).

$$x' = \frac{x - \bar{x}}{\sqrt{\sigma^2 + \epsilon}}$$

- **Batch Normalization** is the normalization of the layer output for each batch.

Overfitting/Underfitting

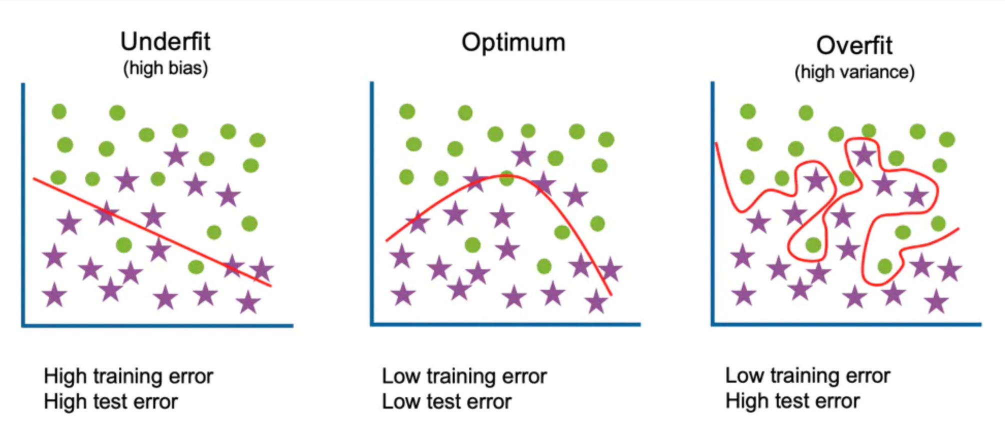
The bias error is an error from erroneous assumptions in the learning algorithm. High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).

The variance is an error from sensitivity to small fluctuations in the training set. High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs (overfitting).

Underfitting - is inability of the model to learn the properties of a training dataset.

Overfitting - occurs when a model is too specialized to traing dataset and is unable to generalize to the unseen data.

Overfitting/Underfitting



Overfitting/Underfitting - Underfitting avoidance

- Decrease Regularization
  - Regularization usually decrease variance.
  - Reducing regularization allow complexity and variance in the model.
- Extend training time
  - Training lead to acquiring knowledge from data.
  - Early stopping does not allow enough learning.
- Feature Selection
  - Use more complex features.
  - Increase a complexity of a model (more neurons, layers, etc.).

Questions?