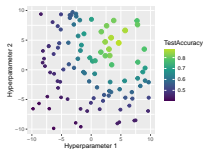


Introduction to Machine Learning

Hyperparameter Tuning Basic Techniques



Learning goals

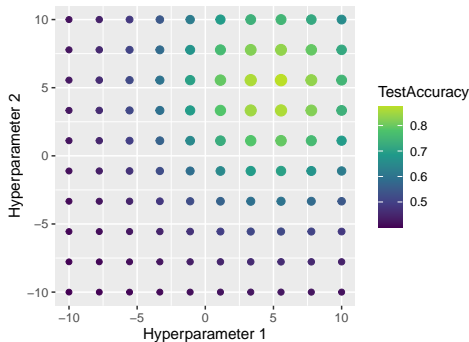
- Understand the idea of grid search
- Understand the idea of random search
- Be able to discuss advantages and disadvantages of the two methods

GRID SEARCH

- Simple technique which is still quite popular, tries all HP combinations on a multi-dimensional discretized grid
- For each hyperparameter a finite set of candidates is predefined
- Then, we simply search all possible combinations in arbitrary order



Grid search over 10x10 points



GRID SEARCH / 2

Advantages

- Very easy to implement
- All parameter types possible
- Parallelizing computation is trivial

Disadvantages

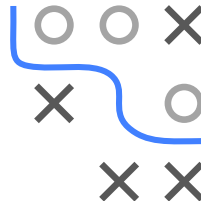
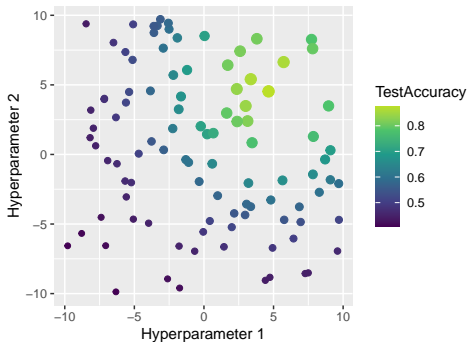
- Scales badly: combinatorial explosion
- Inefficient: searches large irrelevant areas
- Arbitrary: which values / discretization?



RANDOM SEARCH

- Small variation of grid search
- Uniformly sample from the region-of-interest

Random search over 100 points



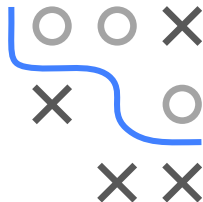
RANDOM SEARCH / 2

Advantages

- Like grid search: very easy to implement, all parameter types possible, trivial parallelization
- Anytime algorithm: can stop the search whenever our budget for computation is exhausted, or continue until we reach our performance goal.
- No discretization: each individual parameter is tried with a different value every time

Disadvantages

- Inefficient: many evaluations in areas with low likelihood for improvement
- Scales badly: high-dimensional hyperparameter spaces need *lots* of samples to cover.



TUNING EXAMPLE

Tuning random forest with grid search/random search and 5CV on the sonar data set for AUC:

Hyperparameter	Type	Min	Max
num.trees	integer	3	500
mtry	integer	5	50
min.node.size	integer	10	100

