Optimization results

Random Forest - Group 17

Iris

Metaparameters:

- Number of trees: 84
- Max depth of trees: 20
- Minimum size of each split: 5
- Ratio of samples: 0.8
- Ratio of rows for training: 0.7

Best purity measure for performance: **Gini**
Worst time: 0.656s on sequential random forest using entropy
Best time: 0.089s on sequential extra trees using Gini
Performance Improvement:

$$\text{Extra trees} \Rightarrow 7.3 \text{ times faster}$$

Purity Measure: *Gini*	**Computing method**	**Total training time**	**Accuracy**
Random Forest	Sequential	0.614s	93.3%
	Parallel	0.189s	95.6%
Extra Trees	Sequential	0.089s	93.3%
	Parallel	0.094s	93.3%

Purity Measure: *Entropy*	**Computing method**	**Total training time**	**Accuracy**
Random Forest	Sequential	0.656s	95.6%
	Parallel	0.201s	93.3%
Extra Trees	Sequential	0.092s	95.0%
	Parallel	0.091s	95.6%

Sonar

Metaparameters:

- Number of trees: 84
- Max depth of trees: 20
- Minimum size of each split: 5
- Ratio of samples: 0.8
- Ratio of rows for training: 0.7

Best purity measure for performance: **Entropy**
Worst time: 11.893s on sequential random forest using Gini
Best time: 0.168s on parallel extra trees using entropy
Performance Improvement:

$$\text{Extra trees} \ + \ \text{Parallel processing} \Rightarrow 70.79 \text{ times faster}$$

Purity Measure: *Gini*	**Computing method**	**Total training time**	**Accuracy**
Random Forest	Sequential	11.893s	77.8%
	Parallel	2.006s	74.6%
Extra Trees	Sequential	0.435s	81.0%
	Parallel	0.177s	73.0%

Purity Measure: *Entropy*	**Computing method**	**Total training time**	**Accuracy**
Random Forest	Sequential	10.269s	79.4%
	Parallel	1.999s	76.2%
Extra Trees	Sequential	0.391s	76.2%
	Parallel	0.168s	79.4%

MNIST

Metaparameters:

- Number of trees: 42
- Max depth of trees: 20
- Minimum size of each split: 20
- Ratio of samples: 0.4
- Ratio of rows for training: 0.7

Best purity measure for performance: **Entropy**
Worst estimated time: 27233s or 7.6h on sequential random forest using Gini
Best time: 240.652s on parallel extra trees using Entropy
Performance Improvement:

$$\text{Extra trees} \ + \ \text{Parallel processing} \Rightarrow 113 \text{ times faster}$$

Purity Measure: *Gini*	**Computing method**	**Total training time**	**Accuracy**
Random Forest	Sequential	27233[1]s	>73.4[2]%
	Parallel	4538[1]s	NA%
Extra Trees	Sequential	479.015s	92.8%
	Parallel	253.498s	92.8%

Purity Measure: *Entropy*	**Computing method**	**Total training time**	**Accuracy**
Random Forest	Sequential	21040[1]s	>73.3[2]%
	Parallel	3506[1]s	NA%
Extra Trees	Sequential	439.417s	92.5%
	Parallel	240.562s	92.5%

[1]Time estimated by measuring the time taken to build a single tree and assuming that the performance improvement of using parallel computing is of aproximately x6 as seen in the datasets of Iris and Sonar.

[2]Accuraccy of a single tree.