Patterns, Volume *3*

Supplemental information

Don't lose samples to estimation

Ioannis Tsamardinos

Experimentation setting

- Simulate hold-out test sets of 50 samples
- All Learners have true accuracy 85%
- They make independent predictions and errors

One Pipeline: No problem

- Estimate performance of a single learner using a hold-out test set of 50 samples; simulated trial #1

Algorithm	Hyper-Parameter	Test Accuracy
K-NN	K=1	0.9

Assume: **true accuracy 85%**

One Pipeline : No problem

- Estimate performance of a single algorithm using a hold-out test set of 50 samples; simulated trial #2

Algorithm	Hyper-Parameter	Test Accuracy
K-NN	K=1	0.7

Assume: **true accuracy 85%**

One Pipeline : No problem

- Estimate performance of a single algorithm using a hold-out test set of 50 samples ; simulated trial #3

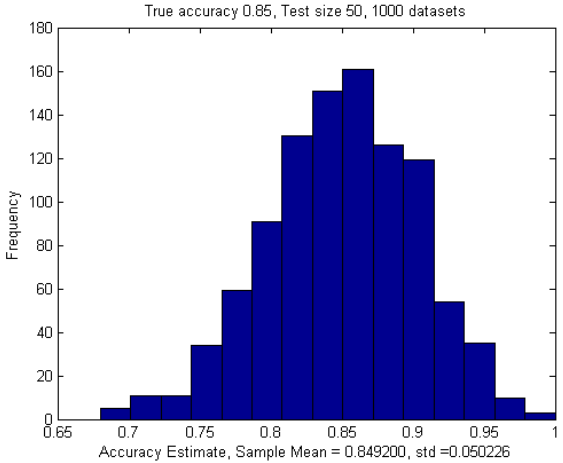
Algorithm	Hyper-Parameter	Test Accuracy
K-NN	K=1	0.87

Assume: **true accuracy 85%**

One Pipeline : No problem

- Estimate performance of a single algorithm using a hold-out test set of 50 samples; 1000 simulated trials

Algorithm	Hyper-Parameter	Test Accuracy
K-NN	K=1	0.842



Assume: **true accuracy 85%**

Mean estimate = 0.842

On average, using held-out test data (or CV) will estimate the correct true accuracy when test data are seen by a single pipeline and corresponding model

Many Configurations: Overestimation!

- Estimate performance of 8 learners using a hold-out test set of 50 samples ; return the estimate of the best one; simulated trial #1

Algorithm	Hyper-Parameter	Test Accuracy
K-NN	K=1	0.87
	K=2	0.90
	K=5	0.86
DT	MaxPChance=0.01	0.7
	MaxPChance=0.05	0.8
	MaxPChance=0.1	**0.9**
SB	l = 0	0.60
	l=1	0.90

Assume: **true accuracies 85% (all learners); independent predictions by each learner**

Many Configurations: Overestimation!

- Estimate performance of 8 learners using a hold-out test set of 50 samples ; return the estimate of the best one; simulated trial #2

Algorithm	Hyper-Parameter	Test Accuracy
K-NN	K=1	**0.95**
	K=2	0.90
	K=5	0.80
DT	MaxPChance=0.01	0.76
	MaxPChance=0.05	0.78
	MaxPChance=0.1	0.78
SB	l = 0	0.80
	l=1	0.88

Assume: **true accuracies 85% and independent predictions by each learner**

Many Configurations: Overestimation!

- Estimate performance of 8 learners using a hold-out test set of 50 samples ; return the estimate of the best one; simulated trial #3

Algorithm	Hyper-Parameter	Test Accuracy
K-NN	K=1	0.68
	K=2	0.75
	K=5	0.79
DT	MaxPChance=0.01	**0.92**
	MaxPChance=0.05	0.85
	MaxPChance=0.1	0.88
SB	l = 0	0.90
	l=1	0.88

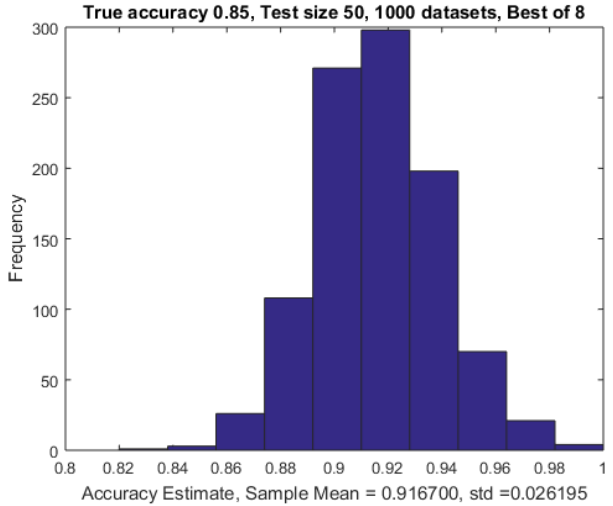
Assume: **true accuracies 85% and independent predictions by each learner**

JADBio®

Many Configurations: Overestimation!

- Estimate performance of a single algorithm using a hold-out test set of 50 samples; 1000 simulated trials

Algorithm	Hyper-Parameter	Estimate
K-NN	K=1	0.916
	K=2	
	K=5	
DT	MaxPChance=0.01	
	MaxPChance=0.05	
	MaxPChance=0.1	
SB	l = 0	
	l=1	



On average, the accuracy of the winning learner is over-estimated, when selected among many tries; see [Tsamardinos et. al. Machine Learning Journal 2018 for a theoretical proof].

Assume: **true accuracy 85%**
Mean Estimate: **0.916**