Reinforcement learning algorithm

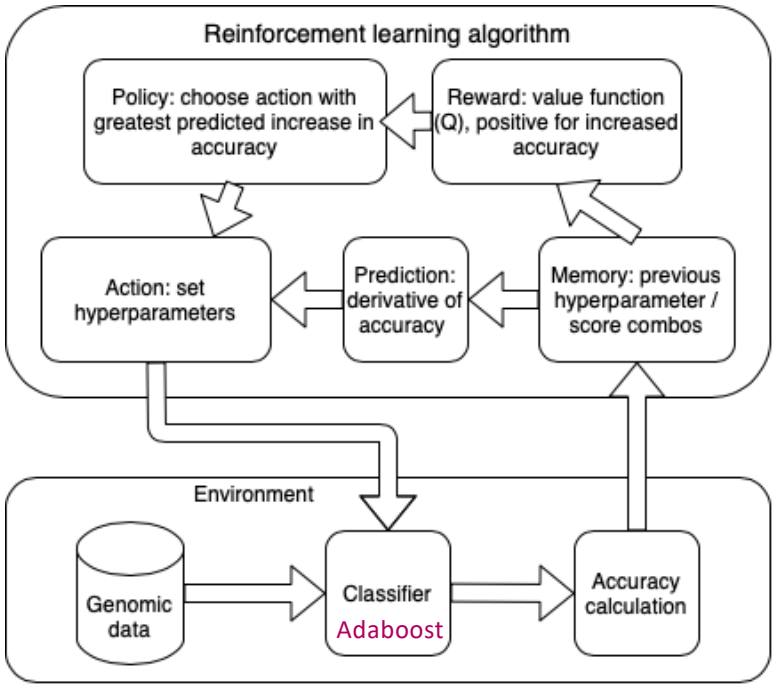
Brian Karlberg

Machine Learning 643

Dr. Xubo Song

Spring 2021

Schematic


Reinforcement learning algorithm
Policy: choose action with greatest predicted increase in accuracy
Reward: value function (Q), positive for increased accuracy
Action: set hyperparameters
Prediction: derivative of accuracy
Memory: previous hyperparameter / score combos
Environment
Genomic data
Classifier
Adaboost
Accuracy calculation

Process

- Generate Q-learning pseudocode
  - Use prediction accuracy to estimate value of next actions
- Measure learning environment wall times and variance
  - Accurate characterization of state quality
- Set reinforcement learning algorithm coefficients
  - Discount factor =0.2
  - Learning rate

```
# Jump to random n_est, lr coords
```

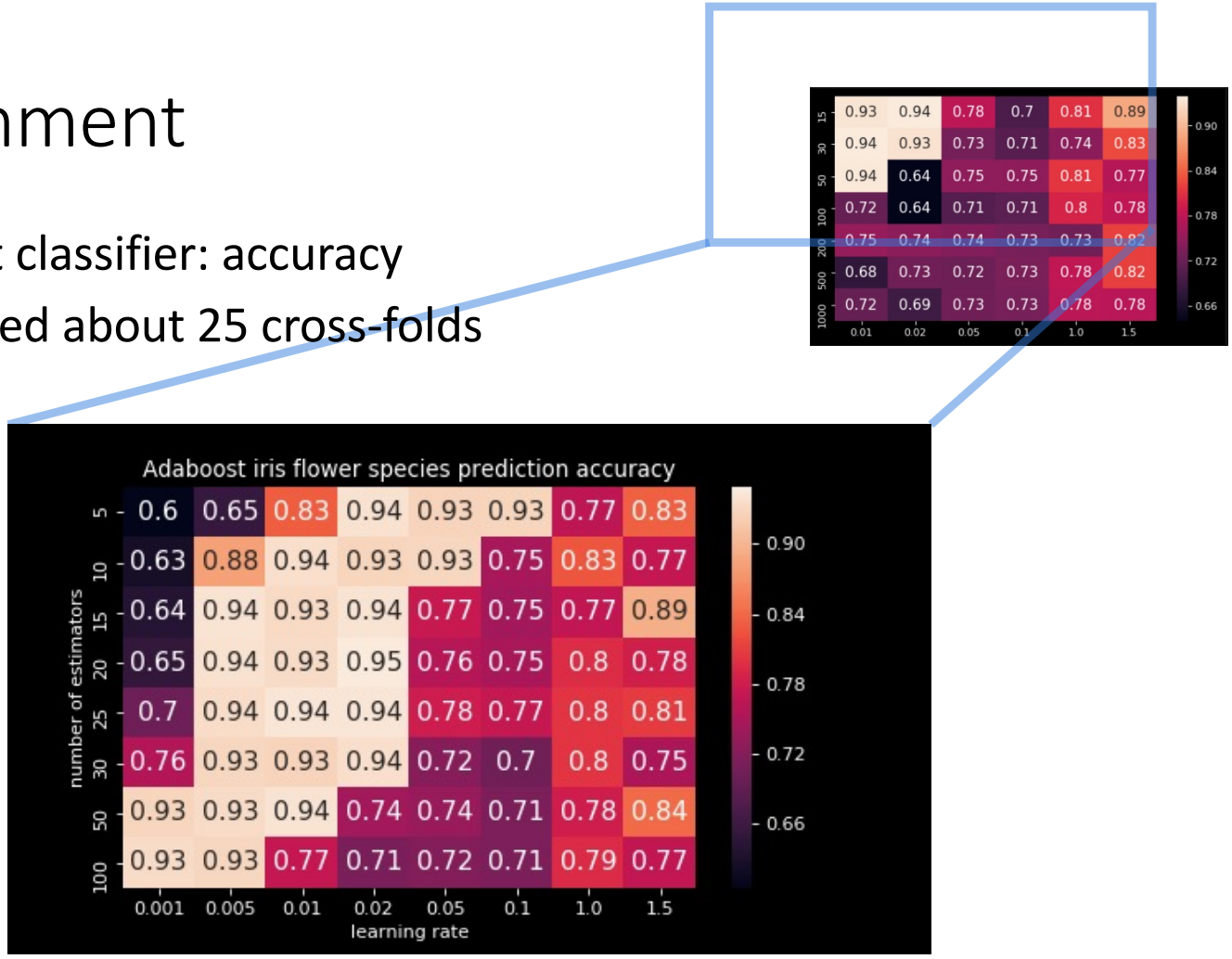
```
est_step = 5
LR_step = 1
```

```
# Q(s,a) = v = Q(s,a) + Eta[reward(s,a) + gamma[max[Q(s^,a^)]] - Q(s,a)]
```

Badia et al. 2020. “Agent57: Outperforming the Atari Human Benchmark.” In *International Conference on Machine Learning*, 507–17. PMLR.

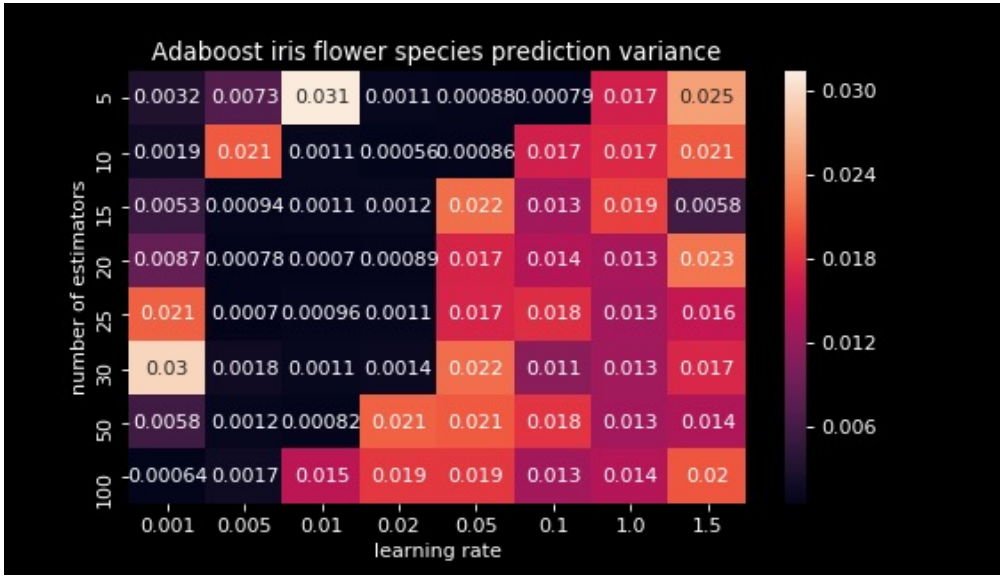
Environment

- Adaboost classifier: accuracy
- EDA – Need about 25 cross-folds



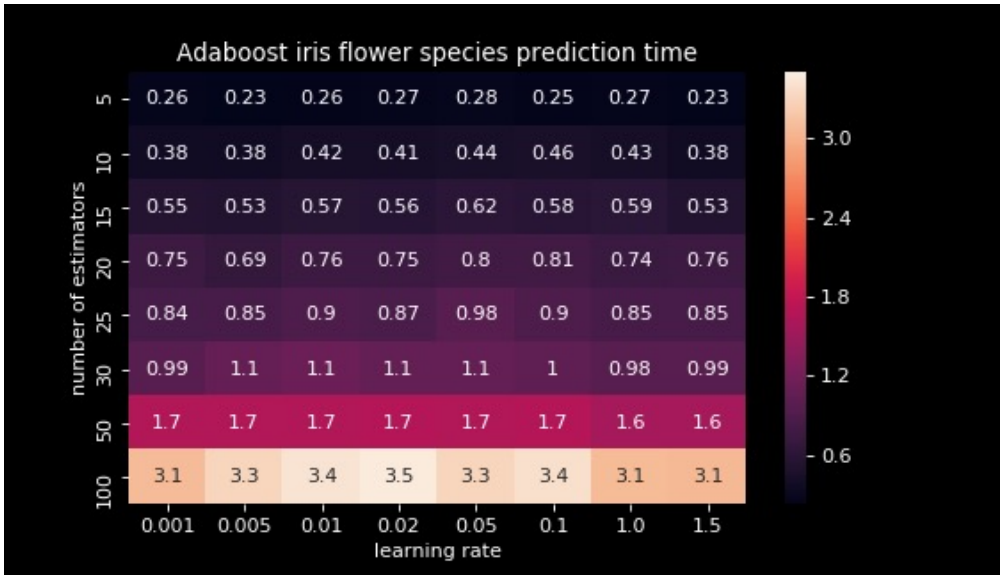
Environment

Adaboost classifier: variance



Environment

Adaboost classifier: time

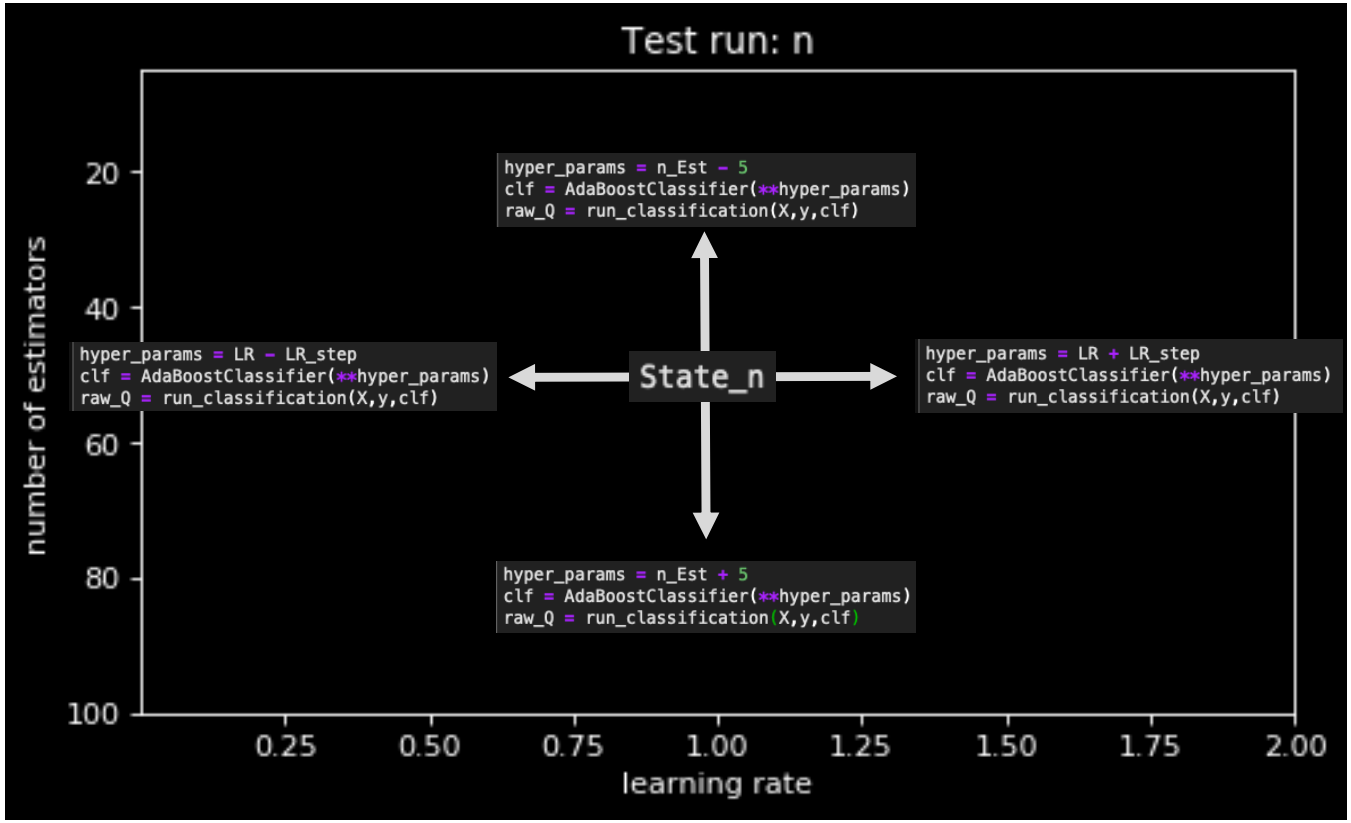


Coding objectives

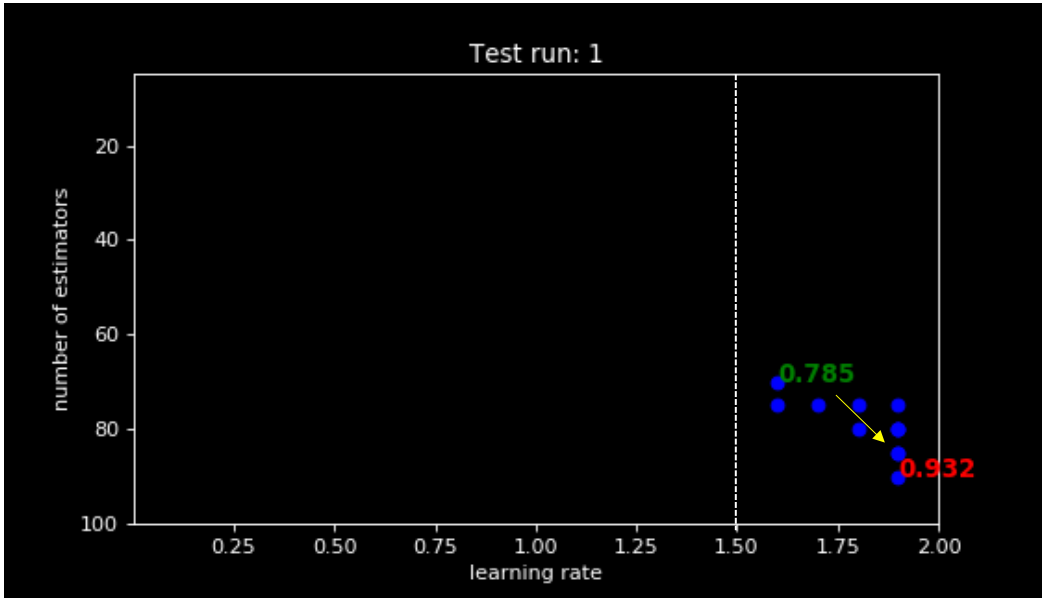
- Use Q-table to avoid leaving hyperparameter search space
- Balance exploration and exploitation
- Don’t leave optimal area once located

Algorithm summary

```
reward = 1 * raw_Q[0] + .0001/raw_Q[1]
```



Results, test data

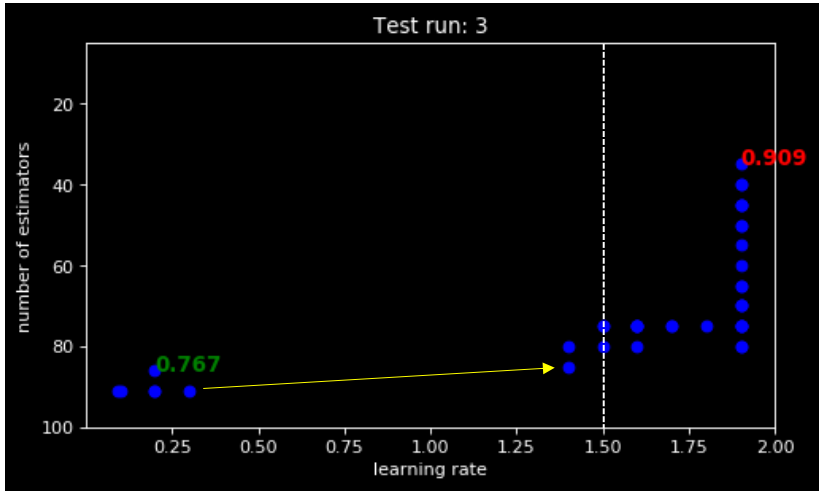


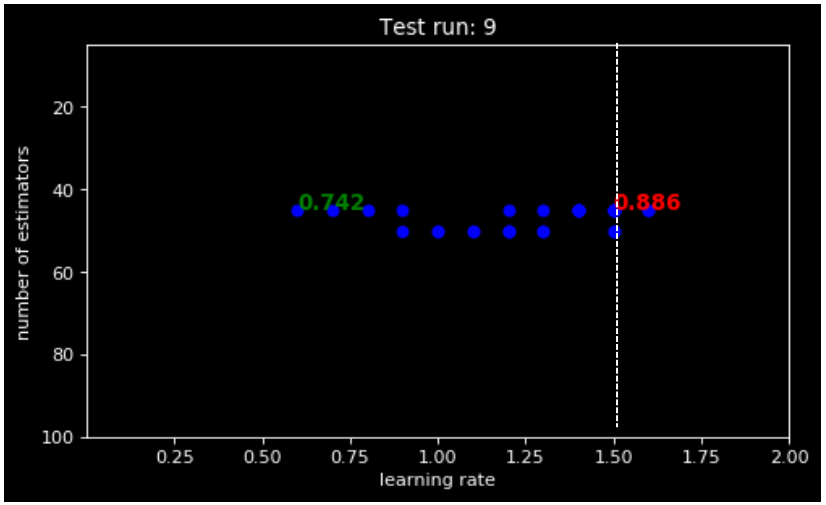
Q_table

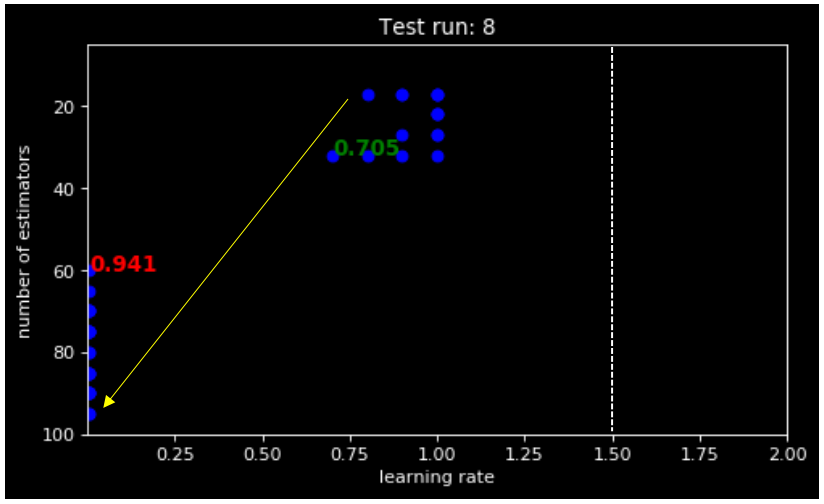
	State	Increase_n_Est	Decrease_n_Est	Increase_LR	Decrease_LR
0	0	0.000000	0.000000	0.000000	0.000000
1	1	0.894297	0.855114	0.861071	0.786295
2	2	0.855428	0.839140	0.895433	0.884702
3	3	0.860401	0.857960	0.956131	0.872203

	State	n_Estimators	Learning_rate	Score	Variance	Time
0	0.0	70.0	1.6	0.785276	0.016199	2.234745
1	1.0	75.0	1.6	0.882108	0.008204	2.434876
2	2.0	75.0	1.7	0.877912	0.005707	2.350607
3	3.0	75.0	1.8	0.911588	0.002245	2.341592
4	4.0	75.0	1.9	0.896848	0.003518	2.336238
5	5.0	80.0	1.9	0.917908	0.002616	2.558930
6	6.0	85.0	1.9	0.918944	0.003286	2.647718
7	7.0	80.0	1.9	0.934748	0.001797	2.485806
8	8.0	80.0	1.8	0.918964	0.003229	2.551576
9	9.0	80.0	1.9	0.913696	0.002109	2.533007
10	10.0	85.0	1.9	0.912644	0.002351	2.746907
11	11.0	90.0	1.9	0.931596	0.001328	2.813750

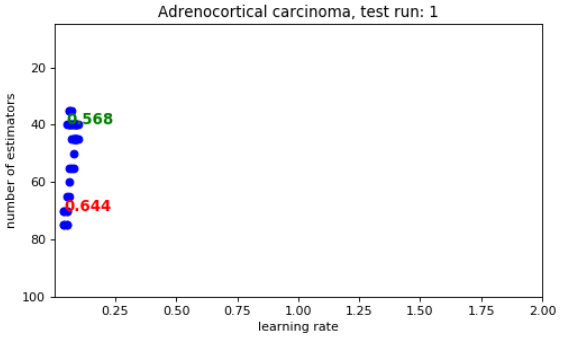
Results, test data

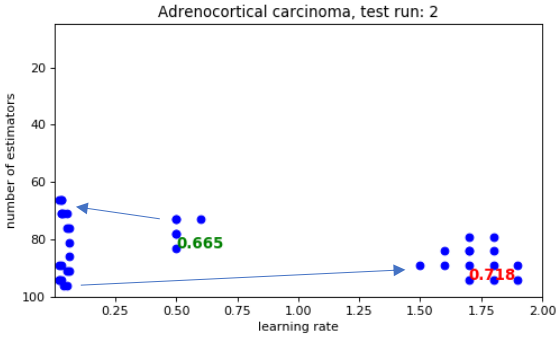


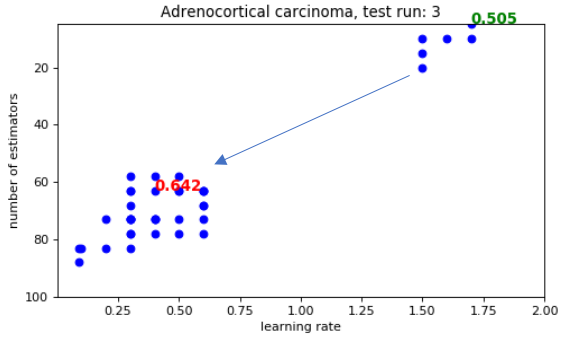


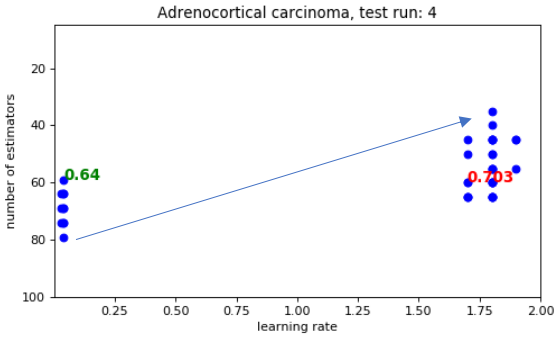


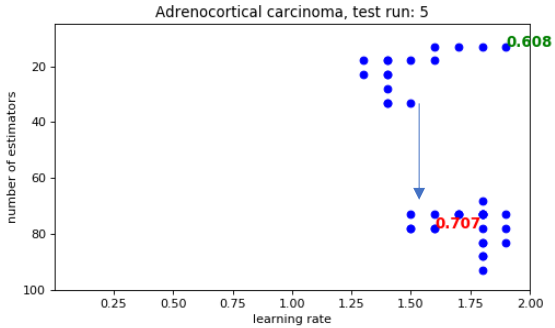
Results, genomic data

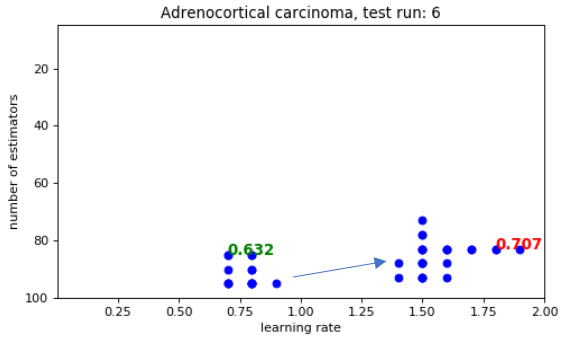












Summary

- Environment setup to conceptually map to value (Q) function
  - Map actions back to environment
- Balance of computational expense to precision of accuracy scores
  - Number of cross-folds
- Design of RL algorithm controls for generalizability