



# Win Some, Learn Some



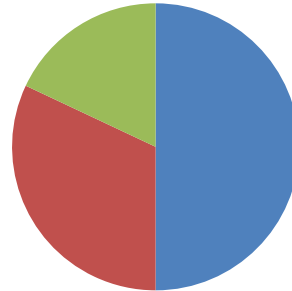
## K Means Clustering

### Our Data

Our data is taken from FightMetric LLC, a public online clearinghouse of all fight statistics collected live at every fight put on by the UFC. Each fighter is represented by the following 8 attributes, which are averaged over their career. Every pair of fighters engaged in a match make up a training example  $X^{(i)}$ .

Significant Strikes Landed per Minute	Significant Striking Accuracy (%)	Significant Strikes Absorbed per Minute	Significant Strike Defense (%)
Average Takedowns Landed per 15 Minutes	Takedown Accuracy (%)	Takedown Defense (%)	Average Submissions Attempted per 15 Minutes

### Style



Clusters	SlpM	Str. Acc	SAPM	Str. Def	TD Avg	TD Acc	TD Def	Sub Avg
Well-rounded	2.71	0.43	2.44	0.58	1.64	0.41	0.60	0.78
The "Striker"	4.20	0.43	3.89	0.59	0.88	0.38	0.64	0.49
The "Grappler"	3.46	0.48	2.37	0.58	4.40	0.54	0.66	1.08

- Well-Rounded 50%
- "Striker" 32%
- "Grappler" 18%

"Grapplers" tend to have an advantage over Well-Rounded fighters (61% victories)  
 Well-Rounded fighters tend to have an advantage over "Strikers" (62% victories)  
 "Grapplers" and "Strikers" seem more evenly matched (53% for "Grapplers")

### Classifiers

We experimented with 3 types of feature mapping from the 16 attributes of each training example in the raw data to 8 features for each training example: ratio, difference, and normalized difference

Classifier	Accuracy with $a_1/a_2$	Accuracy with $a_1 - a_2$	Accuracy with $(a_1 - a_2) / (a_1 + a_2)$
Naïve Bayes	56.89%	46.55%	53.45%
Logistic Regression	<b>69.82%</b>	67.24%	67.24%
Linear-kernel SVM $K(u,v) = u \cdot v + 1$	68.1%	68.97%	65.52%
Normalized linear-kernel SVM $K(u,v) = \frac{1}{N} u \cdot v + 1$	68.97%	67.24%	64.66%
Polynomial-kernel SVM (2nd) $K(u,v) = (u \cdot v + 1)^2$	52.59%	68.97%	67.24%
Norm. polynomial-kernel SVM (2nd) $K(u,v) = (\frac{1}{N} u \cdot v + 1)^2$	54.31%	67.24%	67.24%
Polynomial-kernel SVM (3rd) $K(u,v) = (u \cdot v + 1)^3$	47.41%	51.72%	62.07%
Norm. polynomial-kernel SVM (3rd) $K(u,v) = (\frac{1}{N} u \cdot v + 1)^3$	54.31%	65.52%	68.97%
Sigmoid-kernel SVM $K(u,v) = \tanh(u \cdot v + 1)$	49.14%	56.9%	51.72%
Norm. sigmoid-kernel SVM $K(u,v) = \tanh(\frac{1}{N} u \cdot v + 1)$	42.24%	56.9%	64.66%
Gaussian-kernel SVM $K(u,v) = e^{-(u \cdot v)^2}$	64.66%	64.66%	63.79%
Norm. Gaussian-kernel SVM $K(u,v) = e^{-(\frac{1}{N} u \cdot v)^2}$	62.93%	66.38%	65.52%

### Our Predictions (Logistic Regression)

Fight	*Return on \$100	Market Probabilities	Prediction	Prediction Probabilities
Aldo vs. McGregor	105, 80	0.43, 0.57	Aldo	0.504, 0.496
Weidman vs. Rockhold	65, 135	0.67, 0.33	Weidman	0.53, 0.47
Souza vs. Romero	67, 130	0.66, 0.34	Romero	0.49, 0.51
Maia vs. Nelson	69, 125	0.65, 0.35	Maia	0.70, 0.30
Holloway vs. Stephens	18, 425	0.96, 0.04	Holloway	0.65, 0.35
Faber vs. Saenz	13, 525	0.98, 0.02	Saenz	0.47, 0.53
Torres vs. Lybarger	36, 235	0.87, 0.13	Torres	0.95, 0.05
Alves vs. Covington	100, 83	0.45, 0.55	Alves	0.503, 0.497
Santos vs. Lee	475, 15	0.03, 0.97	Lee	0.36, 0.64
Proctor vs. Mustafaeov	290, 29	0.91, 0.09	Proctor	0.73, 0.27
Makdessi vs. Medeiros	63, 140	0.69, 0.31	Makdessi	0.56, 0.44
McGee vs. Alexandre Jr.	57, 155	0.73, 0.27	McGee	0.82, 0.18

\*betting lines taken from 5dimes

### Further Investigation

- More features (height, reach, past record/recent fights)
- Method of winning (round, decision/KO/submission)
- Running predictions on how the market picks favorites

Ivan Suarez Robles  
 Joseph Wu  
 CS 229