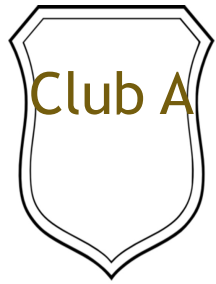# Pricing Football Players using Neural Networks

Sourya Dey

Final Project - Neural Learning and Computational Intelligence

April 19th 2017, University of Southern California
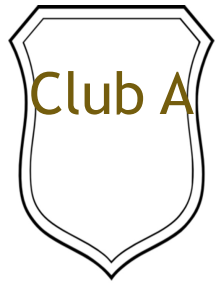
How the Football Transfer Market Works

Nominal Price = $2 Million

Club A

Club B

Club C

Potential youngster

International Star

Clinical goalscorer

Reliable defender

94

How the Football Transfer Market Works

Nominal Price = $2 Million

Club A

Initial Bid = $2.2 Million

Club B

Rival Bid = $2.1 Million

Club C

Potential youngster

International Star

Clinical goalscorer

Reliable defender

How the Football Transfer Market Works

Club A

Nominal Price = $2 Million

Asking Price = $2.5 Million

Initial Bid = $2.2 Million

Club B

Rival Bid = $2.1 Million

Club C

Potential youngster

International Star

Clinical goalscorer

Reliable defender

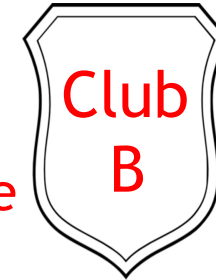# How the Football Transfer Market Works

Nominal Price = $2 Million

Club A

Asking Price = $2.5 Million

Initial Bid = $2.2 Million

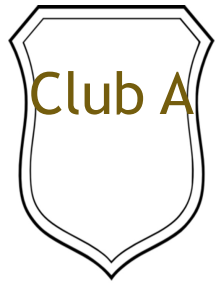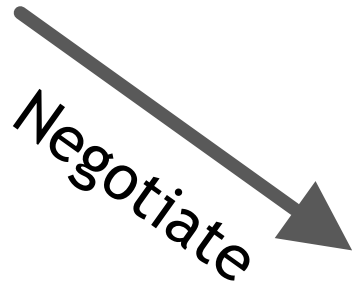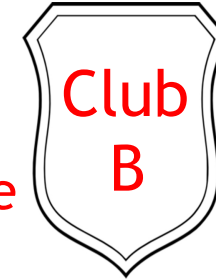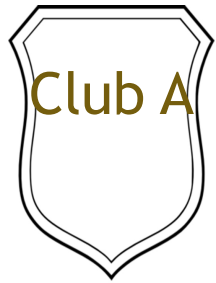Can't match asking price

Club B

Rival Bid = $2.1 Million

Counter Bid = $2.3 Million

Club C

Potential youngster

International Star

Clinical goalscorer

Reliable defender

How the Football Transfer Market Works

Nominal Price = $2 Million

Club A

Initial Bid = $2.2 Million

Can't match asking price

Club B

Asking Price = $2.5 Million

Negotiate

Rival Bid = $2.1 Million

Counter Bid = $2.3 Million

Club C

Potential youngster

International Star

94

Clinical goalscorer

Reliable defender

How the Football Transfer Market Works

Club A

Nominal Price = $2 Million

Asking Price = $2.5 Million

Initial Bid = $2.2 Million

Can't match asking price

Club B

Rival Bid = $2.1 Million

Counter Bid = $2.3 Million

Club C

Negotiate

Potential youngster

International Star

Clinical goalscorer

Reliable defender

94

How the Football Transfer Market Works

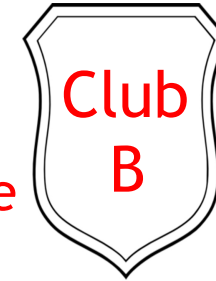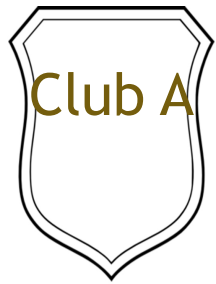Nominal Price = $2 Million

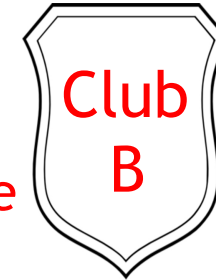Club A

Asking Price = $2.5 Million

Initial Bid = $2.2 Million

Club B

Can't match asking price

Rival Bid = $2.1 Million

Counter Bid = $2.3 Million

Club C

SOLD for $2.4 Million

Negotiate

Potential youngster

International Star

94

Clinical goalscorer

Reliable defender

Seal deal before transfer window closes!

Money Talks

" *The January I moved from Birmingham to Blackburn, I sat there by the phone for 13 days, worrying, waiting for it to go through - I lost a stone.*"

-- Robbie Savage
Former Premier League Midfielder
Football Focus Special, 2011

**Negotiations are a big part of football.
What is the optimum price for a player?**

Data Source: FIFA 2017

A football simulation video game with official licenses for most major teams and players in the world

Player Attributes

Authenticity of Data

Season Ticket Holders

Managers

Scouts

300 data points per player

Feedback from 9000 data reviewers

Stats updated weekly on EA servers as per real-life performance

Criticisms of Data Authenticity

- Subjectivity by EA
- Not enough data on fledgling players
- Imperfect formula

" *If Messi were playing in the Irish league, his attributes would drop simply because he's not on the highest level anymore.* "

-- Michael Mueller-Moehring
EA Sports Producer and Ratings Specialist
Interview with ESPN FC, September 2016

Gathering Data –

Players

NAME	ACCELERATION	AGGRESSION	AGILITY	BALANCE	BALL CONTROL	COMPOSURE	CROSSING	CURVE	DEF	DRI	DRIBBLING	FINISHING	FREE KICK ACCURACY	HEADING ACCURACY	INTERCEPTIONS	JUMPING	LONG PASSING	LONG SHOTS	MARKING	OVA	PAC	PAS	PENALTIES	PHY	POSITIONING	POT	REACTIONS	SHO	SHORT PASSING	SHOT POWER	SLIDING TACKLE	SPRINT SPEED	STAMINA	STANDING TACKLE	STRENGTH	VISION	VOLLEYS	AGE	INTERNATIONAL REPUTATION	SKILL MOVES	WEAK FOOT	VALUE
Bartra	78	79	69	69	78	73	50	56	81	73	70	43	45	78	86	78	42	82	81	77	69	52	70	61	83	81	48	79	61	81	76	65	79	68	75	43		25	2★	3★	3★	$ 18.4M
M. Antonio	88	75	86	79	78	67	73	66	52	81	82	74	60	88	40	84	67	71	44	80	89	69	73	82	77	81	75	73	71	75	53	89	81	56	86	65	69	26	1★	4★	4★	$ 18.4M
Ganso	36	56	67	56	84	85	83	80	45	81	85	72	76	65	47	33	84	79	41	80	35	85	76	64	74	81	75	86	77	37	34	71	45	67	87	75		26	3★	4★	3★	$ 17.8M
J. Brooks	69	76	51	60	82	77	36	30	82	63	66	33	28	85	82	74	69	29	80	80	74	56	45	77	40	85	69	36	70	50	79	78	64	84	84	57	31	23	2★	2★	3★	$ 17.8M
M. Nastasić	66	78	64	58	66	78	46	29	82	51	58	27	32	78	81	67	22	81	80	70	57	47	75	30	85	79	31	69	44	86	78	65	84	78	55	32		23	2★	2★	3★	$ 17.8M
R. Soriano	79	69	83	72	83	78	63	82	82	73	65	77	48	80	78	80	62	72	81	75	84	81	67	80	71	83	75		26	2★	3★	3★	$ 17.8M
B. Schweinsteiger	50	80	67	63	81	88	78	82	75	74	72	70	78	78	83	80	84	80	68	82	51	82	77	73	82	84	76	82	75	52	58	74	77	84	82	31	4★	3★	3★	$ 17.8M
D. Didavi	76	70	75	73	82	73	75	52	79	81	74	58	51	76	87	45	80	77	75	65	72	82	81	82	78	87	49	75	67	58	75	70	73	26	2★	3★	3★	$ 17.8M
D. Blind	65	71	73	75	80	86	75	82	76	46	55	71	86	71	80	64	82	81	61	78	56	76	64	83	85	56	81	68	77	86	84	73	62	26	3★	2★	4★	$ 17.8M
Iturraspe	58	66	53	45	80	65	57	68	79	70	69	60	66	78	84	52	84	69	72	81	60	77	58	73	72	81	83	63	86	67	70	84	80	81	47	27	2★	2★	3★	$ 17.8M
P. Herrmann	87	66	86	82	79	70	79	74	31	82	83	76	50	55	29	69	64	74	27	80	88	74	61	59	82	81	80	75	78	78	40	88	64	26	56	77	74	25	2★	3★	3★	$ 17.8M
J. Vardy	94	92	77	78	78	63	59	51	76	50	83	48	49	74	48	48	81	93	75	79	84	80	66	85	51	90	52	69	72	29	3★	3★	3★	$ 17.8M
J. Hector	76	70	71	68	81	81	79	79	55	65	73	81	61	75	61	77	81	77	75	65	76	83	69	60	79	69	81	77	84	81	70	51	26	3★	3★	3★	$ 17.8M
R. Boudebouz	77	58	81	75	81	49	82	33	81	83	78	80	45	32	61	76	27	80	76	80	87	62	73	81	70	78	81	79	30	59	79	71	26	2★	4★	4★	$ 17.8M
Roque Mesa	79	92	77	89	82	79	67	46	73	77	72	67	48	32	81	77	80	54	72	82	78	76	42	80	67	82	79	61	89	63	80	78	92	79	70	75	34	27	1★	2★	3★	$ 17.8M
T. Walcott	92	65	83	79	78	78	77	38	80	81	82	74	58	35	66	65	69	33	81	93	74	74	66	81	81	77	75	76	36	93	75	39	82	73	72	27	3★	3★	3★	$ 17.8M
Negredo L	64	78	67	57	79	67	53	68	39	73	72	81	69	83	32	79	66	77	36	81	65	66	80	80	80	81	76	81	75	85	32	65	65	34	88	64	81	30	3★	3★	3★	$ 17.3M
D. Drinkwater	68	74	69	78	80	77	64	70	74	75	74	76	70	61	69	76	56	83	73	70	68	76	53	75	74	81	79	67	81	76	76	83	79	74	77	63	26	1★	3★	4★	$ 17.3M
Raúl García	55	84	67	63	84	83	65	63	73	73	60	54	81	58	82	73	81	86	77	85	81	76	61	71	82	61	82	29	2★	3★	3★	$ 17.3M
D. Sidibé	80	76	67	67	77	75	77	77	76	70	82	66	78	80	82	74	43	79	76	85	67	69	80	88	73	69	42	23	2★	2★	2★	$ 17.3M
T. Kolodziejczak	68	83	61	50	72	56	63	59	79	66	72	22	34	72	81	79	65	43	78	80	71	64	49	84	57	84	75	35	71	47	82	74	83	80	85	58	26	24	1★	2★	2★	$ 17.3M
Iborra	54	78	54	38	80	73	53	63	69	67	72	70	81	85	74	82	64	85	82	54	75	72	83	62	82	73	81	85	75	84	88	74	66	28	2★	2★	4★	$ 17.3M
Luiz Gustavo	68	88	66	62	73	72	61	42	83	70	68	46	45	75	84	75	75	71	78	75	67	69	58	82	51	82	80	59	77	80	81	87	80	69	54	28	3★	2★	3★	$ 17.3M

Prices

Occur as quantized values except for a few outliers

Most Common Price = $1.1M

Rounded Min Price = $43K

Rounded Max Price = $37M
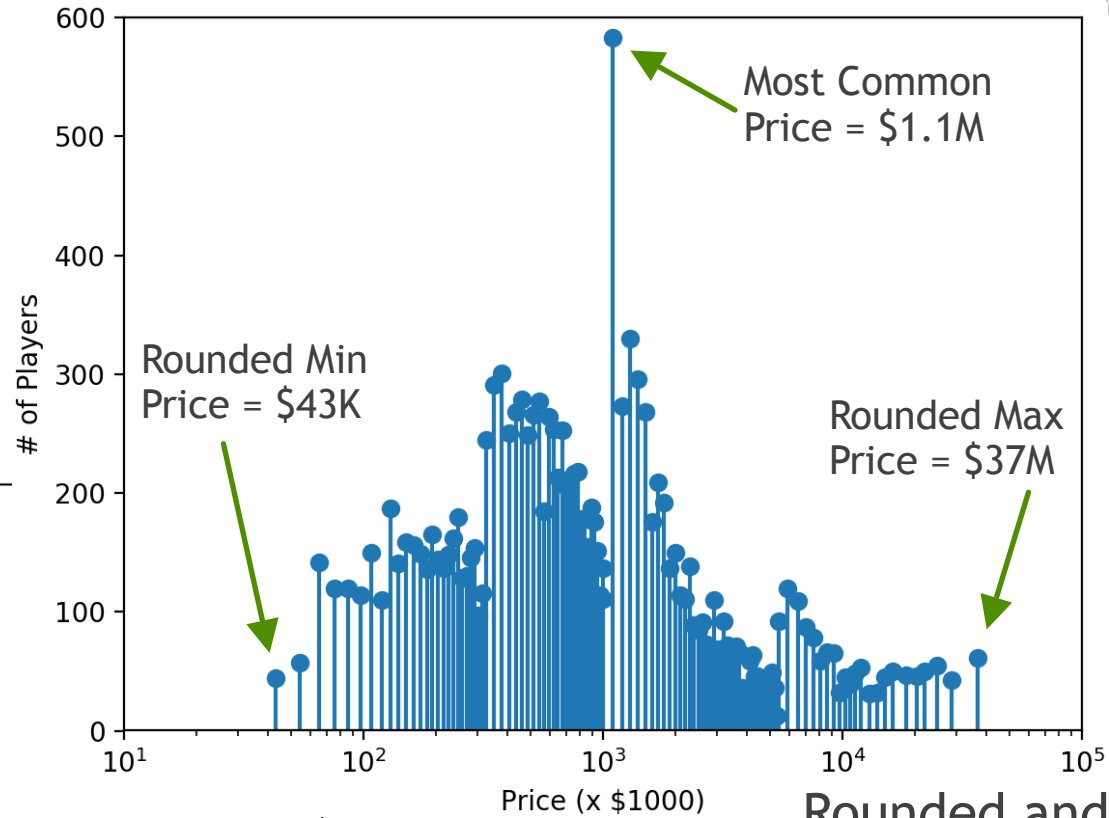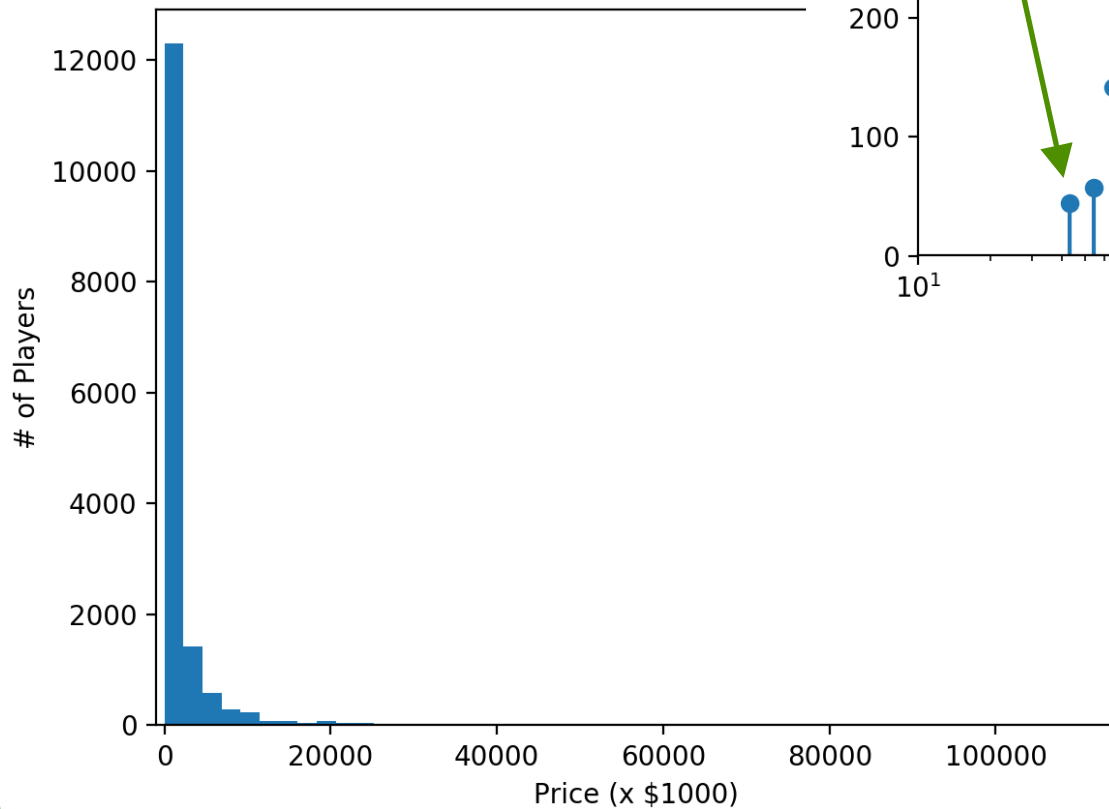
Rounded and Categorized
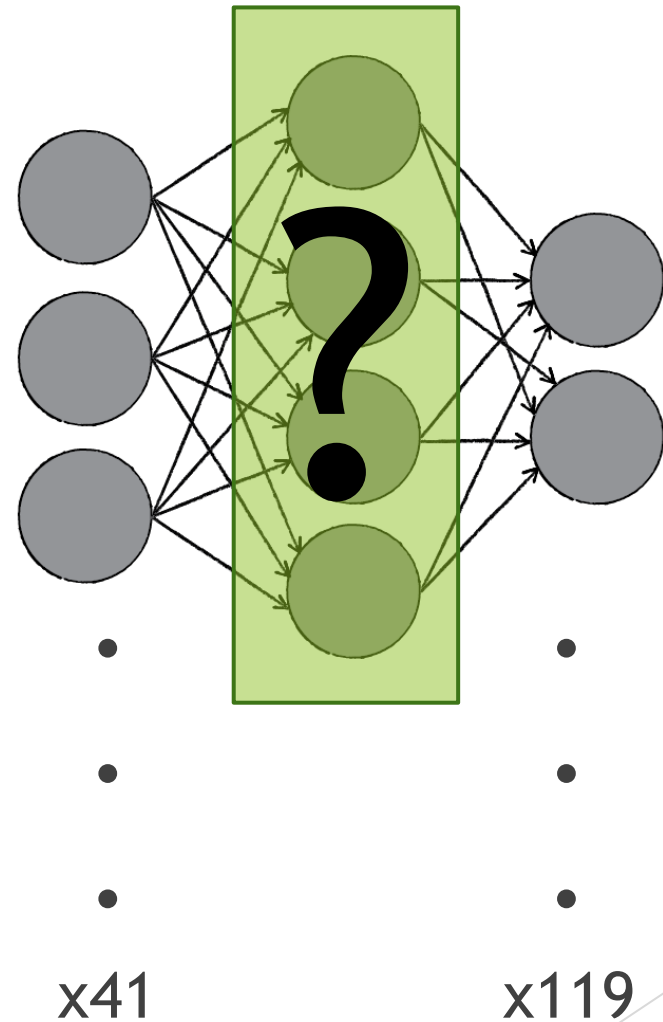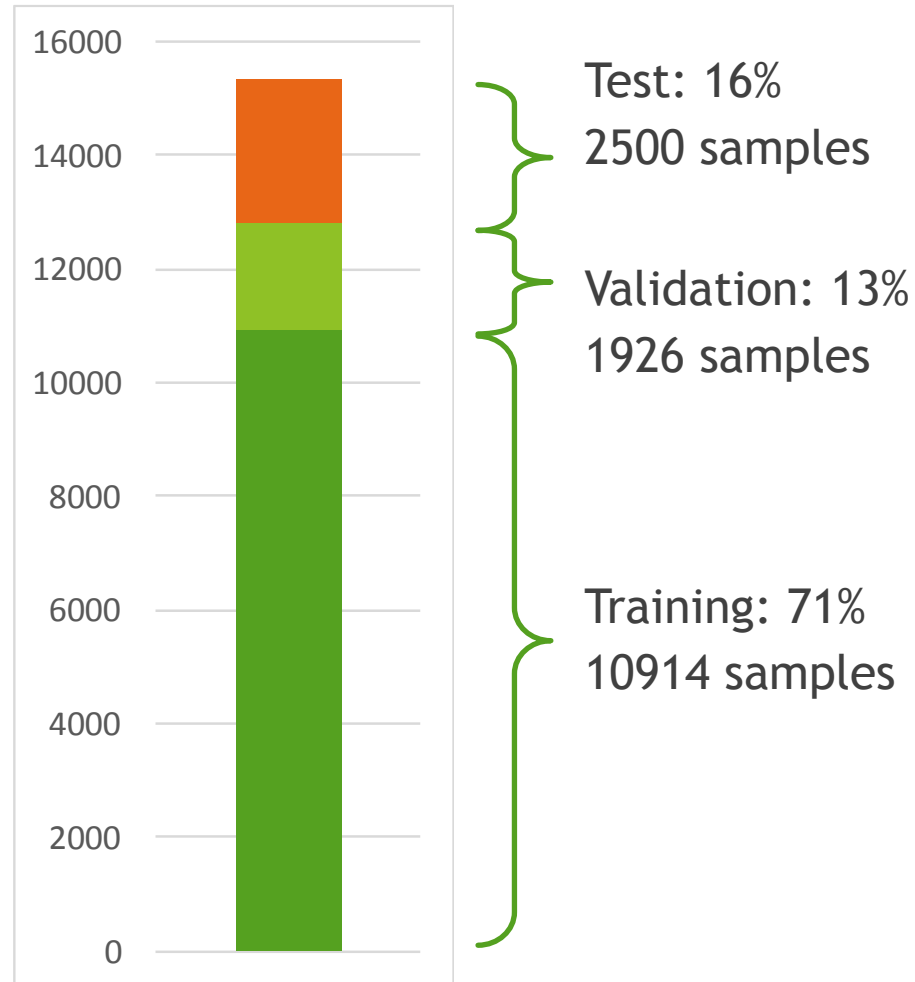
Original Bottom-Heavy Prices

9

Network Structure

- 119 output price values (categories)

- 41 input features
  - 37 on a scale of 0–99
  - Age on a range of 16-43
  - 3 on a scale of 1-5 stars
- Normalized to $\mu = 0, \sigma = 1$

x41   x119

Data Splits

Total:

15340 players
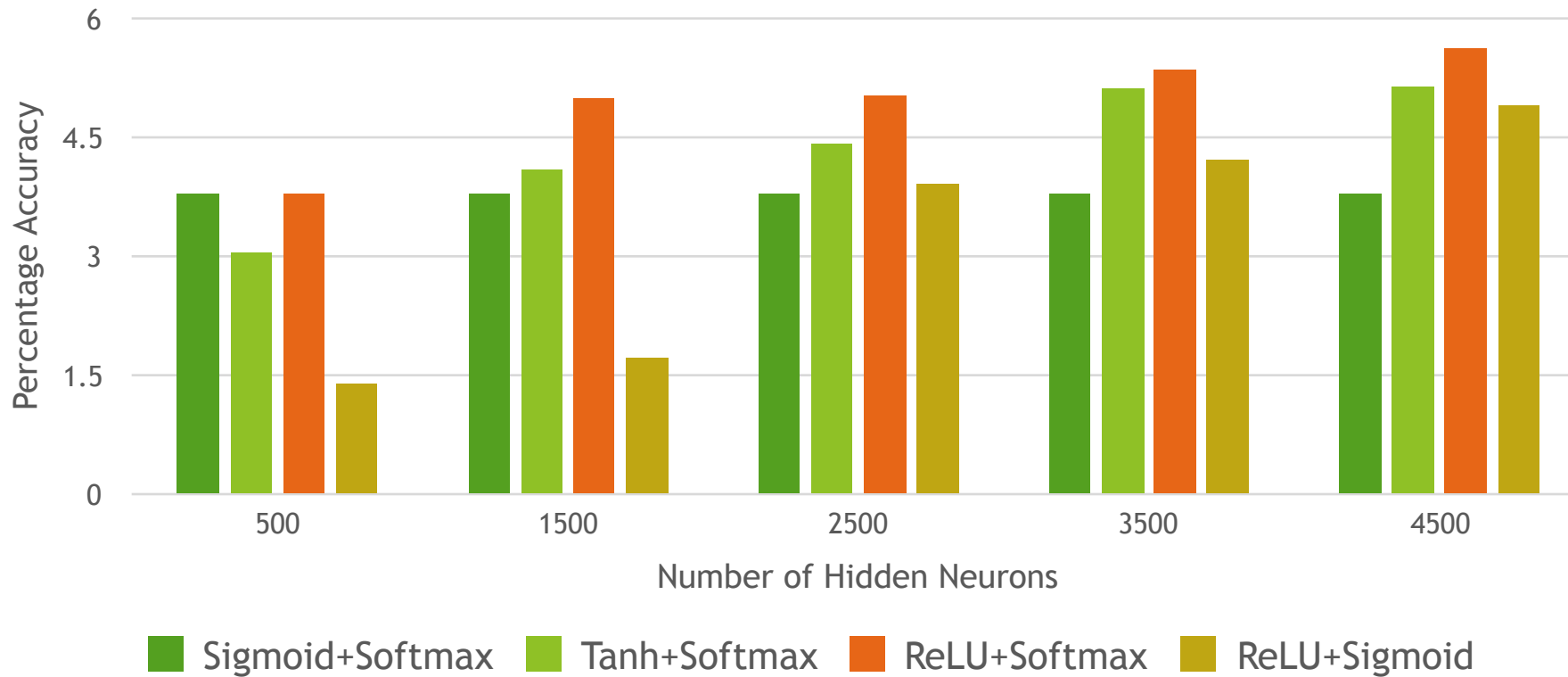

Test: 16%
2500 samples

Validation: 13%
1926 samples

Training: 71%
10914 samples

Network Experiments

- Activation functions
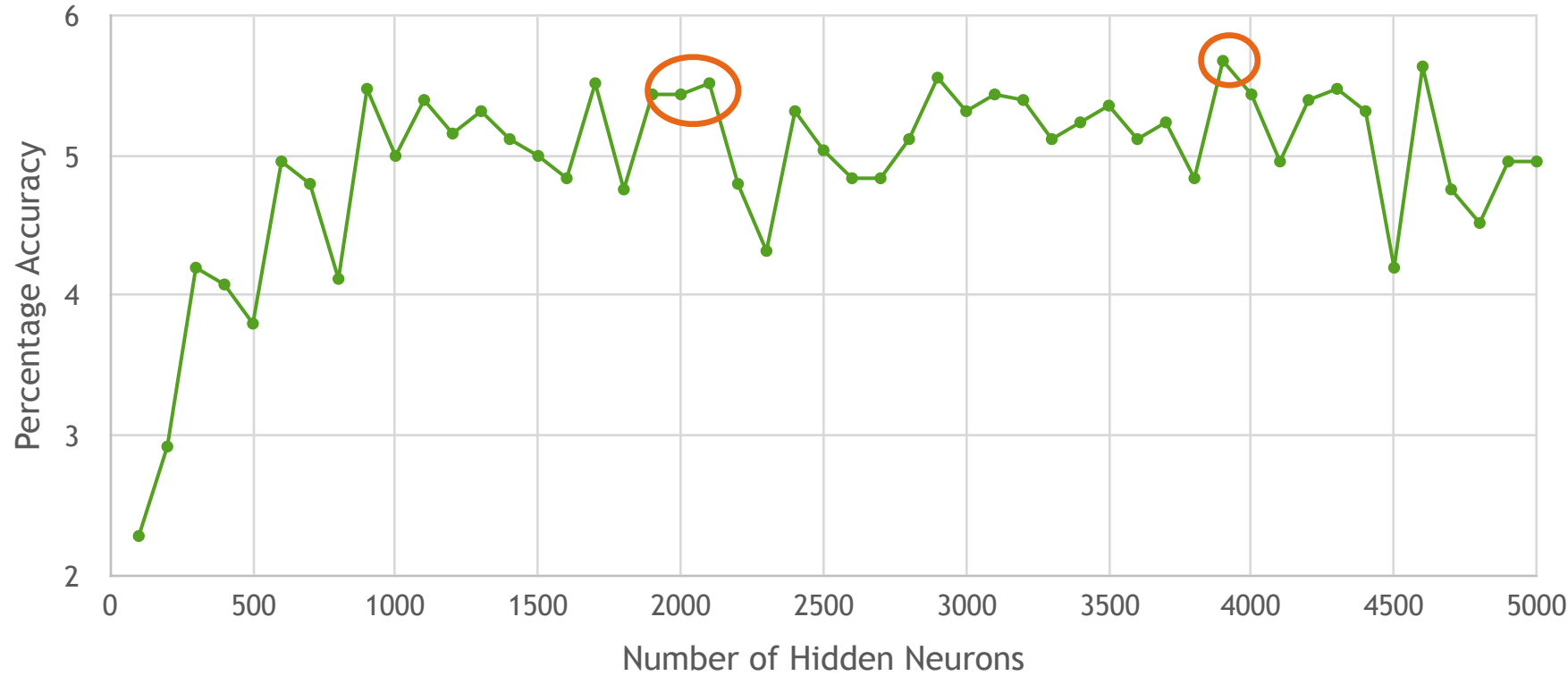- Number and size of hidden layers
- Learning rate and its decay
- Regularization (L2)
- Nesterov Momentum
- Early stopping to prevent overfitting

Activation Functions

ReLU for all hidden layers and softmax
for output layer works best
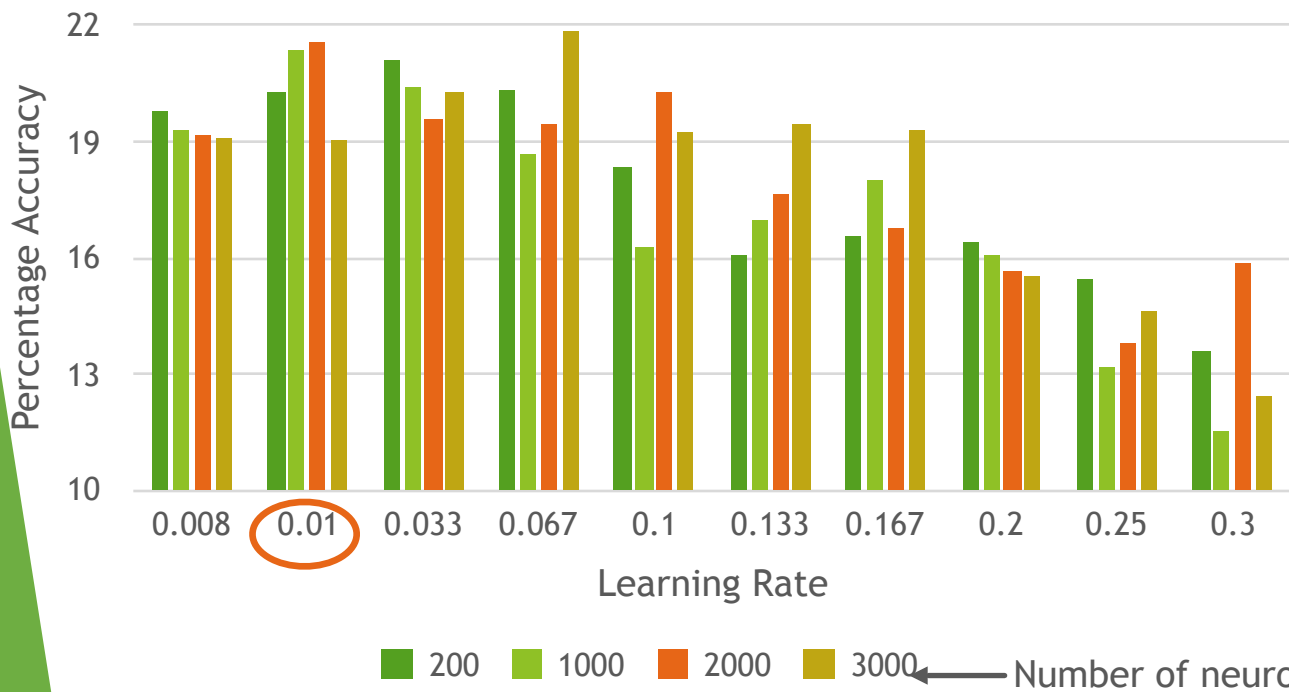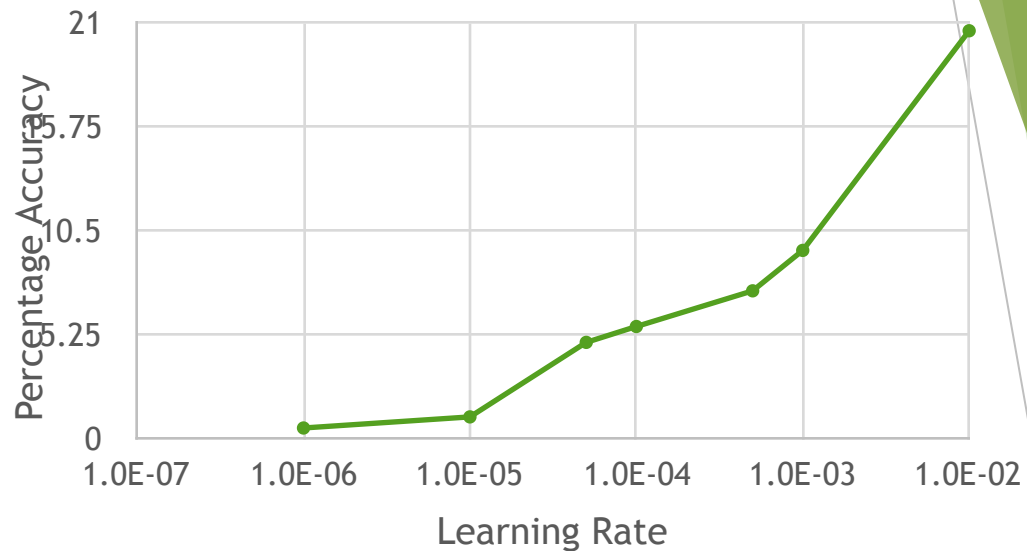
Size of 1st Hidden Layer

No real trend. Max accuracy for 3900 neurons.
I picked 2000 neurons for final model after optimizing other parameters.
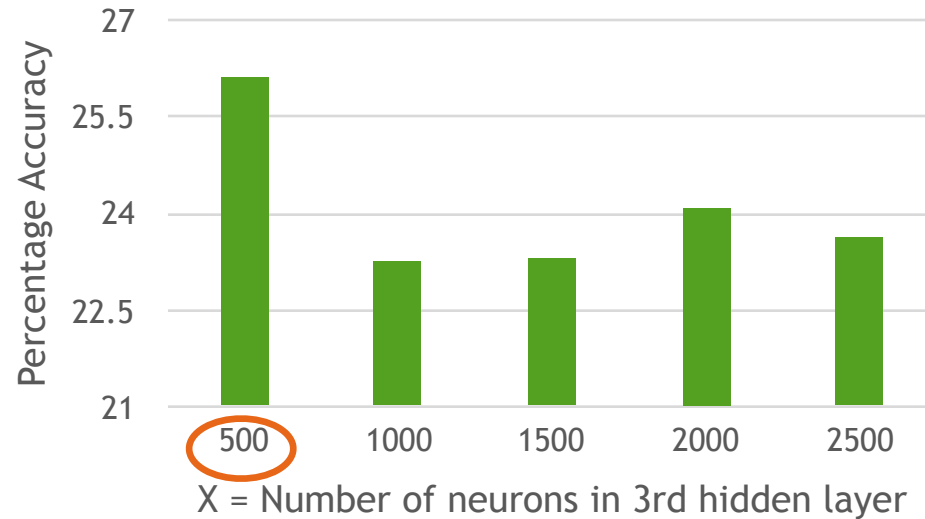
Learning rate η

Picked eta = 0.01
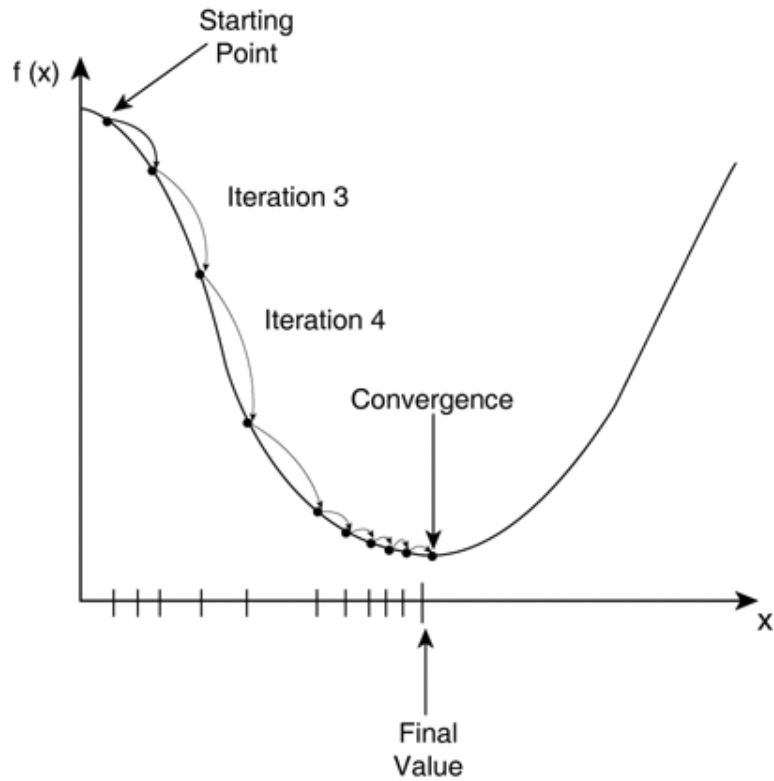
Size of subsequent hidden layers
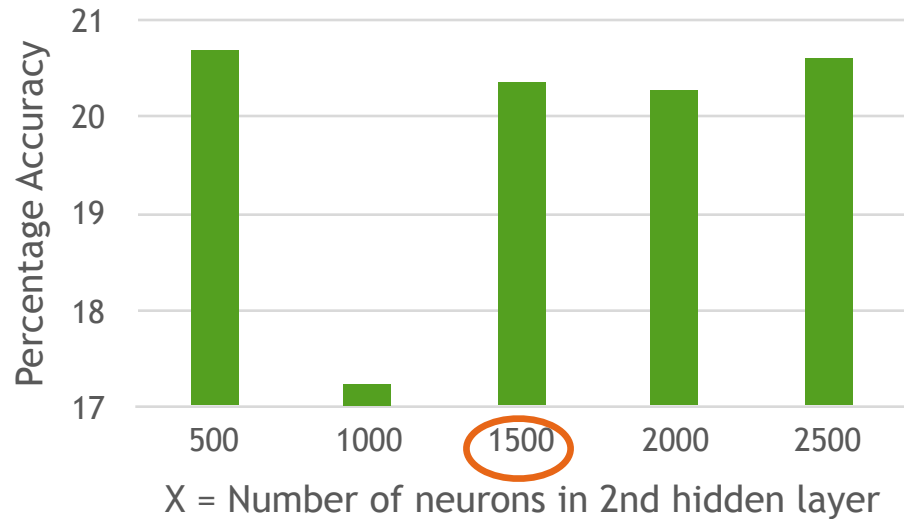
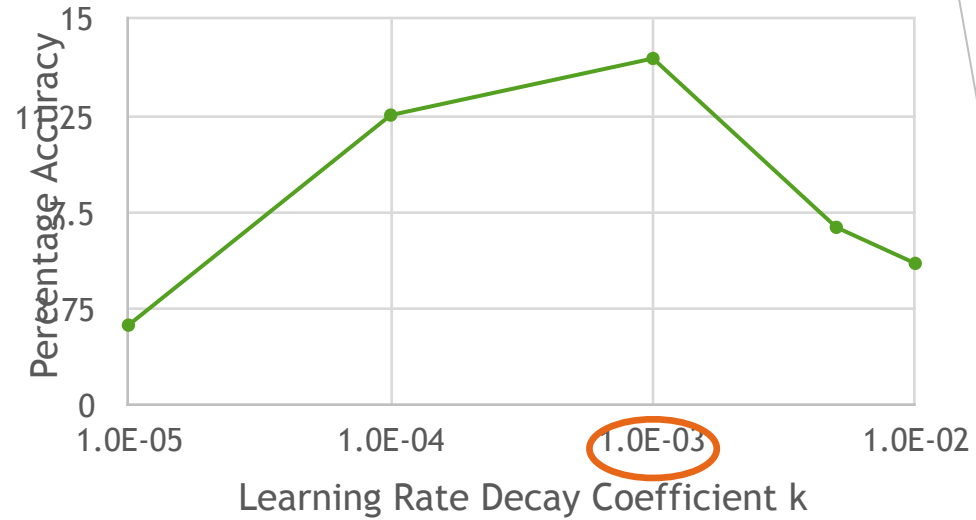Network config = [41,2000,X,119]
Picked X = 1500

Network config = [41,2000,1500,X,119]
Picked X = 500

Final network configuration = [41,2000,1500,500,119]

Learning rate decay (Annealing)

$$\eta_t = \frac{\eta_0}{1 + \boldsymbol{k}t}$$

Picked *k* = 0.001

Nesterov Momentum

Ordinary Gradient
Descent Update

$$w_{t+1} = w_t - \eta \frac{\partial(Cost)}{\partial w}\bigg|_t$$

Momentum Update

$$w_{t+1} = \left(w - \boldsymbol{\mu}\,\Delta\,w\right)_t - \eta \frac{\partial(Cost)}{\partial w}\bigg|_t$$

Nesterov Momentum Update

$$w_{t+1} = \left(w - \boldsymbol{\mu}\,\Delta\,w\right)_t - \eta \frac{\partial(Cost)}{\partial(w - \boldsymbol{\mu}\,\Delta\,w)}\bigg|_t$$

Picked Nesterov μ = 0.99

18
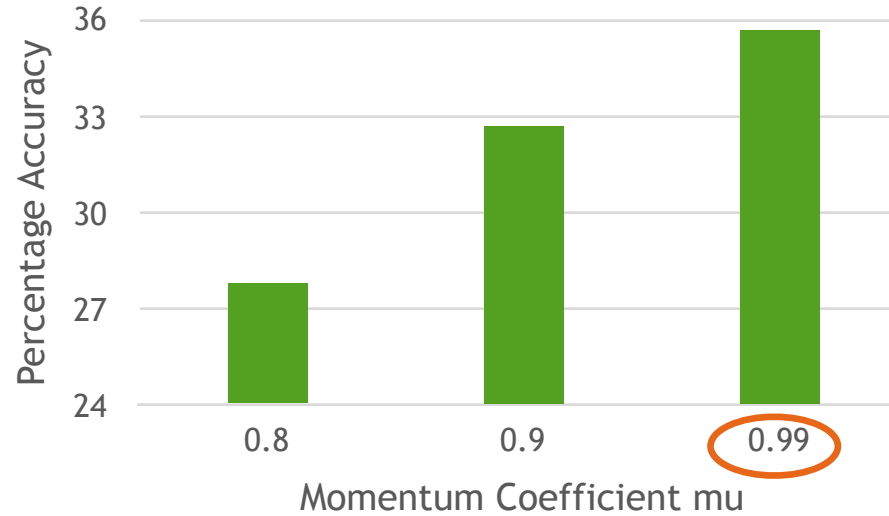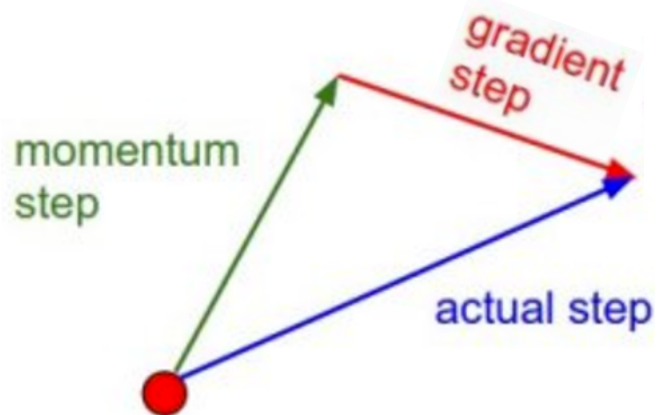
L2 Regularization

$$Cost\ +=\ \lambda \sum w^2$$

Picked λ = 5e-4 = 0.0005

Early Stopping

Prevent overfitting by stopping training when validation accuracy doesn't improve over some **threshold** number of epochs
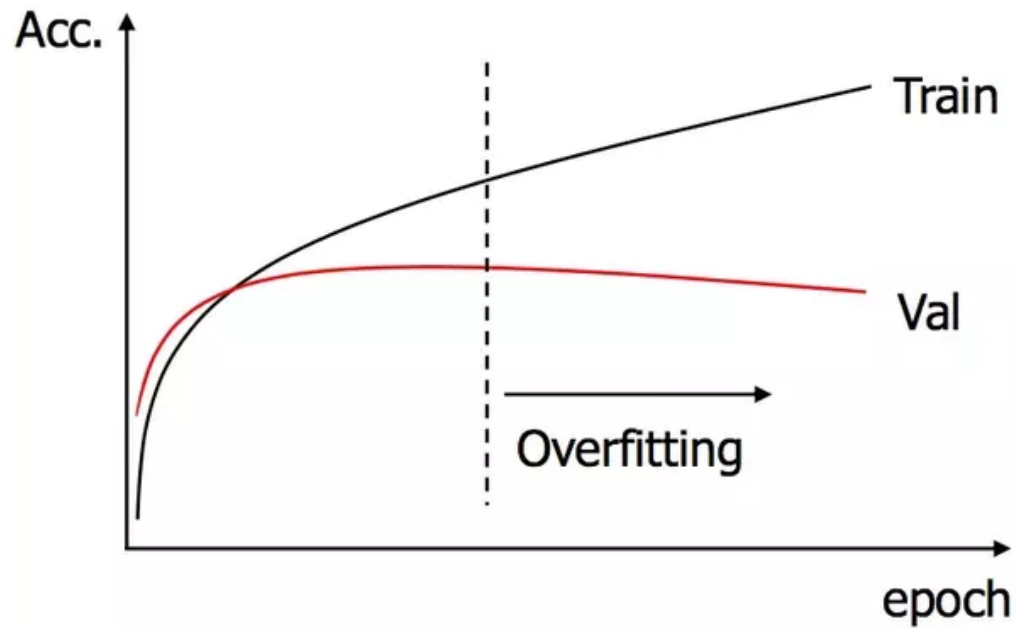
Picked threshold
= 10 epochs

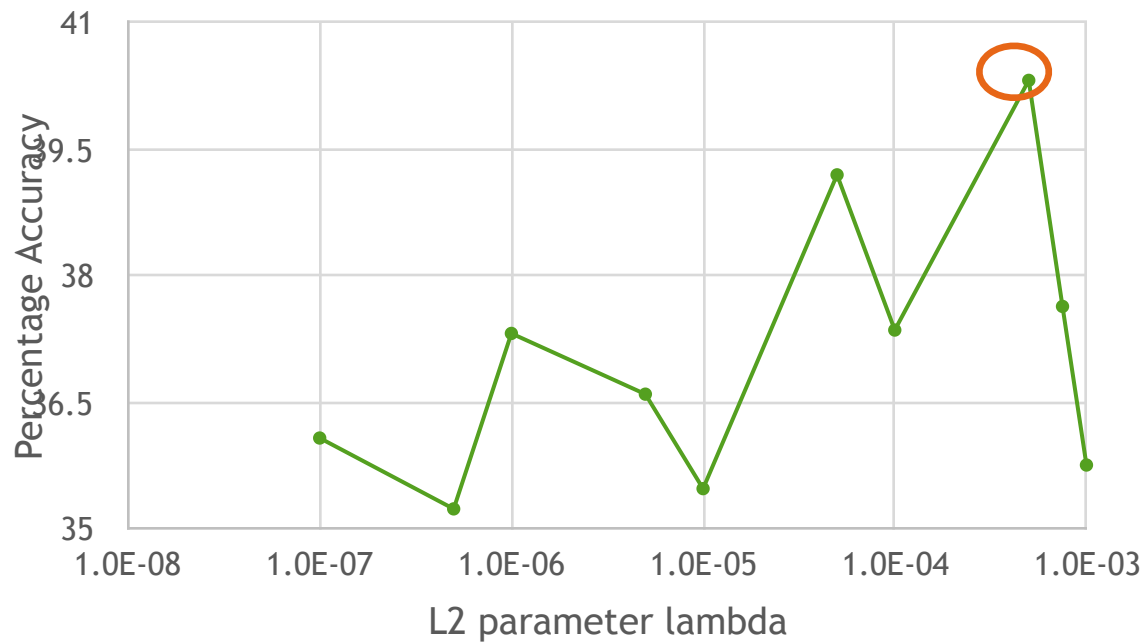# Final Training Conditions

- Network configuration = [41,2000,1500,500,119]
- ReLU activation for all hidden layers, finally softmax output
- L2 Regularization Coefficient = 0.0005
- Learning rate = 0.01, annealing coefficient = 0.001
- Nesterov momentum coefficient = 0.99
- Minibatch size = 20

Early stopping at 60 epochs

Accuracy Metrics

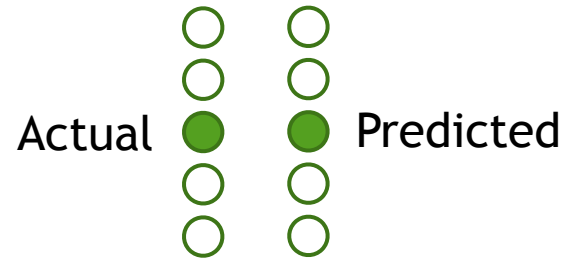Actual — Predicted

(Top-1) Accuracy: Prediction exactly matches actual price

40.32%

Actual — Predicted
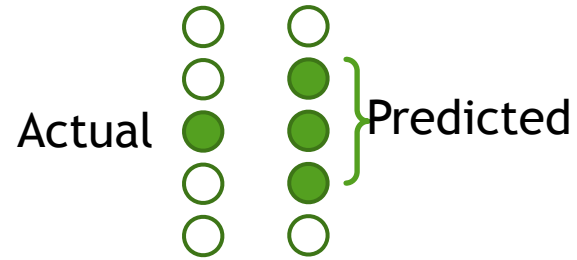
Top-3 Accuracy

71.28%

Actual — Predicted
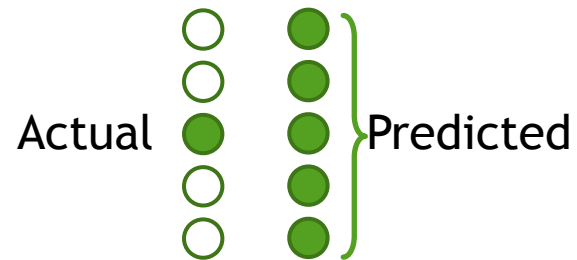
Top-5 Accuracy

87.2%

Average Percentage Error (APE) in Price

$$APE = \left(\text{Avg}_{All\ Test\ Samples} \frac{\left| True\ price - Predicted\ price \right|}{True\ price} \right) \times 100$$

APE = 6.32%

Potential Improvements

- Star outliers hard to predict
- Goalkeeping stats not considered
- Training beyond ~60 epochs doesn't improve
  - Vanishing gradients: Use different eta for each layer

Lionel Messi: $96 Million

Manuel Neuer: $75 Million

Additional Details

- Accompanying report
  - More explanations
  - References
- Code Available on Github
- Tools Used
  - Python 2.7
  - Theano
  - Keras

Ending Quote

" *If you have the money and you find the one player who can make you win and make the difference, no matter how expensive he is, you should do it. But there are not many players in the world who will make a real difference.* "

-- Arsene Wenger
Manager, Arsenal Football Club
Interview with The Times, 2009