SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of AI &DS

Course Name – 19AD602 DEEP LEARNING

III Year / VI Semester

Unit 3-DIMENSIONALITY REDUCTION
Topic: Training a Convnet: hyperparameter optimization

Case Study

A company building a machine learning model for customer churn prediction struggled to achieve satisfactory accuracy. They used hyperparameter optimization techniques (e.g., grid search, random search, or Bayesian optimization) to tune parameters like learning rate, batch size, and number of hidden layers in their neural network. After optimization, their model's accuracy improved by 15%, enabling better-targeted retention strategies.

Activity

1. Choose a dataset (e.g., Titanic or MNIST) and a machine learning algorithm (e.g., Random Forest or Neural Network).
2. Split the dataset into training and testing sets. Train a baseline model.
3. Use hyperparameter optimization (grid search or random search) to find the best values for key hyperparameters, such as learning rate or tree depth.
4. Compare the performance metrics of the baseline and optimized models.

Training a Convnet: hyperparameter optimization

Hyperparameter Tuning - Grid vs Random

Performance

Problem → RF ⇒ no. of trees ? best X
DT
NN
↳ Learning rate

Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning. The prefix 'hyper_' suggests that they are 'top-level' parameters that control the learning process and the model parameters that result from it.

Training a Convnet: hyperparameter optimization

	Hyperparameter	Meaning	Sample Value	
1	Learning rate	Determines the size of the step that the optimization algorithm takes in each iteration of training	0.01	0 to 1
2	Number of hidden layers	Determines the number of layers between the input and output layers of the neural network	3, 1, 2, 4	
3	Dropout rate	Determines the percentage of neurons in the network that are randomly set to zero during training, to prevent overfitting	0.2	
4	Number of iterations	Determines the number of times the training data is passed through the model during training	1000	
5	Batch size	Determines the number of training examples used in each iteration of training	32 ✓	
6	Activation function	Determines the function used to transform the output of a neuron into a non-linear range	ReLU	
7	Regularization strength	Determines the strength of the penalty applied to the weights of the model to prevent overfitting	0.01 ✓	
8	Kernel size	Determines the size of the convolutional kernel used in a convolutional neural network	3×3 ✓	
9	Number of neurons per layer	Determines the number of neurons in each hidden layer of the neural network	128	
10	Momentum	Determines the fraction of the previous gradient direction to add to the current gradient direction during training, to accelerate convergence	0.9	

1 Grid Search

Grid Search starts with defining a search space **grid**. The grid consists of selected hyperparameter names and values, and **grid search** exhaustively searches the best combination of these given values.

1. Learning Rate (2) - [0.1,0.2]
2. Number of Hidden Layers (2) - [3,5]
3. Batch Size (2) - [32,64]

$2 \times 2 \times 2 = 8$

Time

0.1, 3 — 0.1, 3, 32; 0.1, 3, 64

0.1, 5 — 0.1, 5, 32; 0.1, 5, 64

(4)

0.2, 3, 32

0.2, 3, 64

0.2, 5, 32

0.2, 5, 64

(4)

performance

T1 – 40%

T3 – 60%

T8 – 95%

Here's how grid search works:

1. Define a set of hyperparameters and their possible values.
2. Specify a grid of hyperparameter values to search over.
3. Train and evaluate the model for each combination of hyperparameters in the grid.
4. Choose the hyperparameters that give the best performance on the validation set.

Advantages of Grid Search:

- Exhaustive: Grid search is an exhaustive search over the specified hyperparameter values, which guarantees that the optimal combination of hyperparameters is found within the search space.
- Systematic: Grid search is a systematic method that ensures that all possible combinations of hyperparameters are evaluated, which helps to avoid biases and oversights.
- Reproducible: Grid search is a reproducible method that yields the same results every time it is run, as long as the random seed is set.

Drawbacks of Grid Search:

- Computationally expensive: Grid search can be computationally expensive, as it requires training and evaluating the model for each combination of hyperparameters in the grid. This can be particularly time-consuming for large search spaces or complex models.
- Search space limitations: Grid search is limited by the size of the search space and the granularity of the hyperparameters. If the search space is too large or the hyperparameters are too granular, it may be difficult to find the optimal combination of hyperparameters within a reasonable time frame.
- Interactions between hyperparameters: Grid search assumes that hyperparameters are independent, but in reality, there may be interactions between hyperparameters that affect the model's performance. Grid search may not be able to capture these interactions.

Random Search

In random search, we define **distributions** for each hyperparameter which can be defined *uniformly* or with a *sampling method*. The key difference from grid search is in random search, not all the values are tested and values tested are selected at random.

1. Learning rate - range(0.5,1) → [0.5, 0.6, 0.7, 0.8, 0.9, 1]
2. Batch Size - list(32,64,128,256)

1 ⇒ 0.9, 128   2 ⇒ 0.5, 32

10

Training a Convnet: hyperparameter optimization

Advantages of Random Search:

- More efficient: Random search is often more efficient than grid search, as it can find good hyperparameters faster by randomly sampling from a distribution over the search space. This can be particularly useful for large search spaces or complex models.
- Flexible: Random search is a flexible method that can handle different types of hyperparameters and distributions, including continuous, discrete, and categorical hyperparameters.
- Better at handling interactions: Random search can handle interactions between hyperparameters better than grid search, as it can randomly sample different combinations of hyperparameters and capture their interactions.

Drawbacks of Random Search:

- Stochastic: Random search is a stochastic method that yields different results each time it is run, due to the random sampling of hyperparameters. This can make it difficult to reproduce results or compare different runs.
- Sample inefficiency: Random search may be sample inefficient if the search space is small or the hyperparameters are not well-defined, as it may sample many combinations of hyperparameters that perform poorly on the validation set.
- Limited by search space: Random search is limited by the size and quality of the search space, and may not be able to find the optimal combination of hyperparameters if the search space is too small or poorly defined.

Thank you