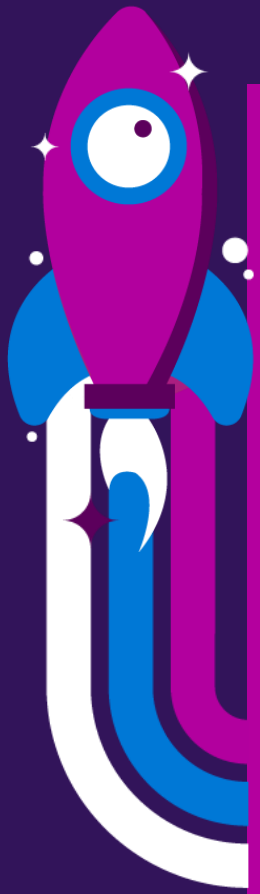
#FusionAIDay



MACHINE LEARNING FOR DEVELOPERS

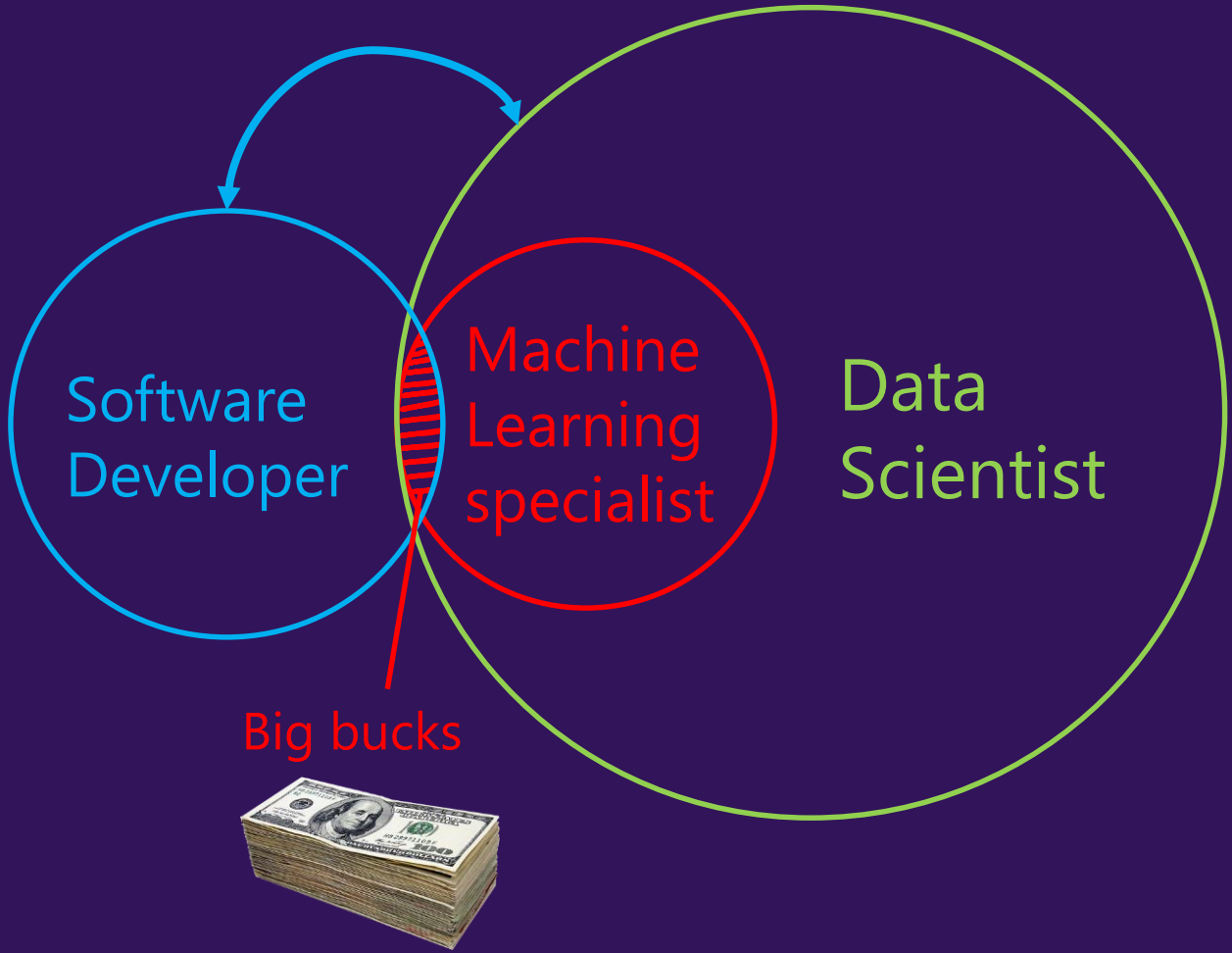
Mason Cusack, @MasonCusack

WHO ARE YOU?

Mathematician?

Data scientist?

Developer first?


Software
Developer
Machine
Learning
specialist
Data
Scientist
Big bucks

BUT I'M A DEVELOPER!

An ML development tool should:

- Forget about fine detail (if I want to)
- Embed ML in my application
- Retain power

Azure Machine Learning Studio

studio.azureml.net

Design

Train

Deploy

WHY ALL THE FUSS ABOUT MACHINE LEARNING?

#FusionAIDay

fusion

Microsoft

Collaborative filtering optimization objective

→ Given $x^{(1)}, \ldots, x^{(n_m)}$, estimate $\theta^{(1)}, \ldots, \theta^{(n_u)}$:

$$\min_{\theta^{(1)},\ldots,\theta^{(n_u)}} \frac{1}{2} \sum_{j=1}^{n_u} \sum_{i:r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)})^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

→ Given $\theta^{(1)}, \ldots, \theta^{(n_u)}$, estimate $x^{(1)}, \ldots, x^{(n_m)}$:

$$\min_{x^{(1)},\ldots,x^{(n_m)}} \frac{1}{2} \sum_{i=1}^{n_m} \sum_{j:r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)})^2 + \frac{\lambda}{2} \sum_{i=1}^{n_m} \sum_{k=1}^{n} (x_k^{(i)})^2$$

Minimizing $x^{(1)}, \ldots, x^{(n_m)}$ and $\theta^{(1)}, \ldots, \theta^{(n_u)}$ simultaneously:

$$J(x^{(1)}, \ldots, x^{(n_m)}, \theta^{(1)}, \ldots, \theta^{(n_u)}) = \frac{1}{2} \sum_{(i,j):r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)})^2 + \frac{\lambda}{2} \sum_{i=1}^{n_m} \sum_{k=1}^{n} (x_k^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

$$\min_{\substack{x^{(1)},\ldots,x^{(n_m)} \\ \theta^{(1)},\ldots,\theta^{(n_u)}}} J(x^{(1)}, \ldots, x^{(n_m)}, \theta^{(1)}, \ldots, \theta^{(n_u)})$$

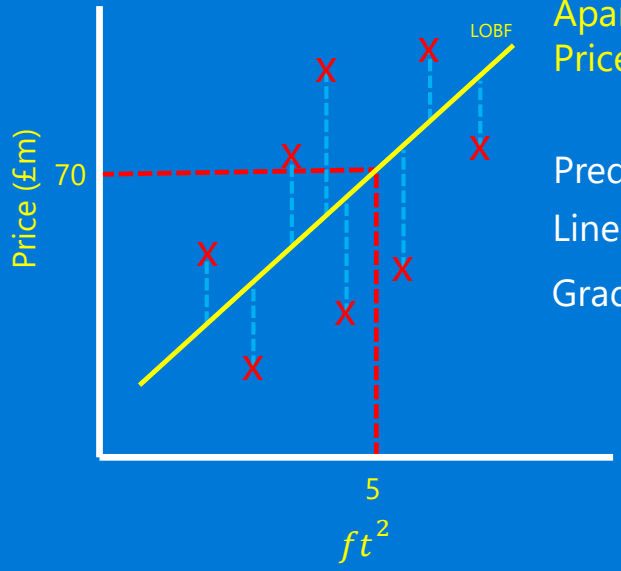
VERY HELPFUL NOT SCARY MATHS

2 minute primer

Supervised
- Classification
- Regression
- Anomaly detection

Unsupervised
- Clustering

Reinforcement
- Agent based learning



Apartment in London = $5ft^2$
Price = £70 million

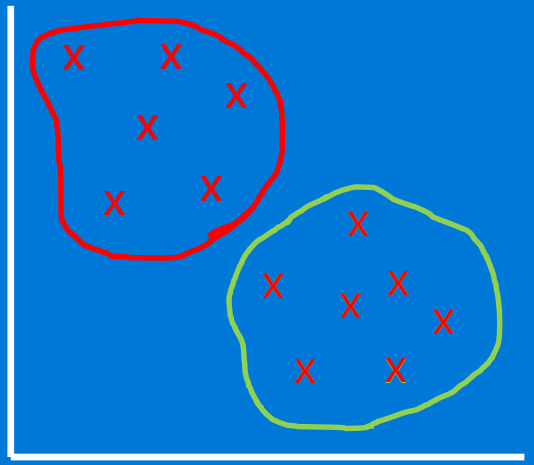
Predict the future

Linear **regression**

Gradient descent

= People who panic buy after a few inches of snow

= Sane people



Gain new insight

Unsupervised

Clustering

MACHINES ARE BETTER THAN HUMANS

#FusionAIDay

fusion

Microsoft

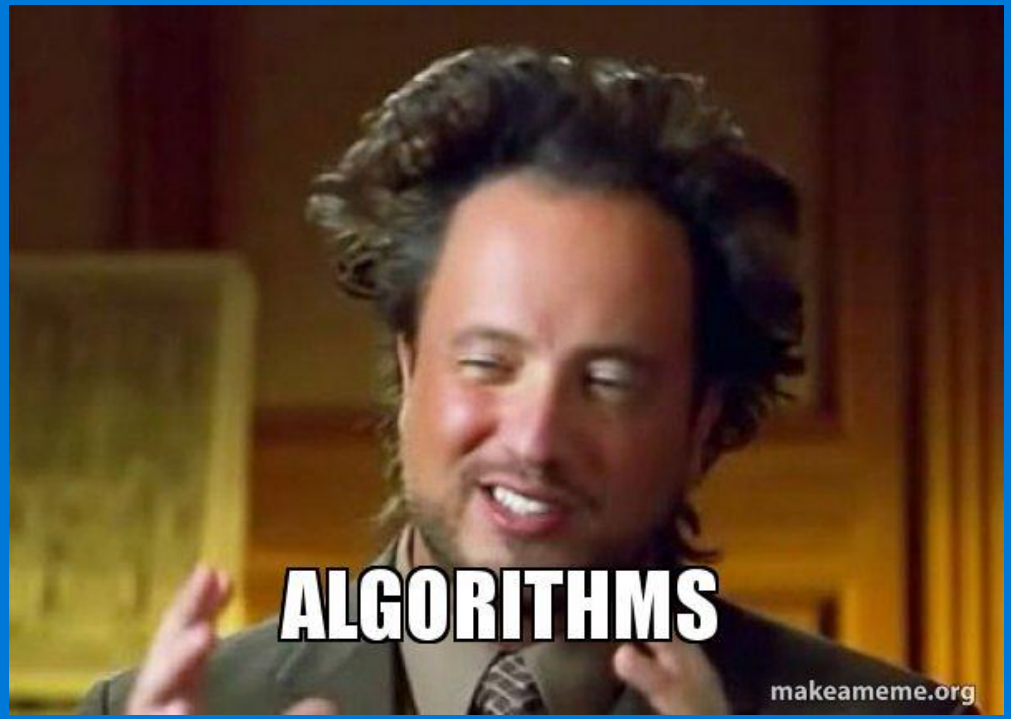
£1000+ USA RUS 20’s

Name	Amount	Issued	Used	Age	Fraudulent
Smith	£2600.45	USA	USA	22	No
Paul	£2294.58	USA	RUS	29	Yes
Peters	£1003.30	USA	RUS	25	Yes
Adams	£8488.32	FRA	USA	64	No
Pali	£200.12	AUS	JAP	58	No
Jones	£3250.11	USA	RUS	43	No
Hanford	£8156.20	USA	RUS	27	Yes
Marx	£7457.11	UK	GER	32	No
Norse	£540.00	USA	RUS	27	No
Edson	£7475.11	USA	RUS	20	Yes

WHAT ARE THE PATTERNS?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1048488	2013	6	20	4	WN	13204	12889	740	0	0	920	0	0	0
1048489	2013	6	20	4	WN	13204	13198	920	-5	0	1105	-7	0	0
1048490	2013	6	20	4	WN	13204	13232	2050	19	1	2230	9	0	0
1048491	2013	6	20	4	WN	13204	13232	655	-2	0	840	10	0	0
1048492	2013	6	20	4	WN	13204	13232	1005	8	0	1145	5	0	0
1048493	2013	6	20	4	WN	13204	13232	1655	9	0	1840	-5	0	0
1048494	2013	6	20	4	WN	13204	13232	1410	1	0	1555	-9	0	0
1048495	2013	6	20	4	WN	13204	13232	1850	25	1	2035	19	1	0
1048496	2013	6	20	4	WN	13204	13342	1400	-2	0	1550	-4	0	0
1048497	2013	6	20	4	WN	13204	13495	850	4	0	930	0	0	0
1048498	2013	6	20	4	WN	13204	13495	1105	7	0	1145	31	1	0
1048499	2013	6	20	4	WN	13204	13495	650	-2	0	725	-3	0	0
1048500	2013	6	20	4	WN	13204	13495	1725	72	1	1810	94	1	0
1048501	2013	6	20	4	WN	13204	13931	2100	5	0	2250	-6	0	0
1048502	2013	6	20	4	WN	13204	13931	1045	0	0	1240	-9	0	0
1048503	2013	6	20	4	WN	13204	14100	2045	28	1	2310	18	1	0
1048504	2013	6	20	4	WN	13204	14100	845	-1	0	1110	-15	0	0
1048505	2013	6	20	4	WN	13204	14107	1905	51	1	2025	40	1	0
1048506	2013	6	20	4	WN	13204	14122	910	-6	0	1125	-4	0	0
1048507	2013	6	20	4	WN	13204	14122	2020	28	1	2230	32	1	0
1048508	2013	6	20	4	WN	13204	14307	2130	25	1	15	10	0	0
1048509	2013	6	20	4	WN	13204	14307	1010	6	0	1250	4	0	0
1048510	2013	6	20	4	WN	13204	14307	850	3	0	1135	-12	0	0
1048511	2013	6	20	4	WN	13204	14307	1425	1	0	1710	7	0	0
1048512	2013	6	20	4	WN	13204	14492	1030	-3	0	1205	-6	0	0
1048513	2013	6	20	4	WN	13204	14683	1750	84	1	1935	78	1	0
1048514	2013	6	20	4	WN	13204	14683	1155	6	0	1340	12	0	0
1048515	2013	6	20	4	WN	13204	14730	1700	89	1	1905	103	1	0
1048516	2013	6	20	4	WN	13204	14843	2105	14	0	2355	6	0	0
1048517	2013	6	20	4	WN	13204	14843	1325	99	1	1615	88	1	0
1048518	2013	6	20	4	WN	13204	14843	650	1	0	935	-1	0	0
1048519	2013	6	20	4	WN	13204	14843	1115	3	0	1405	-3	0	0
1048520	2013	6	20	4	WN	13204	15016	1830	22	1	1950	55	1	0
1048521	2013	6	20	4	WN	13204	15016	1005	16	1	1130	10	0	0
1048522	2013	6	20	4	WN	13232	10140	1000	7	0	1155	-1	0	0

HOW?



repeat until convergence{

$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \ldots, \theta_1)$ or for short $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$

} (simultaneously update for every j = 0, …, n)

What's that?

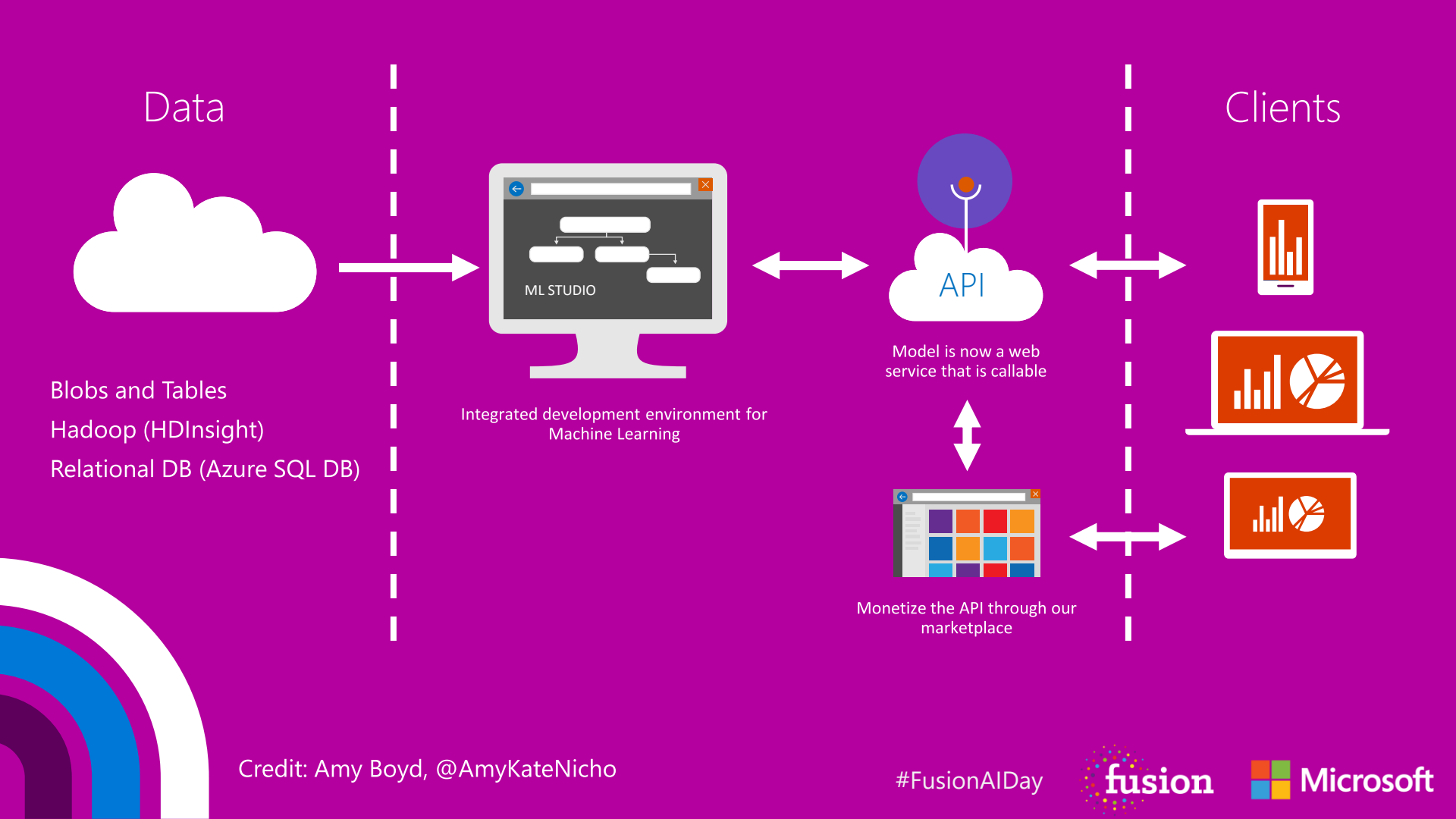
Translation:

```
for(j = 0; j<n; j++){

        theta[j] = Math.something();
};
```

GRADIENT DESCENT FOR MULTIVARIATE LINEAR REGRESSION

DON’T BELIEVE THE HYPE



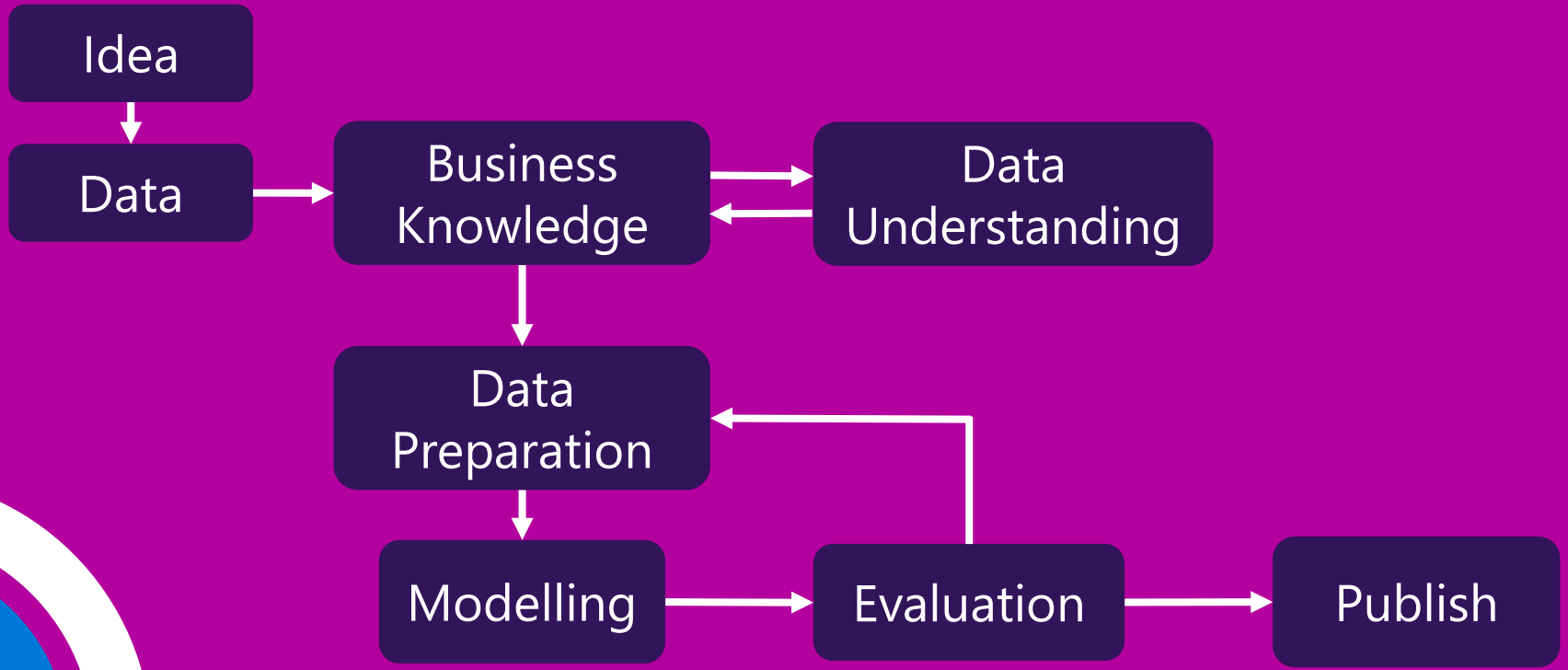
Credit: Amy Boyd, @AmyKateNicho

#FusionAIDay

fusion

Microsoft

AZURE ML PROCESS MODEL



Credit: Amy Boyd, @AmyKateNicho

ANYWAY, ENOUGH TALK

Azure Machine Learning Studio

Aka.ms/AIWorkshopML

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.

START

Finding unusual data points

Discovering structure

Three or more

Predicting categories

Predicting values

Two

ANOMALY DETECTION

- **One-class SVM** — >100 features, aggressive boundary
- **PCA-based anomaly detection** — Fast training

CLUSTERING

- **K-means**

MULTI-CLASS CLASSIFICATION

- Fast training, linear model — **Multiclass logistic regression**
- Accuracy, long training times — **Multiclass neural network**
- Accuracy, fast training — **Multiclass decision forest**
- Accuracy, small memory footprint — **Multiclass decision jungle**
- Depends on the two-class classifier, see notes below — **One-v-all multiclass**

REGRESSION

- **Ordinal regression** — Data in rank ordered categories
- **Poisson regression** — Predicting event counts
- **Fast forest quantile regression** — Predicting a distribution
- **Linear regression** — Fast training, linear model
- **Bayesian linear regression** — Linear model, small data sets
- **Neural network regression** — Accuracy, long training time
- **Decision forest regression** — Accuracy, fast training
- **Boosted decision tree regression** — Accuracy, fast training, large memory footprint

TWO-CLASS CLASSIFICATION

- **Two-class SVM** — >100 features, linear model
- **Two-class averaged perceptron** — Fast training, linear model
- **Two-class logistic regression** — Fast training, linear model
- **Two-class Bayes point machine** — Fast training, linear model

- Accuracy, fast training — **Two-class decision forest**
- Accuracy, fast training, large memory footprint — **Two-class boosted decision tree**
- Accuracy, small memory footprint — **Two-class decision jungle**
- >100 features — **Two-class locally deep SVM**
- Accuracy, long training times — **Two-class neural network**

Created by the Azure Machine Learning Team

Email: AzurePoster@microsoft.com

Download this poster: http://aka.ms/MLCheatSheet

Microsoft