PARAMETERS VS HYPER PARAMETER

- **Parameters :** Parameters that are estimated by the model from the given data
  Ex: weights
- **Hyper parameters :** Parameters that cannot be estimated by the model from the given data
  Ex: learning rate

How

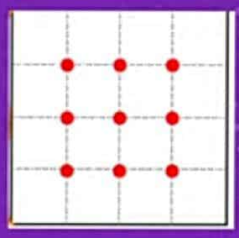
- ✓ **Manual hyper parameter tuning**
- ✓ **Automated hyper parameter tuning**

HYPER PARAMETER TUNING METHOD

1. **Random Search:** Each iteration tries a random combination of hyperparameters from this grid, records the performance, and lastly returns the combination of hyperparameters which provided the best performance

2. **Grid Search:**

- Each iteration tries a combination of hyperparameters in a specific order.
- It fits the model on each and every combination of hyperparameter possible and records the model performance.
- Finally, it returns the best model with the best hyperparameters



HYPER PARAMETER TUNING METHOD

3. **Bayesian Optimization:**

- It helps us find the minimal point in the minimum number of steps.
- Bayesian optimization also uses an acquisition function that directs sampling to areas where an improvement over the current best observation is likely

4. **Tree-structured Parzen estimators:**

- Similar to Bayesian optimization. Instead of finding the values of p(y|x) where y is the function to be minimized (e.g., validation loss) and x is the value of hyperparameter the TPE models P(x|y) and P(y).
- Instead of finding the values of p(y|x) where y is the function to be minimized (e.g., validation loss) and x is the value of hyperparameter the TPE models P(x|y) and P(y)

HYPER PARAMETER TUNING ALGORITHM

1. **Hyperband**

- Hyperband is a variation of random search, but with some explore-exploit theory to find the best time allocation for each of the configurations

2. **Population-based training**

- It is a hybrid of two most commonly used search techniques, Random Search and manual tuning applied to Neural Network models
- PBT starts by training many neural networks in parallel with random hyperparameters. But these networks aren't fully independent of each other

3. **Bayesian Optimization and HyperBand**

It mixes the Hyperband algorithm and Bayesian optimization

HYPER PARAMETER OPTIMIZATION TOOLS

1. **Scikit-learn**

- Grid search & Random search

2. **Hyperopt**

- It allows the user to describe a search space in which the user expects the best results allowing the algorithms in hyperopt to search more efficiently
- Random Search
- Tree of Parzen Estimators (TPE)
- Adaptive TPE

3. **Scikit-optimize**
- To find optimal solutions for hyperparameter search problems in less time

HYPER PARAMETER OPTIMIZATION TOOLS

4. Optuna

- It has the pruning feature which automatically stops the unpromising trails in the early stages of training
- Lightweight, versatile, and platform-agnostic architecture
- Pythonic search spaces
- Easy parallelization
- Quick visualization

5. Ray Tune

- Tune is a popular choice of experimentation and hyperparameter tuning at any scale
- Provided SOTA algorithms such as ASHA, BOHB, and Population-Based Training.
- Supports Tensorboard and MLflow.
- Supports a variety of frameworks such sklearn, xgboost, Tensorflow, pytorch, etc.

HYPER PARAMETER OPTIMIZATION TOOLS

- **Keras Tuner**

- The Keras Tuner is a library that helps you pick the optimal set of hyperparameters for your TensorFlow program

- You can define a hypermodel through two approaches:

- By using a model builder function
- By subclassing the HyperModel class of the Keras Tuner API

Deep Learning Terminology - 1

Dense layer

- Dense layer is the regular deeply connected neural network layer. It is most common and frequently used layer.
- Dense layer does the below operation on the input and return the output.

output = activation(dot(input, kernel) + bias)

Let us consider sample input and weights as below and try to find the result –

- ✓ Input as 2 x 2 matrix [[1, 2], [3, 4]]
- ✓ Kernel as 2 x 2 matrix [[0.5, 0.75], [0.25, 0.5]]
- ✓ Bias value as 0
- ✓ Activation.

Deep Learning Terminology - 2

Depth

- The number of layers (including any embedding layers) in a neural network that learn weights.
- For example, a neural network with 5 hidden layers and 1 output layer has a depth of 6.

Deep Learning Terminology - 3

Early stopping

- A method for regularization that involves ending model training before training loss finishes decreasing.
- In early stopping, you end model training when the loss on a validation dataset starts to increase