|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hyperparameter | Stochastic expression | Choices | Lower bound | Upper bound |
| Learning rate | Uniform |  | 1e-5 | 1e-3 |
| Drop rate | Uniform |  | 0 | 0.5 |
| Regularization | Log uniform |  | Log(1e-5) | Log(1) |
| Batch size | Choice | 16, 32, 64, 128, 256 |  |  |
| Hidden layers architecture | Choice | (128, 128), (256, 256),  (512, 512), (128, 128, 128), (256, 256, 256), (512, 512, 512) |  |  |
|  | Uniform |  | 0.8 | 0.99 |
| Learning rate decay | Log uniform |  | Log(1e-5) | Log(1) |
| Addition layer (res) | Choice | True, False |  |  |
| Bypass layer | Choice | True, False |  |  |

For hidden layers architecture, the number of entries in a tuple is the number of hidden layers. The number in each entry of the tuple is the number of nodes in the hidden layer.