Dataset

Junior League Players Dataset(divided by seasons)
players' stats when playing in junior leagues, not related to AHL/NHL

OHL (Ontario Hockey League)	Features: Rank, age, nationality, name, position, GP, G, A, P, PIM, Plus/ Minus, PPG, SHG, GWG, GGP, AGP, PGP	Amount: Season 2012-2016, inclusive 2,198 records data source: http://www.quanthockey.com/

sample:

rank	age	nationality	name	pos	GP	G	A	P	PIM	plusMinus	PPG	SHG	GWG	GGP	AGP	PGP
1	19	US	Vincent Trocheck	F	63	50	59	109	58	49	10	5	5	0.794	0.937	1.73

QMJHL (Quebec Major Junior League)	Features: Rank, age, nationality, name, position, GP, G, A, P, PIM, Plus/ Minus, PPG, SHG, GWG, GGP, AGP, PGP	Amount: Season 2012-2016, inclusive 2,339 records data source: http://www.quanthockey.com/

sample:

rank	age	nationality	name	pos	GP	G	A	P	PIM	plusMinus	PPG	SHG	GWG	GGP	AGP	PGP
1	19	US	Conor Garland	F	62	39	89	128	97	26	13	1	6	0.629	1.435	2.065

WHL(Western Hockey League)
Players' stats when playing in WHL, and stats about whether they get NHL contract or not

---

Features: player, team, GP, G, A, TP, PPG, PIM, Plus/Minus
Features(After preprocessing): player, position, GP, G, A, TP, PPG, PIM, Plus/Minus, Contract

Amount: 2,736 records

data source: http://www.eliteprospects.com/

sample:

#	PLAYER	TEAM	GP	G	A	TP	PPG	PIM	Plus/Minus
1	Hunter Shinkaruk(C/W)	Medicine Hat Tigers	66	49	42	91	1.38	38	17

sample (After cleaning):

player	position	GP	G	A	TP	PPG	PIM	Plus/Minus	contract
Tyler Johnson	C	71	53	62	115	1.62	48	27	0

AHL Players Dataset
Players' stats when playing in AHL

Features: Rank, age, nationality, name, position, GP, G, A, P, PIM, Plus/ Minus, PPG, SHG, GWG, GGP, AGP, PGP

Amount: Season 2012-2016, inclusive
4,000 records

data source: http://www.quanthockey.com/

David Wilson's Dataset
NHL players' career stats

Features: name, careerGP, NumSeasons, AvgGP, CareerTOI, AvgTOI, CareerPts, Round, Age, D1_pos, D2_pos, Great160, Great120, Draft_Year, Birth_Date

It also contains ZeroGP NHL players info, and NHL players' info divided by stats

Amount: 3,076 records

Question & Method Tried

David's work and Schucker's paper mainly focus on predicting NHL drafted players' performance in NHL career games, not from Junior leagues to NHL/AHL.

Q1. Will a player get an NHL contract from a Junior League?

Experiment: Tried on WHL dataset

SVM:
When using Guassian kernel, setting penalty C=1, gamma=1 in the process of cross validation, SVM gives the minimum validation error. And applying this model to testing data, the accuracy is about **88.54%**

```
In [35]: clf1.score(test_fea_norm, test_target)

Out[35]: 0.8854816733067728
```

```
Out[33]: SVC(C=1, cache_size=200, class_weight=None, coef0=0.0, degree=3, gamma=1,
           kernel='rbf', max_iter=-1, probability=False, random_state=None,
           shrinking=True, tol=0.001, verbose=False)
```

Logistic Regression using Stochastic Gradient Descent:
When choosing learning rate as 0.001, it gives the best performance. And the test accuracy is about **60.20%**

```
In [9]: res

Out[9]: 0.6019900497512438
```

Conclusion: SVM works better than LR in this dataset.

Q2. if they get an NHL contract, will they play any NHL games, or will they be just in the AHL?

Q3. if they play an NHL games, will they get more than 160 NHL games?

Q4. Can we predict the number of games?

Q5. Can we predict the TOI?

Q6. Given the predictions, can we rank junior players and compare with NHL draft order.

Q7. For players who play both in the NHL and the AHL, can we predict their scores in the NHL from the AHL or vice versa?