INFO 7375 - Neural Networks & AI

Homework to Chapter - 10

Submitted By:

Abdul Haseeb Khan
NUID: 002844724
khan.abdulh@northeastern.edu

What are underfitting and overfitting?

Overfitting is when high variance makes a model sensitive to small fluctuations in the training set and results in modeling random noise in the training data, which is explicitly called overfitting.

Underfitting is when high bias from wrong assumptions in the learning algorithm causes the model to miss the relevant relations between features and target outputs, which is explicitly called underfitting.

What may cause an early stopping of the gradient descent optimization process?

An early stopping may be triggered by an unexpected loss function reversal, where the loss trend turns in the opposite direction.

Concretely, training halts when the training error is no longer decreasing while the validation error starts to rise, which signals the onset of overfitting.

Parameters are saved periodically and, after stopping, are restored to the point just before the validation error began to increase.

Describe the recognition bias vs variance and their relationship.

Bias occurs when an algorithm is used but does not fit properly, causing a shift in the classification results, while variance indicates the amount of unexpected variation in the estimation when different training data is used and reflects the difference between the actual values and the predicted values. Their relationship is captured by the bias variance tradeoff, which states that variance can be reduced by increasing bias, with high variance tied to overfitting and high bias tied to underfitting in pursuit of better generalization.

Bias occurs in a machine learning model when an algorithm is used but does not fit properly, and by changing bias, the classification can be shifted to better fit the target values.

Variance indicates the amount of unexpected variation in the estimation when different training data is used, indicating the difference between the actual values and the predicted values

The variance is an error from sensitivity to small fluctuations in the training set, and high variance may result in modeling the random noise in the training data, which is called overfitting. The bias error is an error from wrong assumptions in the learning algorithm, and high bias can cause an algorithm to miss the relevant relations between features and target outputs, which is called underfitting. The bias variance tradeoff describes that variance can be reduced by increasing bias, and effective regularization strives to minimize both sources of error to achieve better generalization. Bias variance tradeoff and overfitting underfitting are equivalent.

In classification terms, high bias means the separation tends to be unevenly closer to one category, just right means the separation adequately classifies objects, and high variance means some correctly classified objects are far from others.

For high bias, the suggested solutions are a bigger neural network, training longer, and changing the ANN architecture. As a basic recipe, high bias is framed as a training data problem addressed by training longer and architecture search, while high variance is framed as a data set problem addressed by regularization and architecture search.

Describe regularization as a method and the reasons for it.

Regularization is a set of strategies and techniques that prevent overfitting in neural networks and thus improve the accuracy of a Deep Learning model on completely new data from the problem domain. It works by adding an extra punishing term to the loss function, which constrains the model toward simpler solutions that generalize better.

In the bias-variance framework, regularization trades increased bias for reduced variance, with an effective regularizer making a profitable trade that reduces variance significantly without overly increasing bias, which directly supports better generalization. In simple terms, this results in simpler models and follows the Occam's razor principle that the simplest models are most likely to perform better, achieved by constraining the model to a smaller set of possible solutions.

The core reason to use regularization is to address high variance by reducing overfitting while keeping the training error as low as possible, which is part of the basic recipe when a model exhibits a data set problem indicative of variance. Summarizing the motivation, it prevents overfitting, reduces variance through a controlled increase of bias, and yields simpler models that are more likely to perform better on unseen data.

Describe dropout as a method and the reasons for it.

Dropout randomly ignores some layer outputs during training with probability p, and during test time all units are present but activations are scaled down by p to match the expected magnitude seen during training. It is used to prevent overfitting by injecting noise into hidden units, breaking co-adaptation, encouraging sparse representations, and making the model more robust.

Dropout falls into noise injection techniques and can be seen as noise injected into the hidden units of the network, where during training some number of layer outputs are randomly ignored with probability p. During test time all units are present, but they are scaled down by p because after dropout the next layers receive lower values during training and keeping all units at test would otherwise produce values a lot higher than expected. By using dropout the same layer alters its connectivity and searches for alternative paths to convey information so that each update is performed with a different view of the configured layer, conceptually approximating training a large number of neural networks with different architectures in parallel, and the probability p depends on the architecture.

Both regularization and dropout are widely adopted methods to prevent overfitting, with dropout randomly muting some neurons in the forward process in order to make the network more concise. Dropout has the effect of making the training process noisy, which breaks up situations where network layers co-adapt to correct mistakes from prior layers and makes the model more robust, and as a form of noise injection adding randomness reduces variance and lowers generalization error. It increases the sparsity of the network and in general encourages sparse representations, which serves as a powerful tool in the regularization arsenal.