
Basketball
Curiosity
BLACK HAWKS

# Problem


Prediction of NBA shots results

# Data

**ProBasketballAPI.com**

NBA API FOR NBA STATISTICS, DRAFTKINGS NBA DATA, AND SPORTSVU DATA.

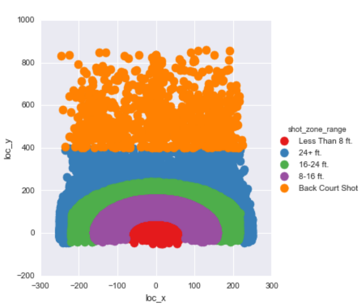
>400,000 shot data from 2 seasons collected into one dataset:

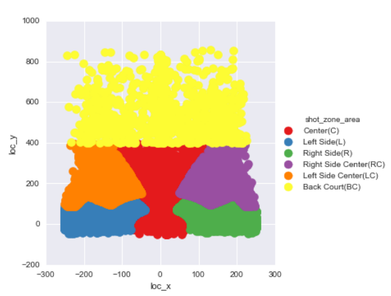
| | action_type | minutes_remaining | period | seconds_remaining | shot_distance | shot_made_flag | shot_zone_area | loc_x | loc_y | shot_zone_basic | shot_zone_range | shot_type | position | season | is_home |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **51643** | Driving Finger Roll Layup Shot | 2 | 2 | 20 | 3 | 1 | Center(C) | 37 | 9 | Restricted Area | Less Than 8 ft. | 2PT Field Goal | Guard | 2014 | 0 |
| **149572** | Layup Shot | 6 | 2 | 41 | 2 | 1 | Center(C) | 20 | 11 | Restricted Area | Less Than 8 ft. | 2PT Field Goal | Guard | 2014 | 1 |
| **149573** | Jump Shot | 6 | 2 | 1 | 3 | 0 | Center(C) | 32 | 12 | Restricted Area | Less Than 8 ft. | 2PT Field Goal | Guard | 2014 | 1 |
| **149574** | Jump Shot | 4 | 2 | 44 | 24 | 0 | Left Side Center(LC) | -218 | 116 | Above the Break 3 | 24+ ft. | 3PT Field Goal | Guard | 2014 | 1 |
| **149575** | Jump Shot | 2 | 2 | 31 | 20 | 0 | Center(C) | 39 | 197 | Mid-Range | 16-24 ft. | 2PT Field Goal | Guard | 2014 | 1 |

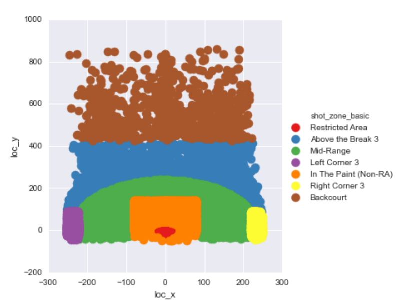
# Preprocess Bayes

## Binning Cropping Binarizing

| | label | distance | field_area | action_type | numpos | time_interval | period | season | is_home |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 4.0 | 0.0 | Alley Oop Dunk Shot | 50 | 8.0 | 1 | 2014 | 0 |
| **1** | 1 | 4.0 | 0.0 | Layup Shot | 50 | 11.5 | 3 | 2014 | 0 |
| **2** | 1 | 20.0 | -2.0 | Jump Shot | 42 | 2.0 | 1 | 2014 | 0 |
| **3** | 1 | 12.0 | 2.0 | Jump Shot | 42 | 1.0 | 1 | 2014 | 0 |
| **4** | 0 | 12.0 | 0.0 | Jump Shot | 42 | 11.0 | 2 | 2014 | 0 |




```
period
1     26.119432
2     24.909347
3     24.574799
4     23.621421
5      0.642255
6      0.105660
7      0.022206
8      0.004880
```

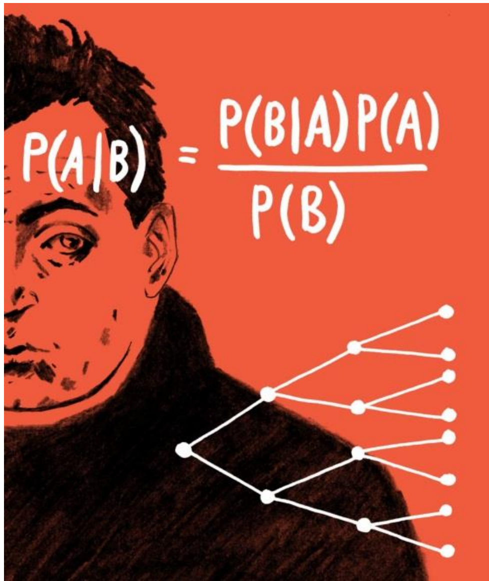
# Naïve Bayes

$$\text{P(H)} = \frac{\#(made\_shots)}{\#(total\_shots)} \approx 0.45$$

$$\text{P(E|H)} = \frac{\#(made\_shots \cap Evidence)}{\#(made\_shots)}$$

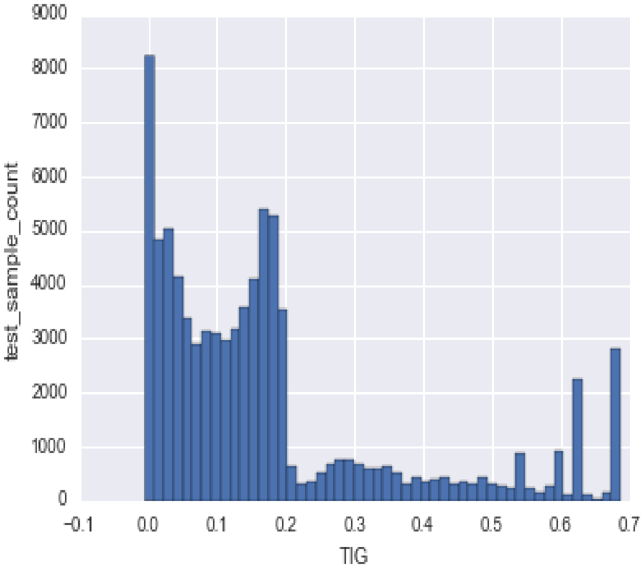
$$\mathrm{P(H|E)} = \frac{\mathrm{P(Evidence|Made_{shot}) * P(Made_{shot})}}{\mathrm{P(Evidence)}}$$

$$\mathrm{P(H|\overline{E})} = \mathrm{P(Made\_shot)} * \prod_{\mathrm{i}} \frac{\mathrm{P(E_i|Made\_shot)}}{\mathrm{P(E_i)}}$$

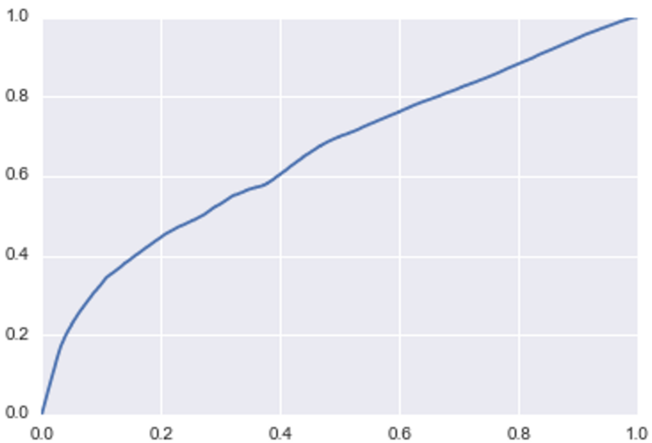


# Bayes Results

| | distance | angle | numpos | time_interval | period | season | is_home | Prediction | TIG | True Classification | Classification Distance |
|---|---|---|---|---|---|---|---|---|---|---|---|
| label | | | | | | | | | | | |
| 0 | 17.023524 | 0.029804 | 28.911082 | 5.768318 | 2.491930 | 2015 | 0.494541 | 0.420556 | 0.137418 | 0.655516 | 0.420556 |
| 1 | 13.013198 | -0.001095 | 30.510206 | 5.871841 | 2.464734 | 2015 | 0.508843 | 0.584432 | 0.229281 | 0.564494 | 0.415568 |



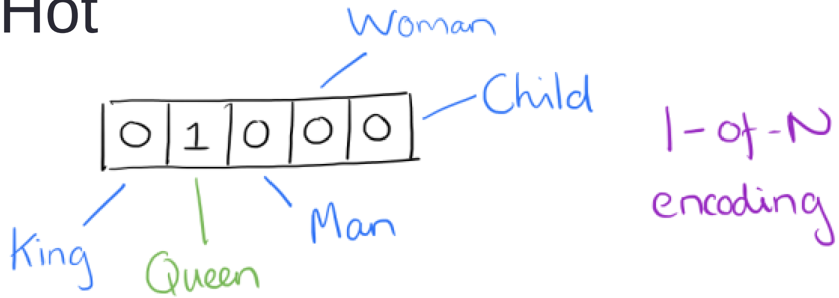

ROC (AUC = 0.656783963551)



# NN preprocess

- ReScaling
- Categories -> Binary / Ordinal


- Words -> One/Two/Three Hot



# Neural Network

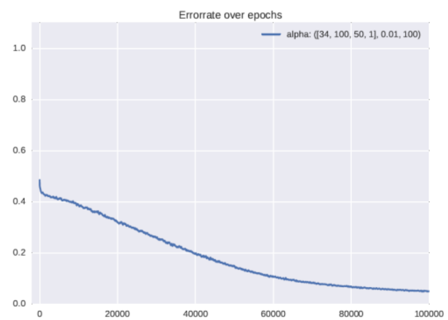
- Activation Function – Sigmoid
- Error Function – Quadratic
- Gradient – **SGD** (>400,000 samples)
- Learning rate – Search Optimized

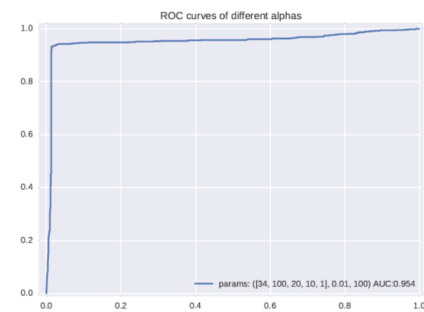
# NN Results and Sensitivity

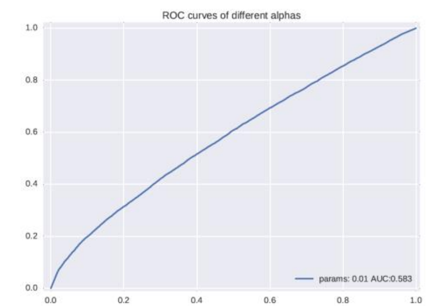
- Better Performance (AUC ~ 0.7)
- Low alpha values are better

| | TrainAUC | ValAUC |
|---|---|---|
| **alpha** | | |
| **0.001** | 0.691078 | 0.686941 |
| **0.010** | 0.681293 | 0.676275 |
| **0.100** | 0.622076 | 0.620140 |
| **1.000** | 0.557134 | 0.557724 |

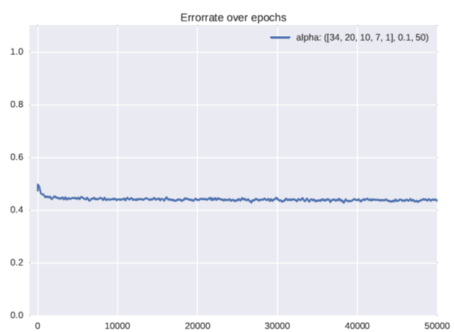
- |Train|=2000

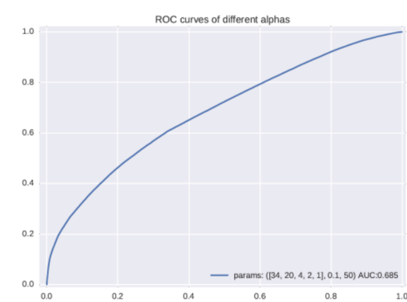


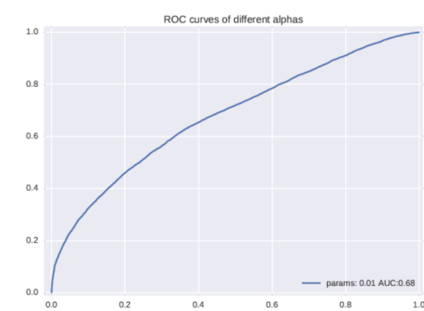




- |Train|=284641



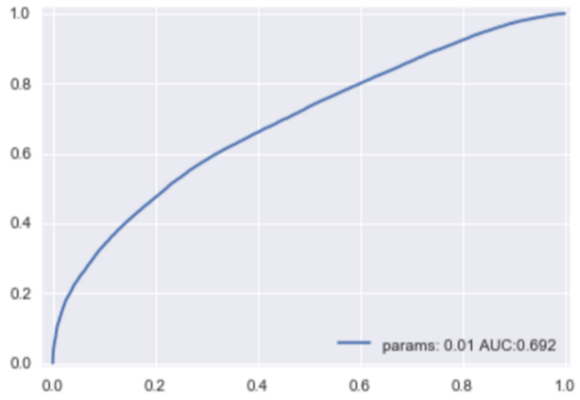
# NN optimization

- Genetic Algorithm to find best parameters set

- Target value – AIC $\underset{\theta}{\operatorname{argmin}}\{-Z * lan(ValidationAUC) + 2 * K\}$

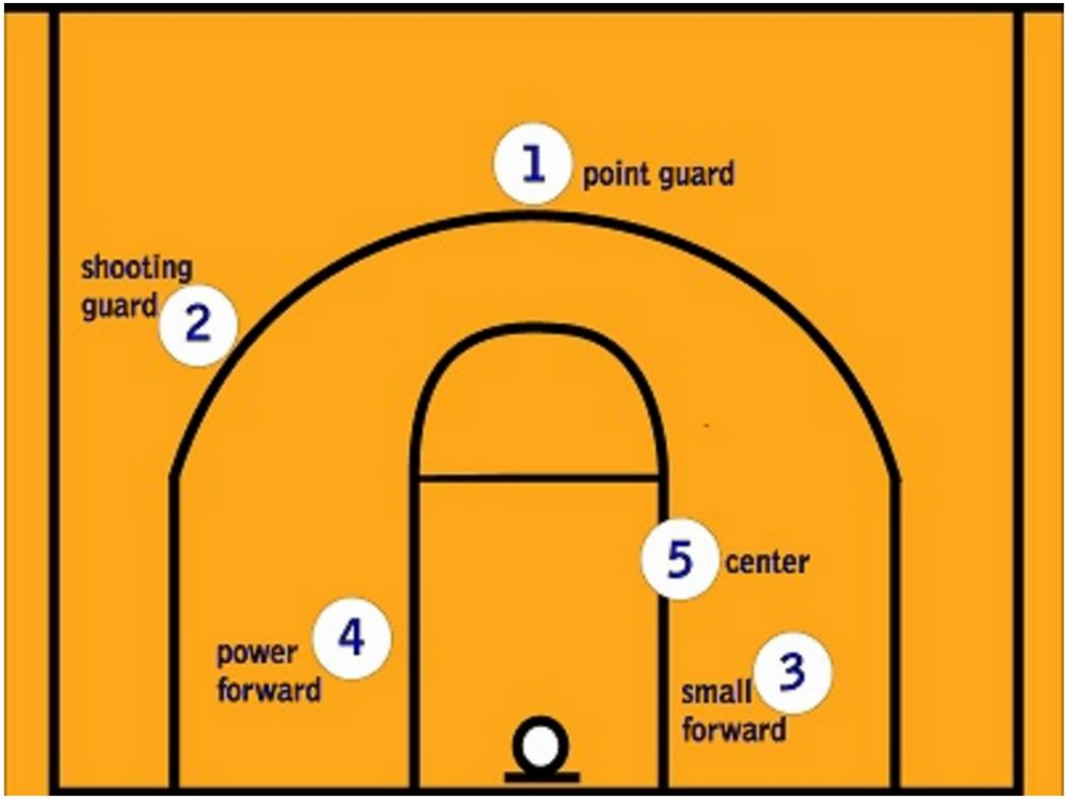
- Best : {34}, 50, 35 , [ 17, 8 ] , {1} - AUC = 0.692

Input Optional Output



# Curiosity – Problem Definition

- Given a period of game we have 5 players with different roles in different position to shoot.

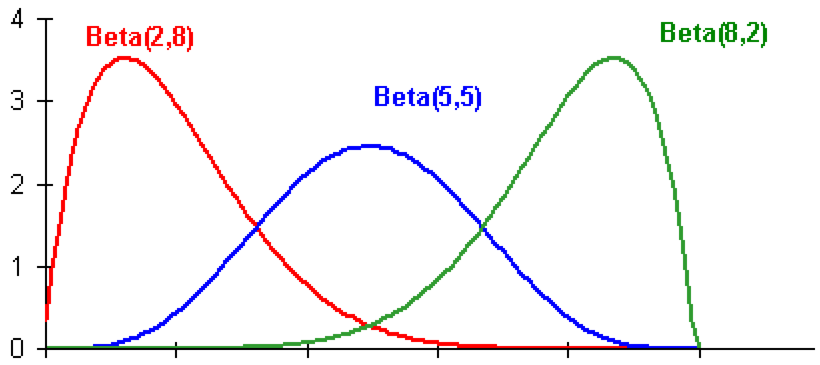


# Implementation

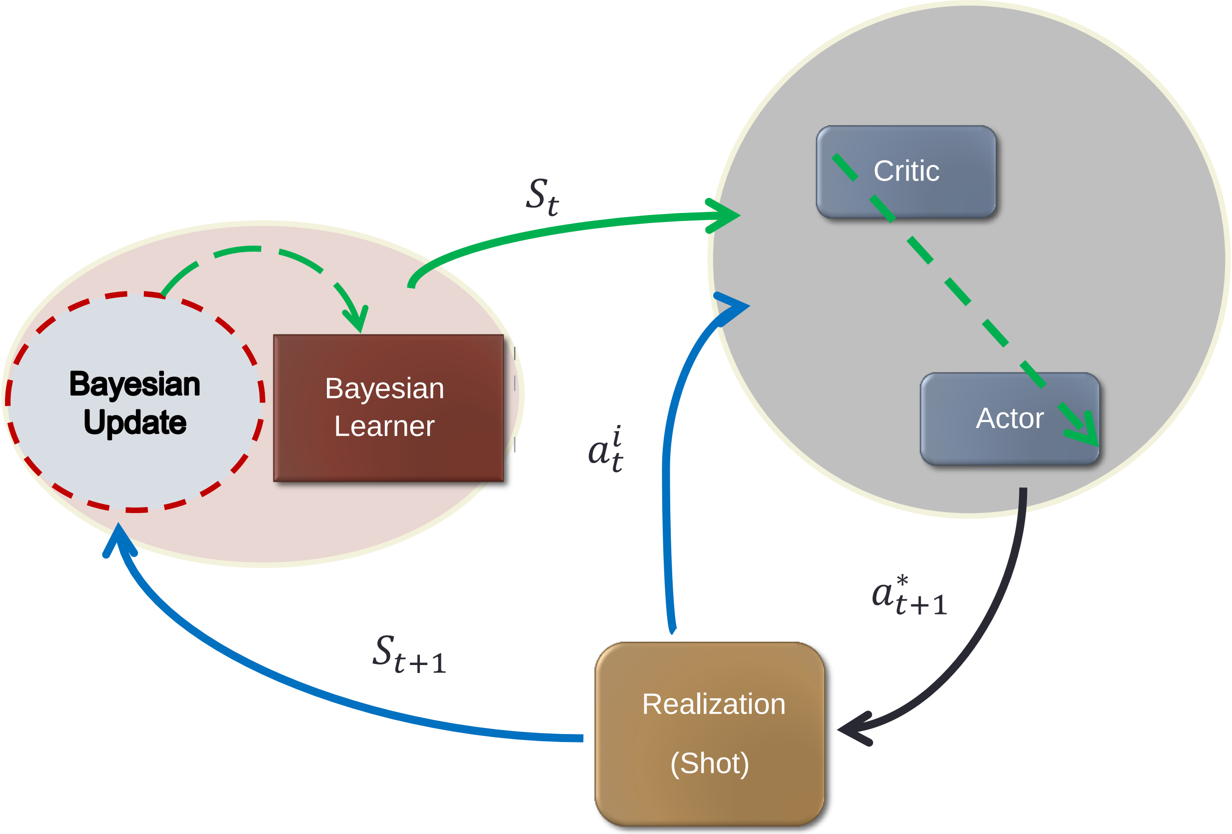
- Bayes Learner will use only 3 key features
  (Location, Position, Action type)

- Features were reduced to small option space

- Beta Distribution was used, to improve Expected Information Gain (and Entropy) calculations:



# Loop



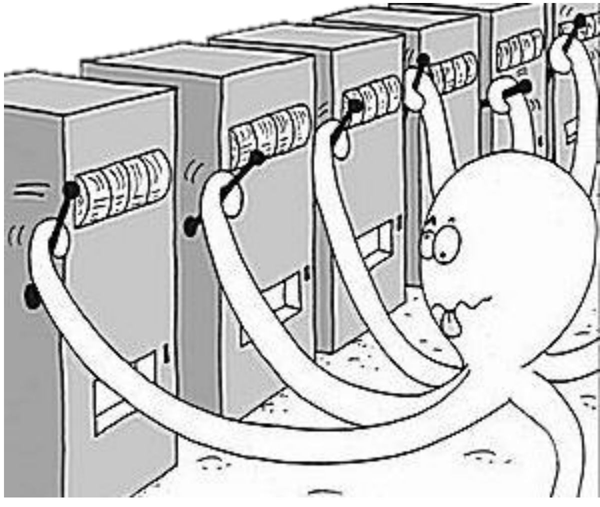
$S_t$
Critic
Bayesian Update
Bayesian Learner
$a_t^i$
Actor
$a_{t+1}^*$
$S_{t+1}$
Realization
(Shot)

# Actor - Critic

- Critic calculated expected value of each action:

$$Q(S_t, a_i) = \lambda * EIG(a_i, S_t) + (1 - \lambda) * \Pr(Shot|a_i, S_t)$$

- Actor makes weighted (Thompson sampling) decision, proportianally to the expected value

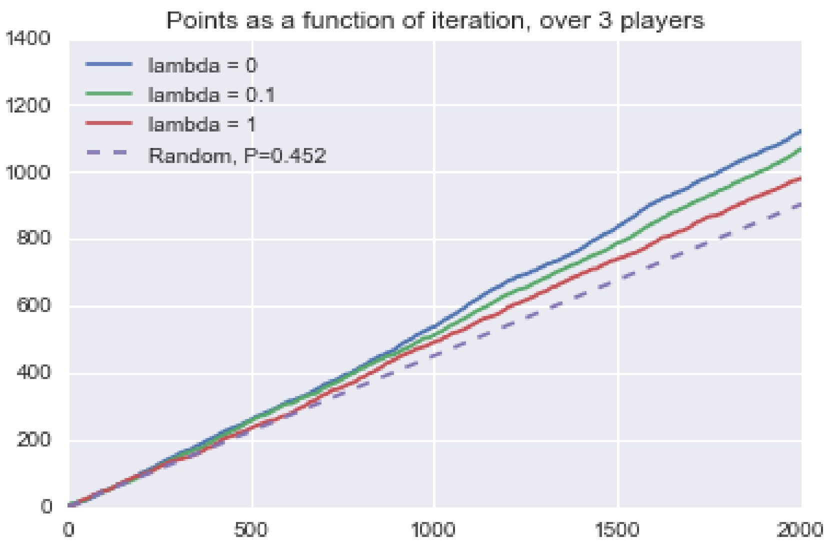


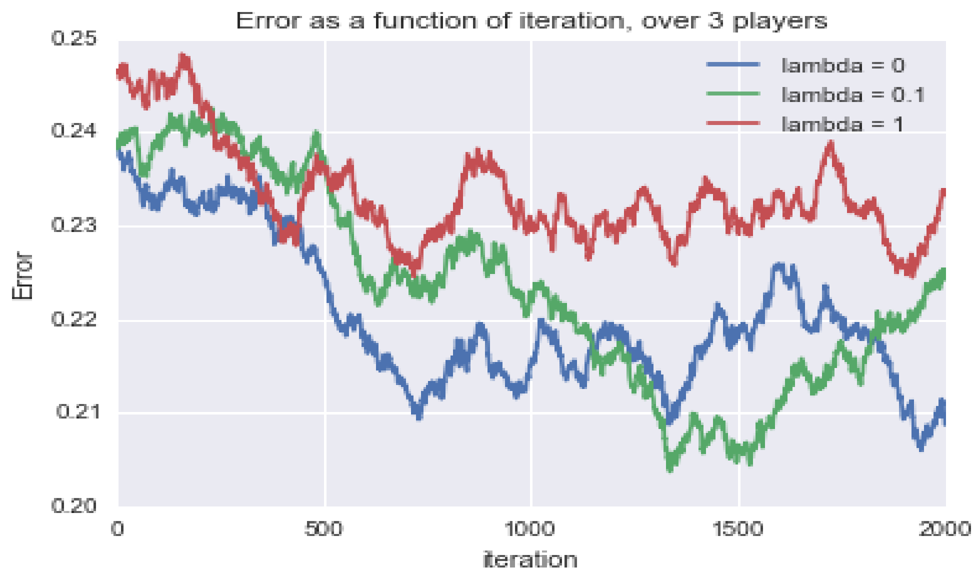
# Comparison

- Different $\boldsymbol{\lambda}$ :
  - Greedy player ($\lambda = 0$)
  - Balanced player ($\lambda = 0.1$)
  - Curious player ($\lambda = 1$)

# Results

Starting with **zero** knowledge and playing **2000** rounds we got:

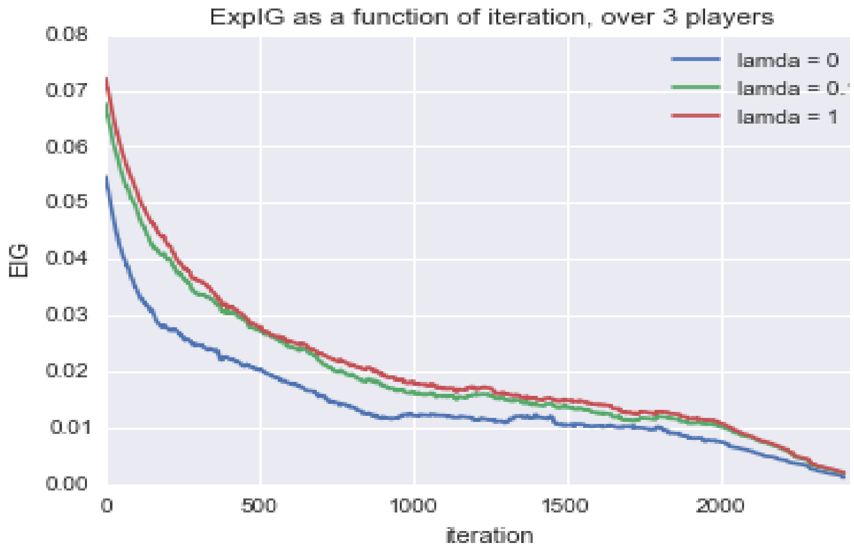


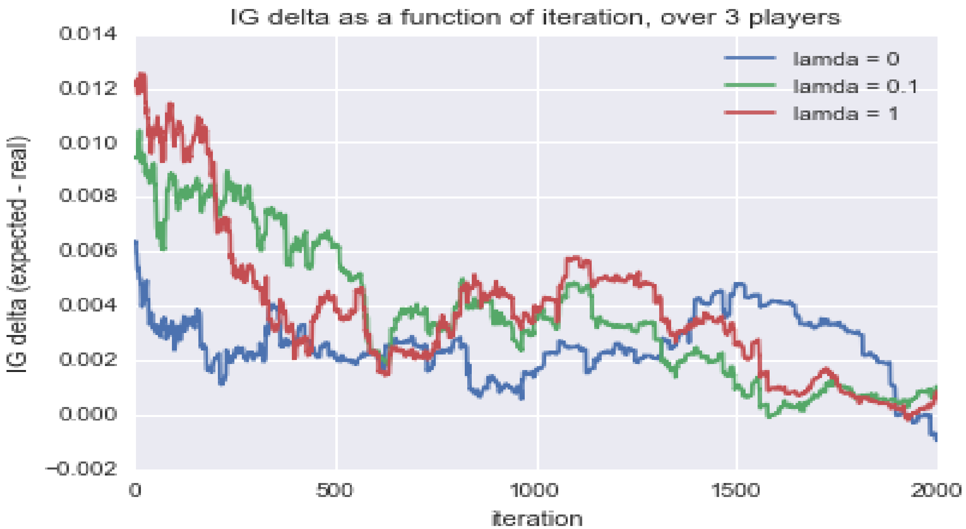

Greedy is going for the point

Curious is going for the knowledge

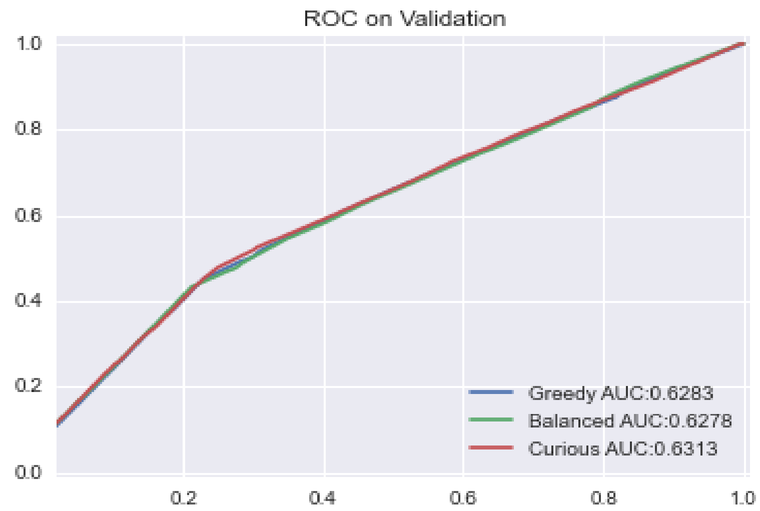
# Results



ExpIG as a function of iteration, over 3 players
EIG
iteration
lamda = 0
lamda = 0.1
lamda = 1
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
500
1000
1500
2000

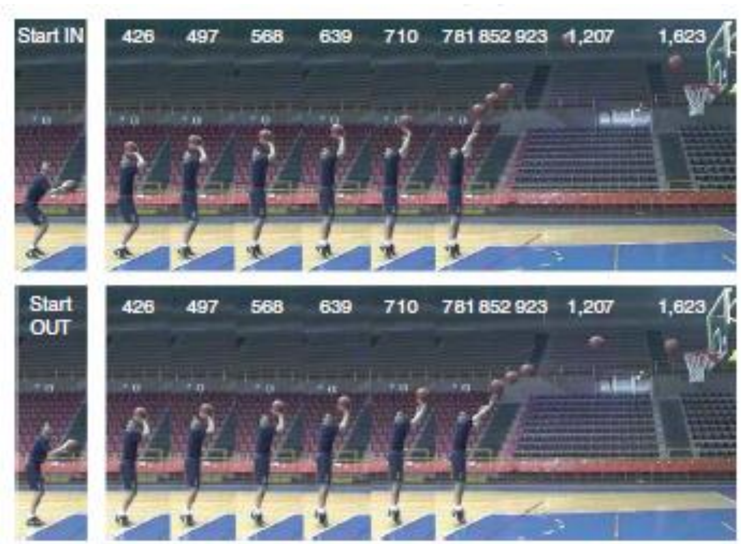
IG delta as a function of iteration, over 3 players
IG delta (expected - real)
iteration
lamda = 0
lamda = 0.1
lamda = 1
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
−0.002
0
500
1000
1500
2000

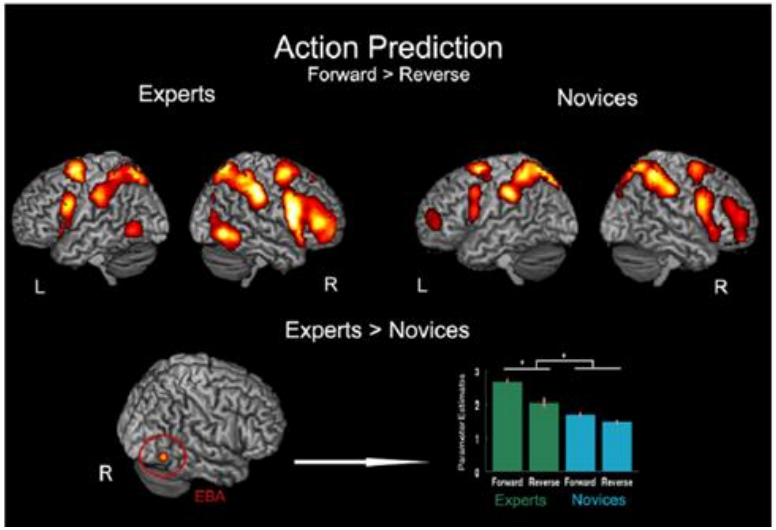



ROC on Validation
Greedy AUC:0.6283
Balanced AUC:0.6278
Curious AUC:0.6313
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8
1.0

# Brain

Action anticipation in in elite basketball players




- **Visual** (Abstract and Detailed) and **Motor** (Mirroring) brain areas were more active among experts
- The ‘**Surprise**’ effect was noticed when experts were wrong

# From noob to pro