The concluding session

Which ML algorithm to use?

- The size, quality, and nature of data.
- The available computational time
- The urgency of the task.
- What you want to do with the data.
- How the math of the algorithm was translated into instructions for the computer you are using.

Even an experienced data scientist cannot tell which algorithm will perform the best before trying different algorithms.

One out of Supervised Algorithms?

- Supervised learning algorithms make predictions based on a set of examples.
- Supervised learning is a popular and useful type of machine learning.
- When the data are being used to predict a category, supervised learning is also called classification.
- When a value is being predicted, as with stock prices, supervised learning is called regression.

One out of Unsupervised Algorithms?

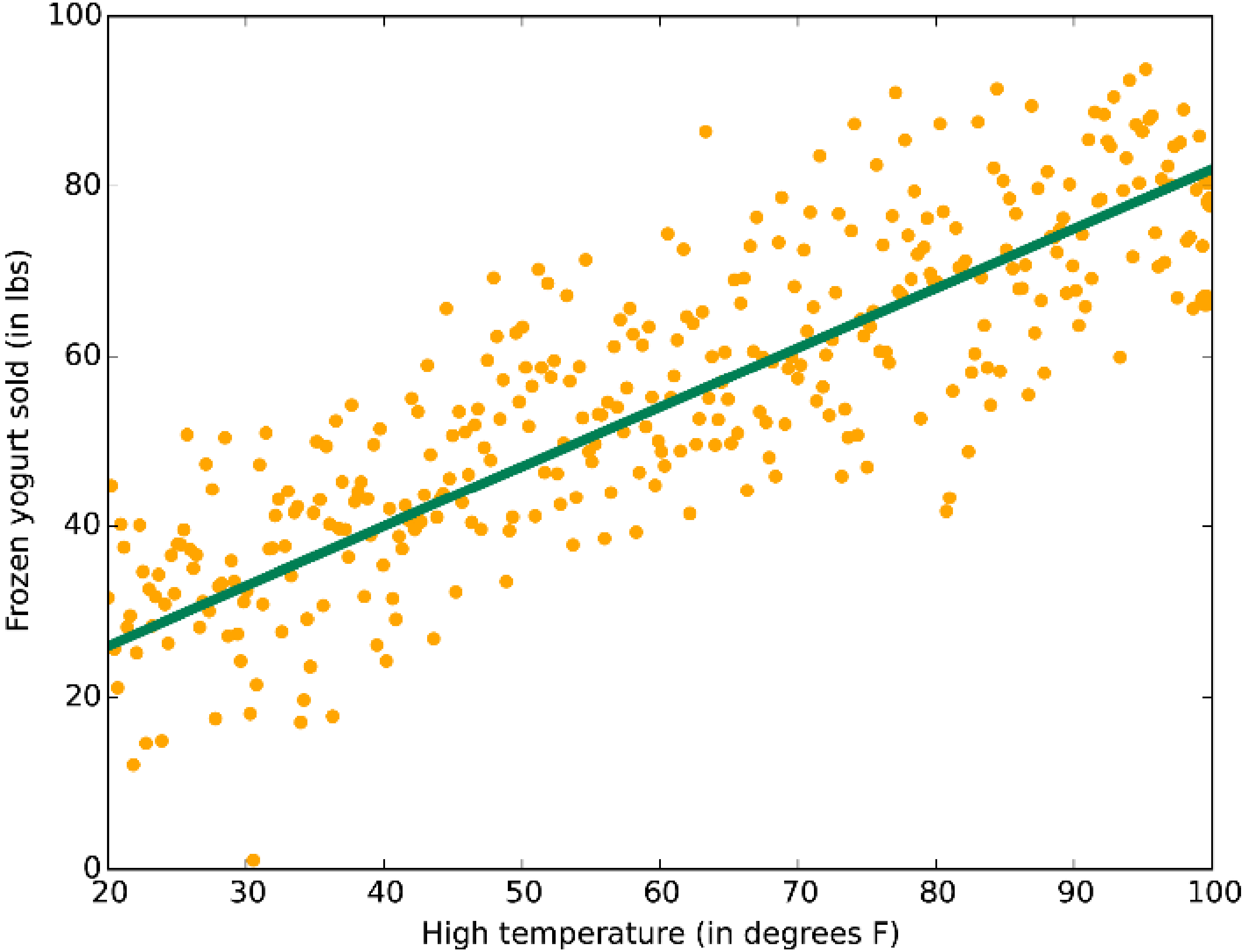
- In unsupervised learning, data points have no labels associated with them.
- Instead, the goal of an unsupervised learning algorithm is to organize the data in some way or to describe its structure. Grouping into clusters.
- Finding different ways of looking at complex data so that it appears simpler or more organized. (Dimension reduction)

OR, Reinforcement Learning?

- Trying different scenarios to discover which actions yield the greatest reward, rather than being told which actions to take.
- In reinforcement learning, the algorithm gets to choose an action in response to each data point.
- Algorithm receives a reward signal a short time later, indicating how good the decision was.
- Based on this, the algorithm modifies its strategy in order to achieve the highest reward.
- Reinforcement learning is common in robotics
- Also a natural fit for Internet of Things applications.

Let us revisit some algorithms

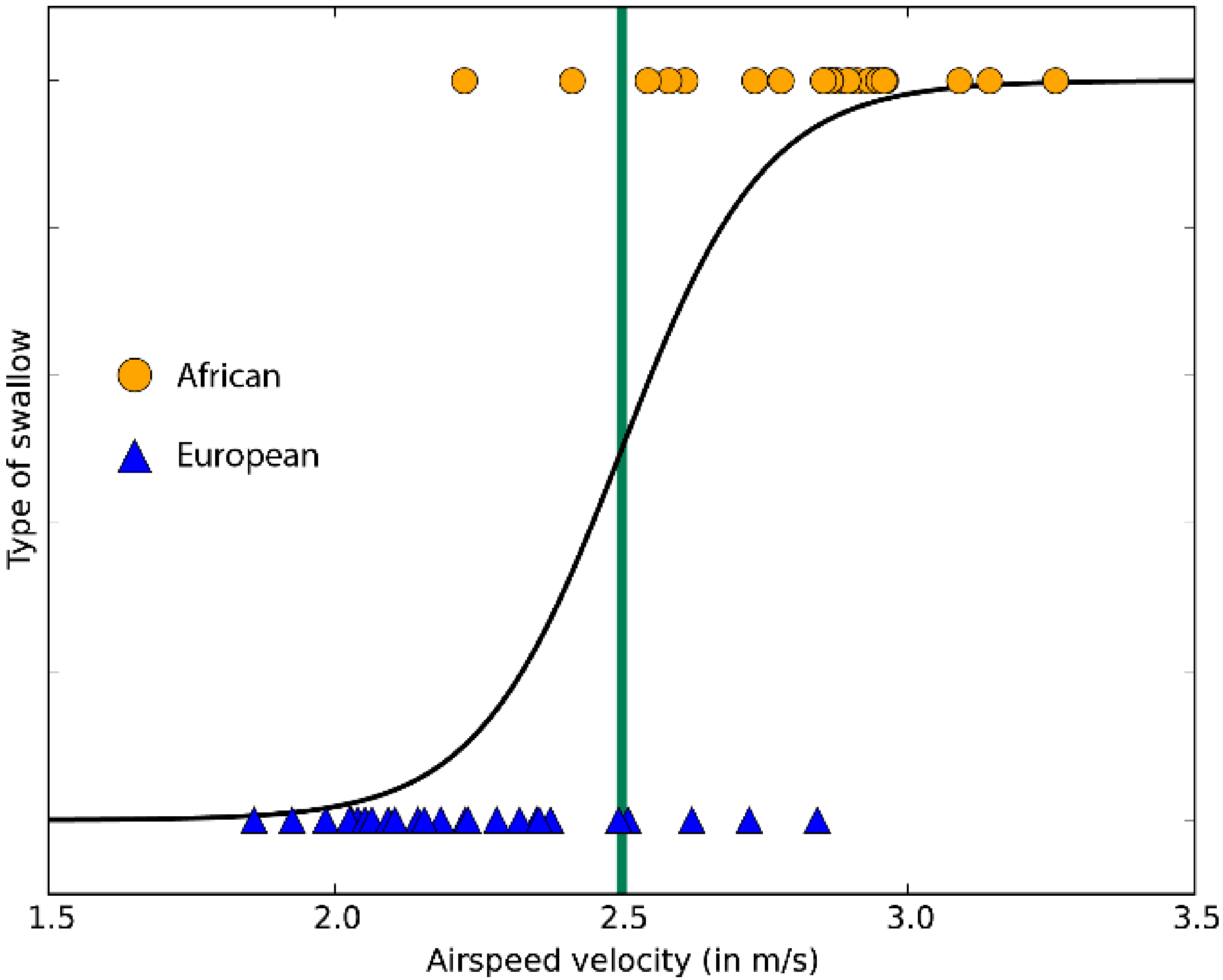
Linear Regression



Good Data with a linear trend

It's a workhorse, simple and fast, but it may be overly simplistic for some problems.

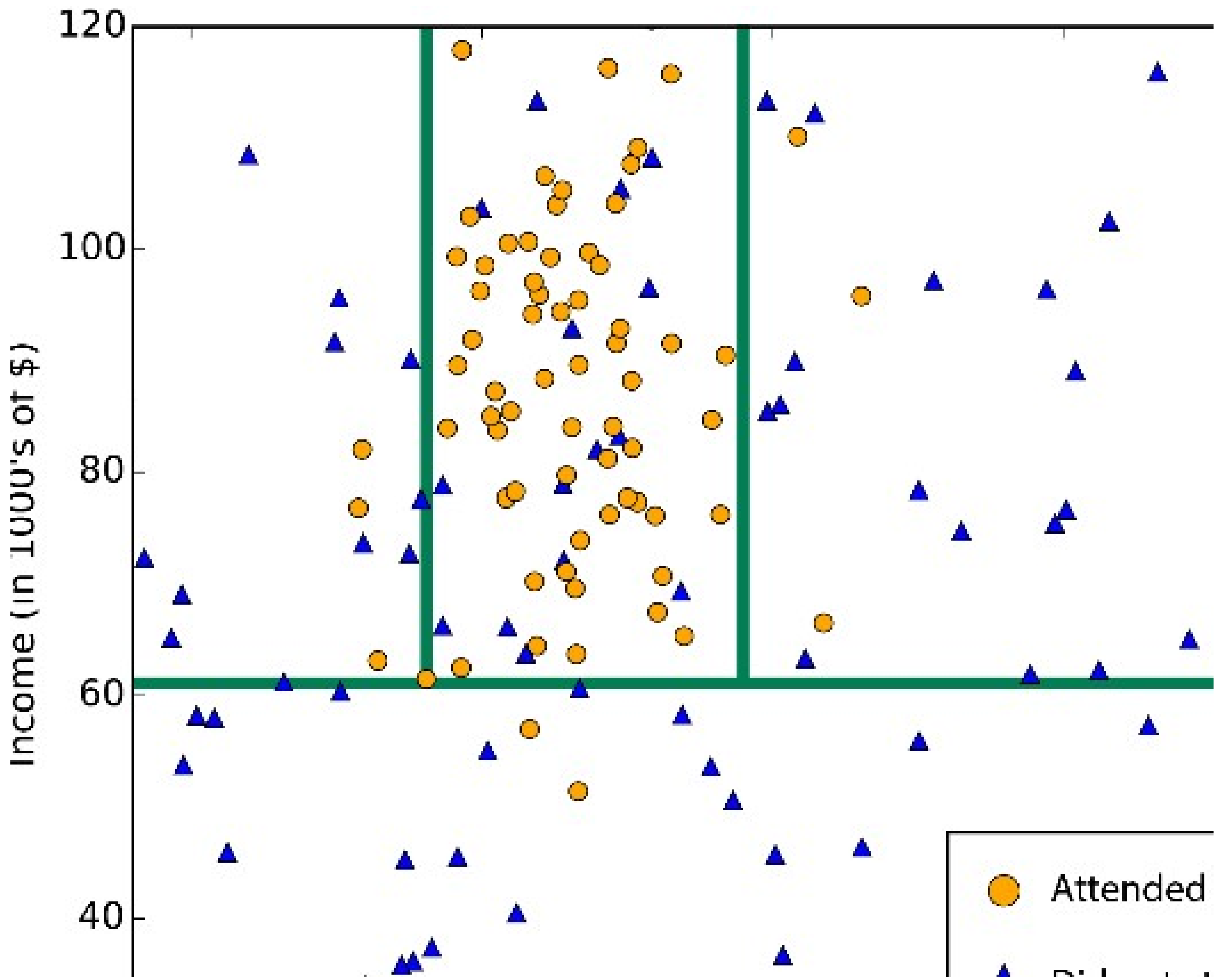
Logistic Regression



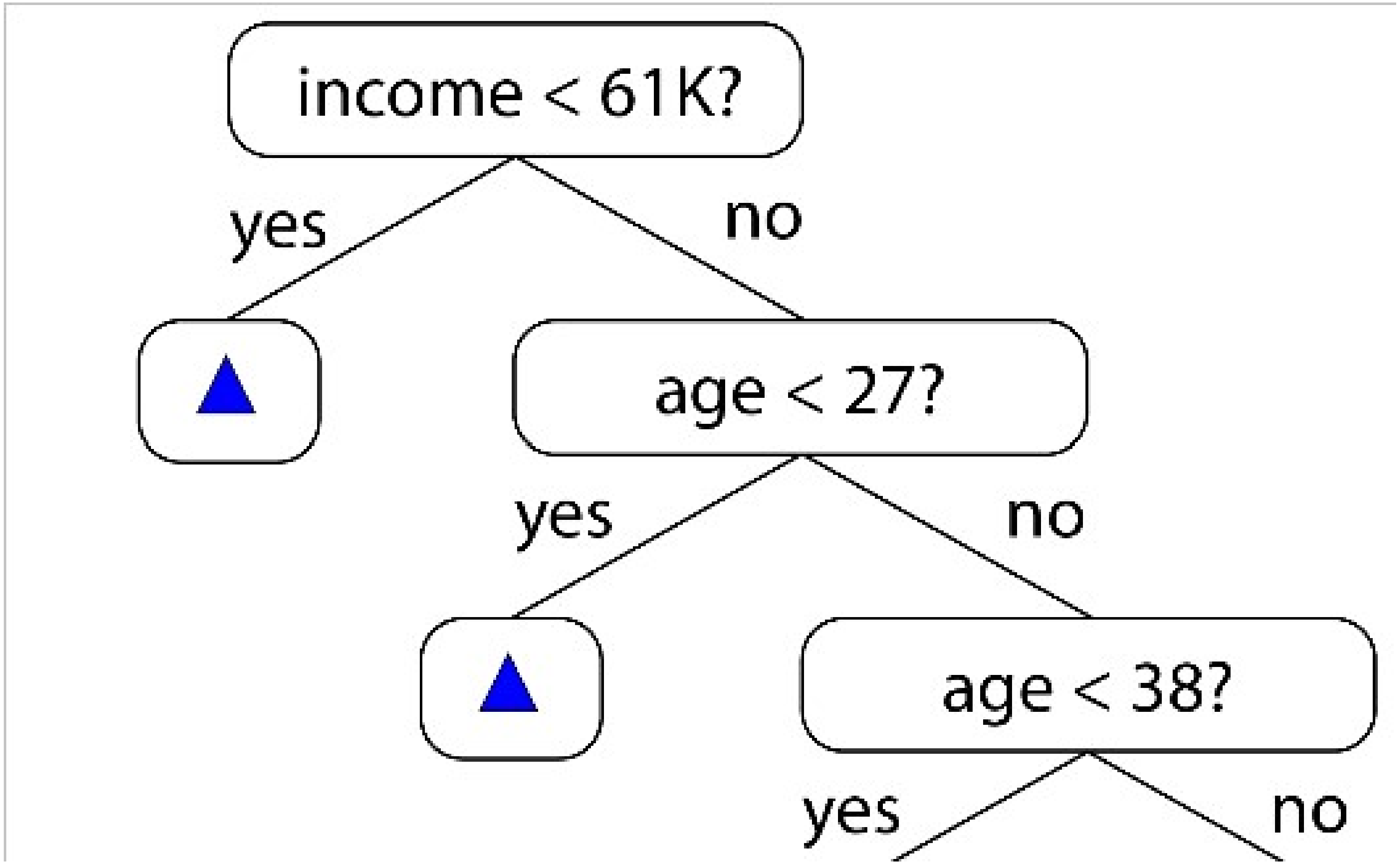
***A logistic regression to two-class data with just one feature** - the class boundary is the point at which the logistic curve is just as close to both classes*

Decision Trees

- Decision tree, forests (regression, two-class, and multiclass), and boosted decision trees are all based on decision trees, a foundational machine learning concept.
- Many variants of decision trees, but they all do the same thing—subdivide the feature space into regions with mostly the same label.
- These can be regions of consistent category or of constant value, depending on whether you are doing classification or regression.



120
100
80
60
40
Income (in 1000's of $)
Attended

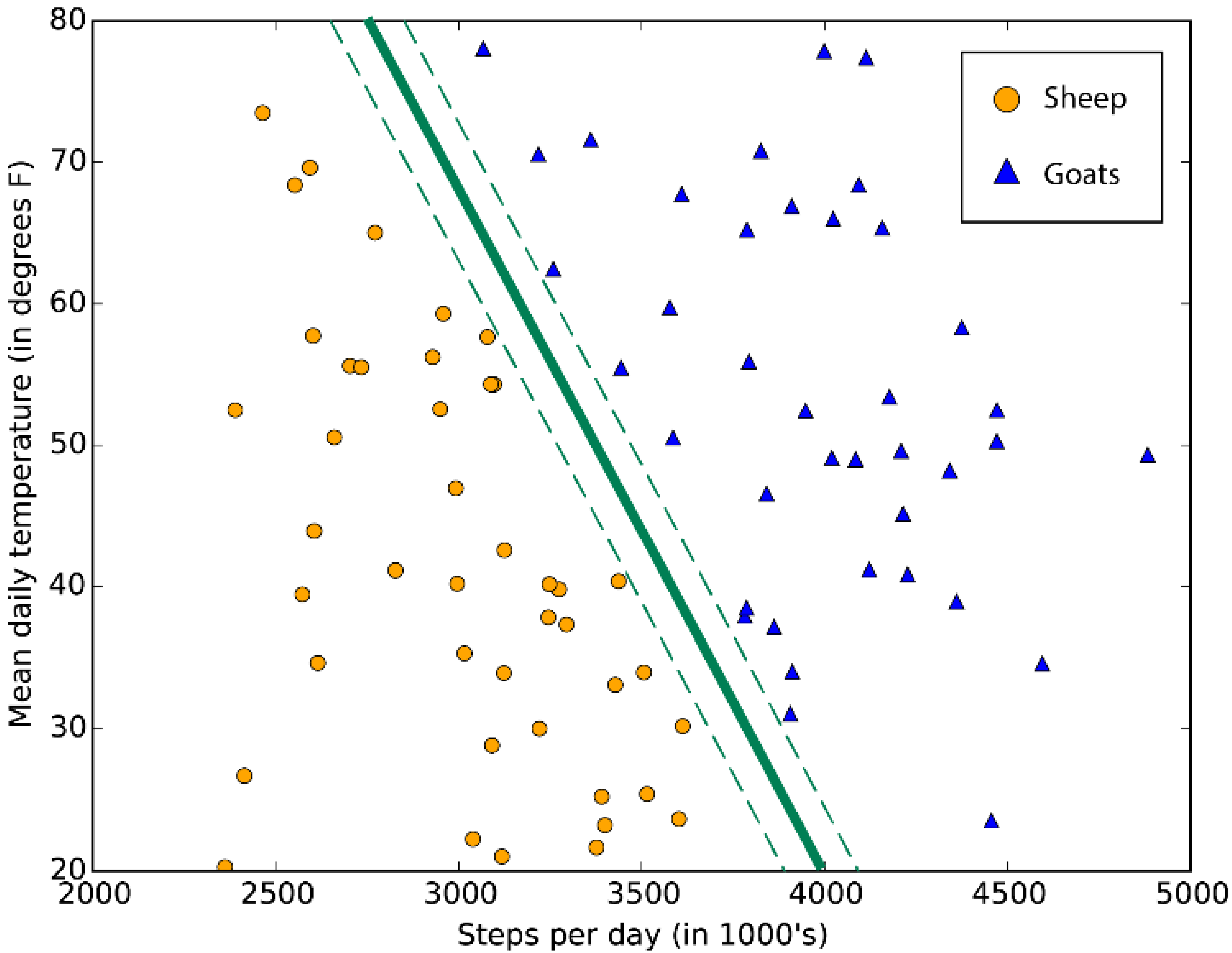

income < 61K?
yes
no
age < 27?
yes
no
age < 38?
yes
no

Decision trees…

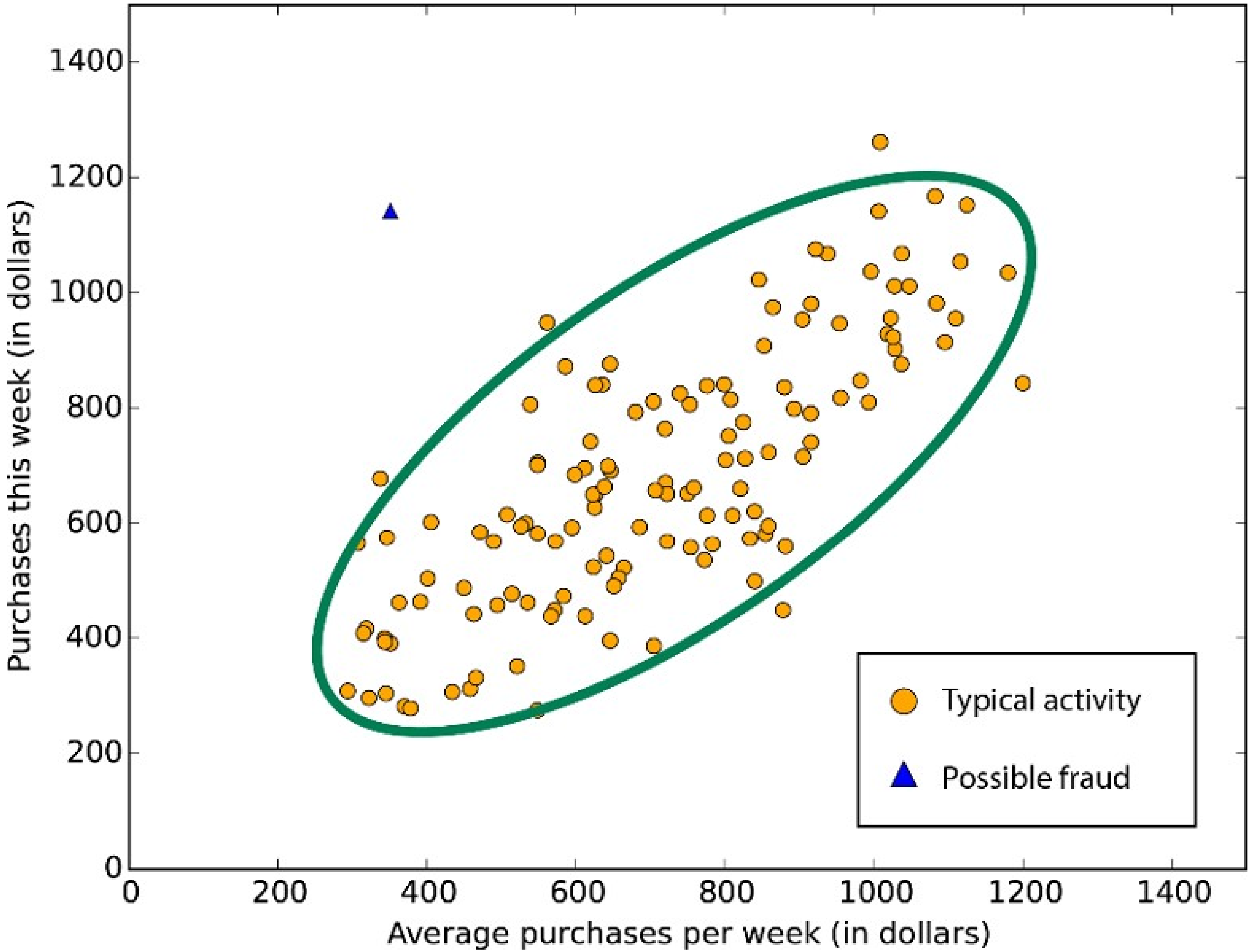
- Boosted decision trees avoid Overfitting by limiting how many times they can subdivide and how few data points are allowed in each region.
- In order to avoid Overfitting, a large set of trees are constructed with special mathematical care taken that the trees are not correlated.

SVM

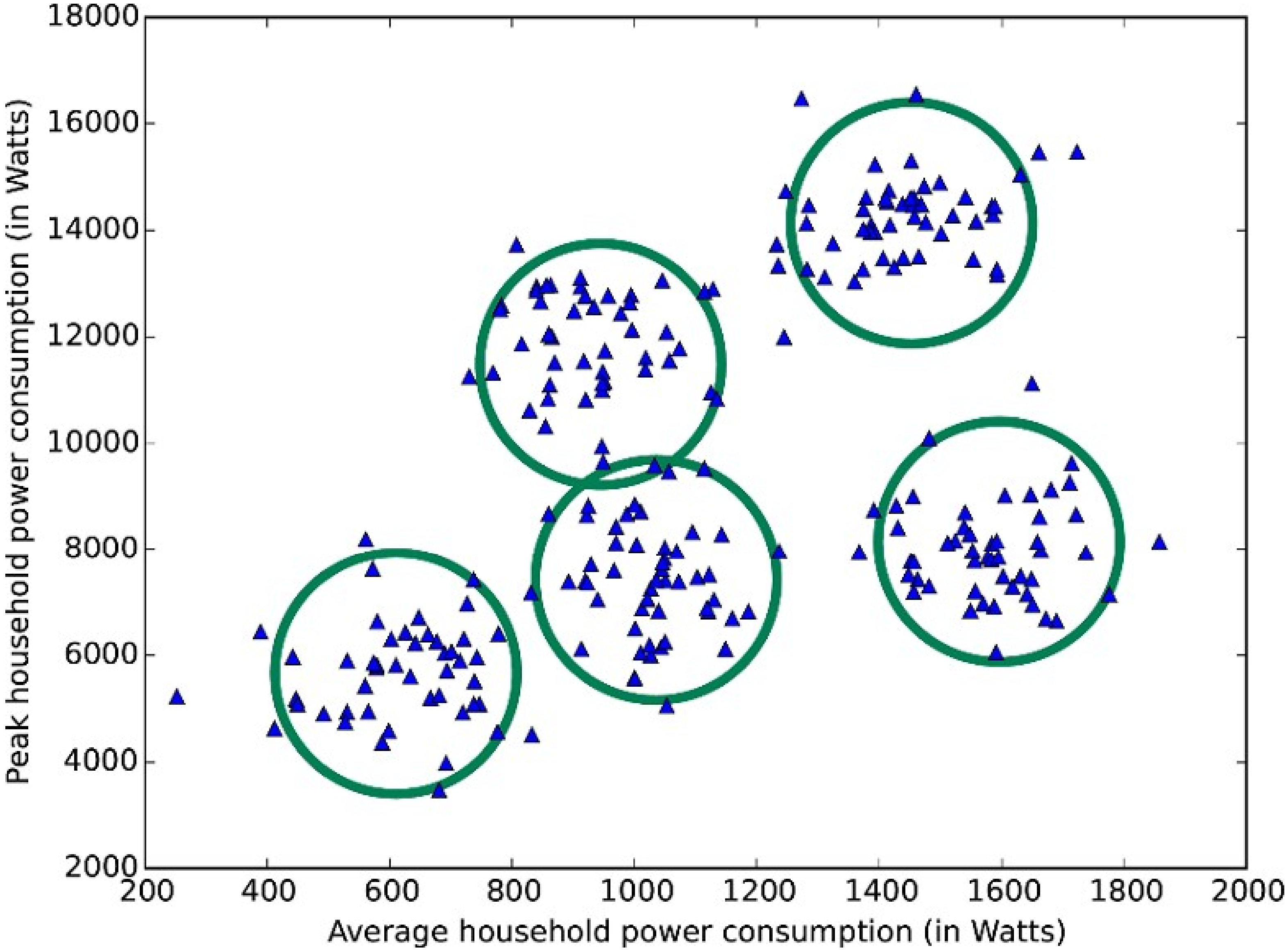
- Support vector machines (SVMs) find the boundary that separates classes by as wide a margin as possible.
- When the two classes can't be clearly separated, the algorithms find the best boundary they can.
- Because it makes this linear approximation, it is able to run fairly quickly. Where it really shines is with feature-intense data, like text or genomic.
- In these cases SVMs are able to separate classes more quickly and with less overfitting than most other algorithms, in addition to requiring only a modest amount of memory.



A typical support vector machine class boundary maximizes the margin separating two classes



PCA-based anomaly detection *- the vast majority of the data falls into a stereotypical distribution; points deviating dramatically from that distribution are suspect*



A data set is grouped into five clusters using K-means

Neural Networks

- Neural networks flourished in the mid-1980s due to their parallel and distributed processing ability.
- Were impeded by the ineffectiveness of the back-propagation training algorithm that is widely used to optimize the parameters of neural networks.
- Support vector machines (SVM) and other simpler models, which can be easily trained by solving convex optimization problems, gradually replaced neural networks in machine learning.

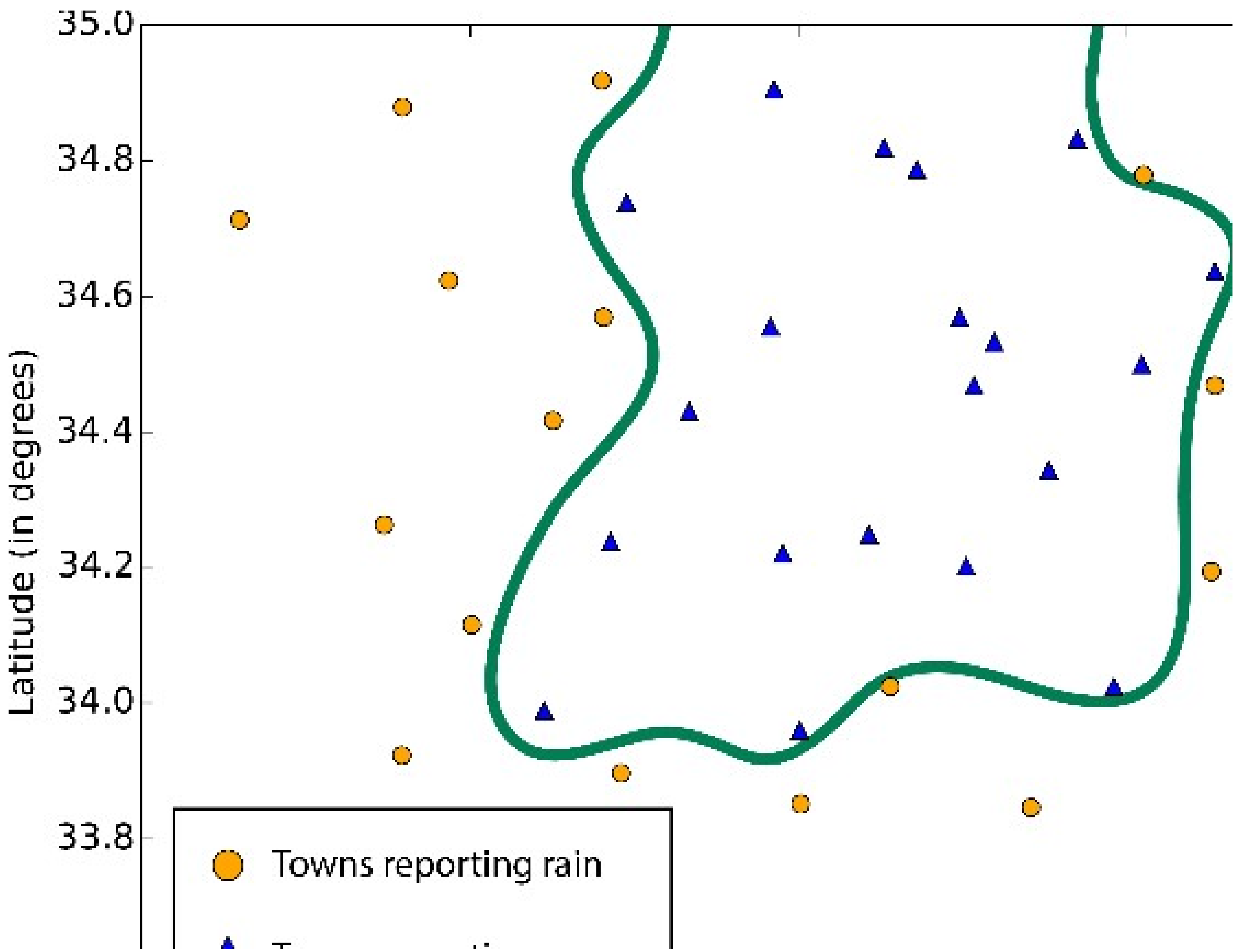
But…

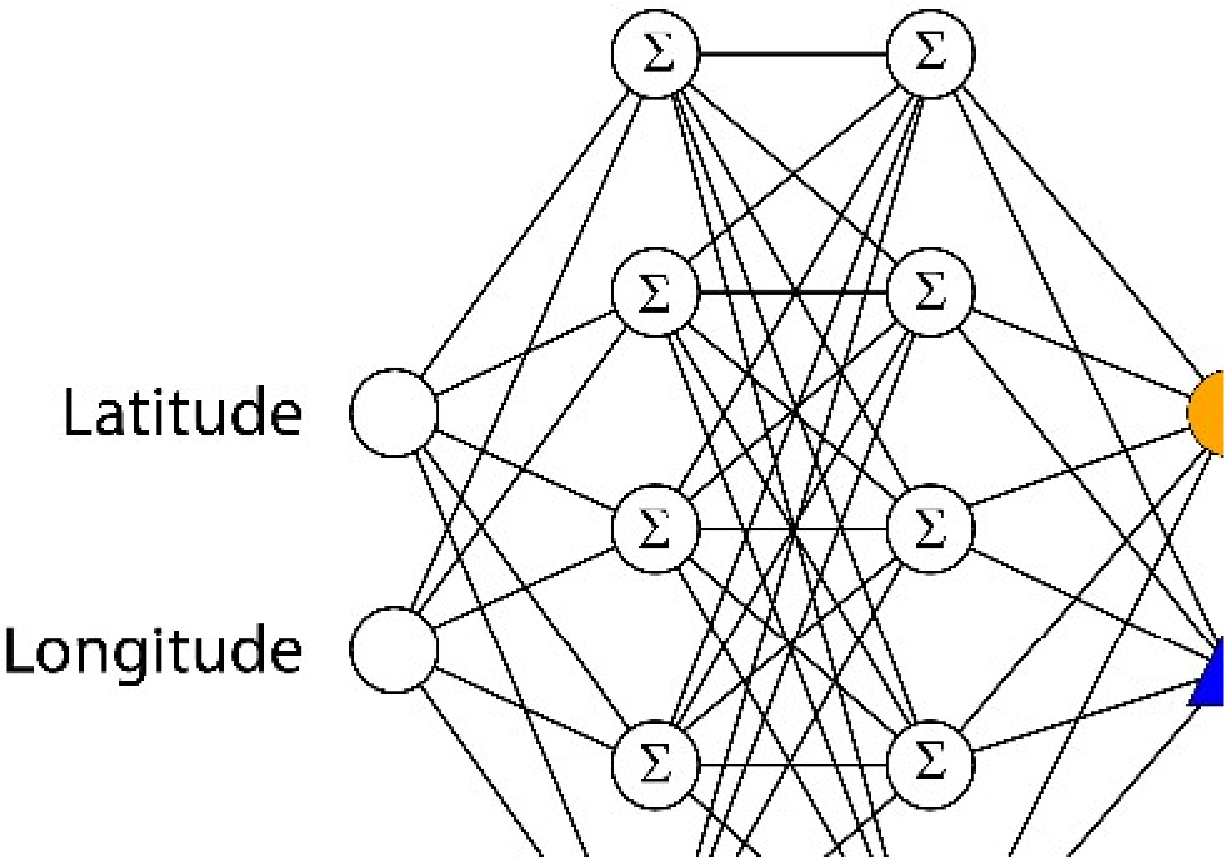
Neural Networks

- In recent years, new and improved training techniques layer-wise greedy training have led to a resurgence of interest in neural networks.
- Increasingly powerful computational capabilities, such as graphical processing unit (GPU) and massively parallel processing (MPP), have also spurred the revived adoption of neural networks.
- The resurgent research in neural networks has given rise to the invention of models with thousands of layers.

Neural Networks

- Neural networks have evolved into deep learning neural networks.
- Deep neural networks have been very successful for supervised learning.
- When used for speech and image recognition, deep learning performs as well as, or even better than, humans.
- Applied to unsupervised learning tasks, such as feature extraction, deep learning also extracts features from raw images or speech with much less human intervention.


35.0
34.8
34.6
34.4
34.2
34.0
33.8
Latitude (in degrees)
Towns reporting rain



The boundaries learned by neural networks can be complex and irregular:
The two-class averaged perceptron is neural networks' answer to skyrocketing training times. It uses a network structure that gives linear class boundaries. It is almost primitive by today's standards, but it has a long history of working robustly and is small enough to learn quickly.

Neural Networks…

- Neural networks are brain-inspired learning algorithms covering multiclass, two-class, and regression problems.
- This combination of simple calculations results in the ability to learn sophisticated class boundaries and data trends, seemingly by magic.

Choosing Algorithm

- Accuracy
- Training time
- Linearity
- Number of parameters
- Number of features
- Special cases
- Ease of use

Let us discuss these factors

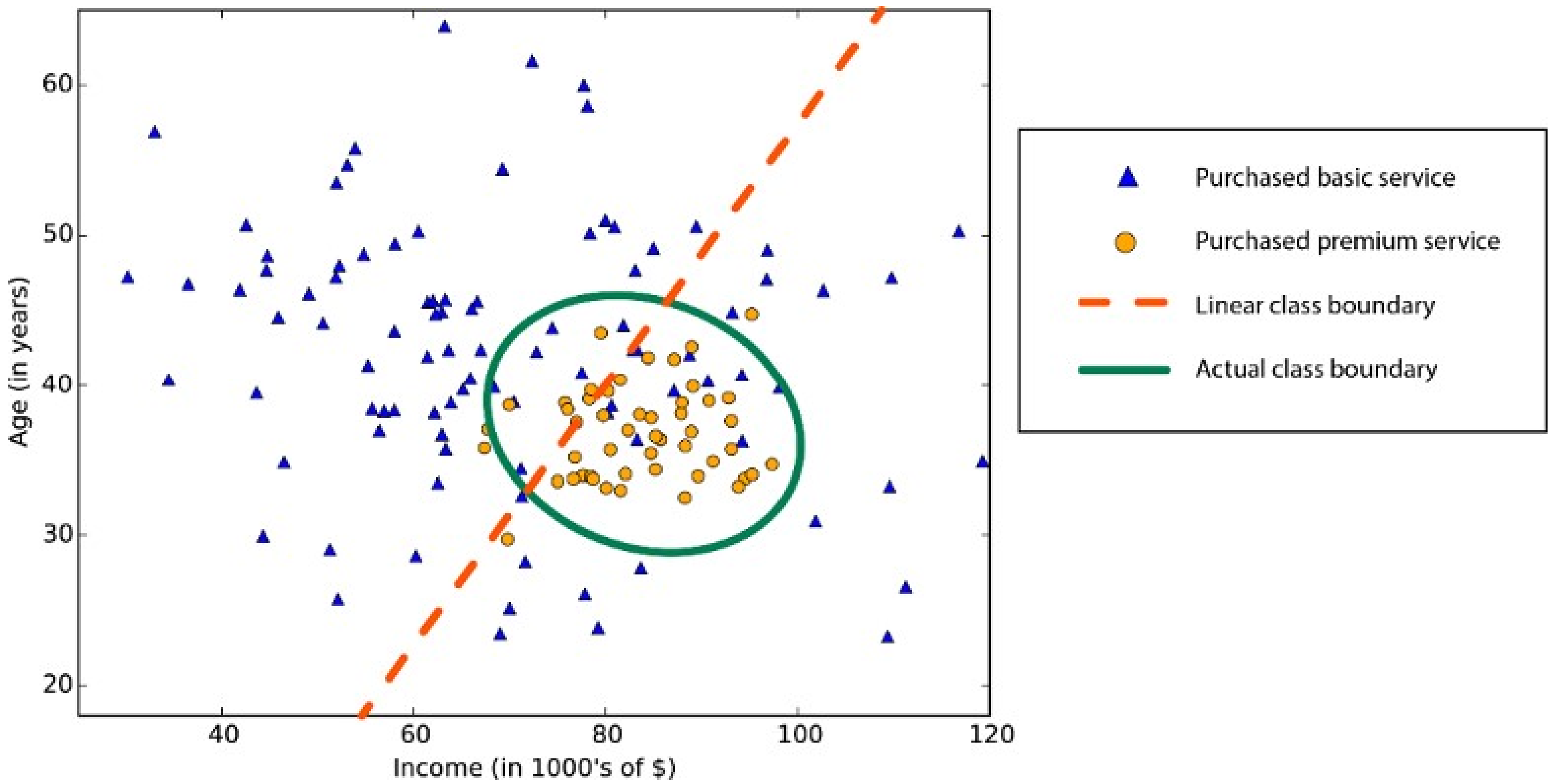
Accuracy

- Many users give the accuracy the most important and priority
- Beginners tend to focus on algorithms they know best.
- Try multiple quick algorithms as dry run on dataset to strengthen your understanding of the data

Linearity

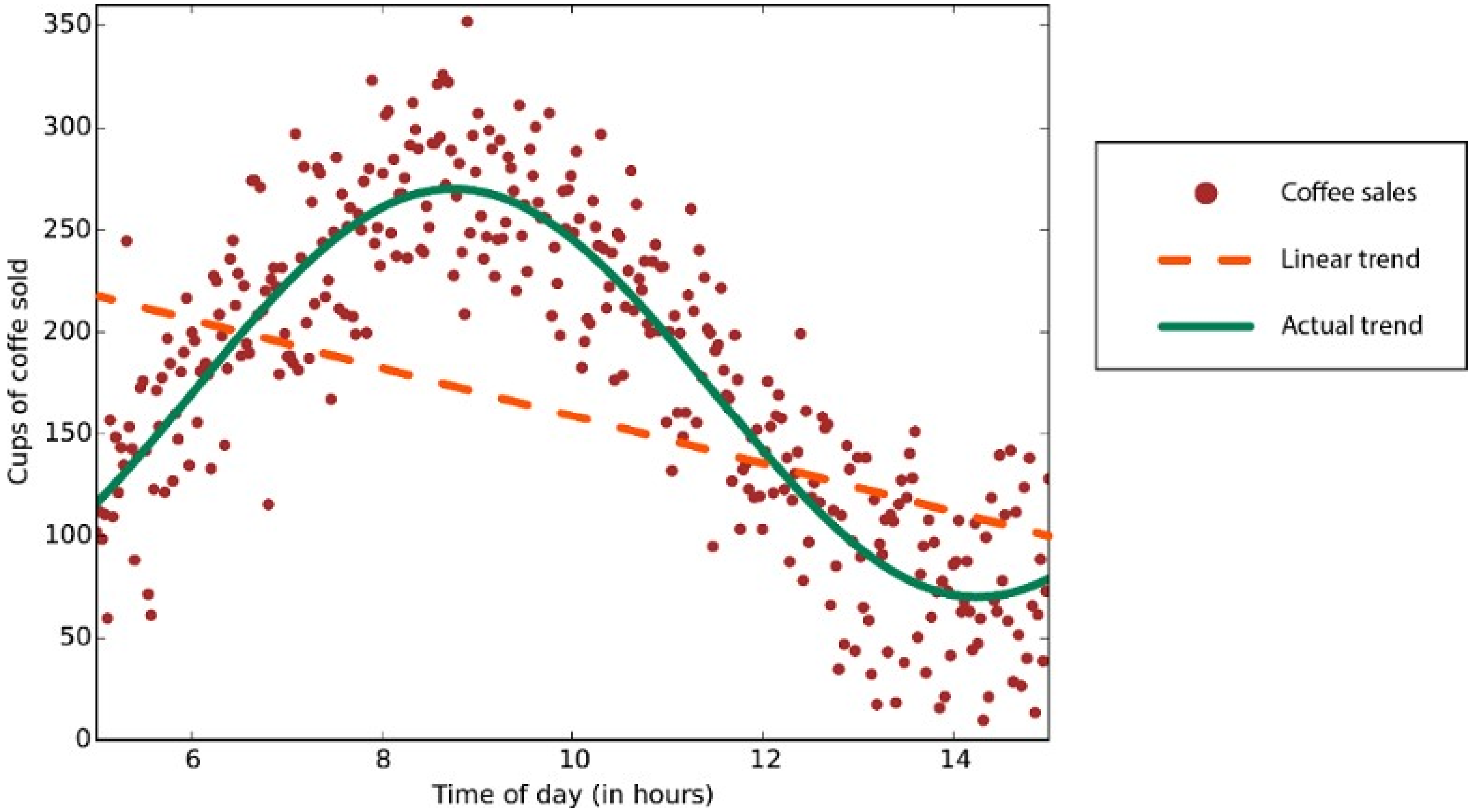
- Linear classification algorithms assume that classes can be separated by a straight line (or its higher-dimensional separation). that data trends follow a straight line.
- These assumptions aren't bad for some problems, but on others they bring accuracy down
- Logistic regression and Support Vector Machines

Linearity



***Non-linear class boundary** - relying on a linear classification algorithm would result in low accuracy*

Linearity



Data with a nonlinear trend *- using a linear regression method would generate much larger errors than necessary*

Number of Parameters

- Parameters are the knobs a data scientist gets to turn when setting up an algorithm.
- # Parameters affect the algorithm's behavior, such as error tolerance or # iterations, or options between variants of how the algorithm behaves.
- Training time and accuracy of the algorithm can sometimes be quite sensitive to getting just the right settings.

Number of Parameters

- Algorithms with large numbers parameters require the most trial and error to find a good combination.
- Having many parameters typically indicates that an algorithm has greater flexibility.
- It can often achieve very good accuracy, provided the right combination of parameter settings.

Number of features

- For certain types of data, the number of features can be very large compared to the number of data points.
- This is often the case with genetics or textual data.
- The large number of features can bog down some learning algorithms, making training time unfeasibly long.
- Support Vector Machines are particularly well suited to this case

Algorithm	Accuracy	Training time	Linearity	Parameters	Notes
Binomial classification (2 classes) ● - shows excellent accuracy, fast training times, and the use of linearity ○ - shows good accuracy and moderate training times					
logistic regression		●	●	5	
decision forest	●	○		6	
decision jungle	●	○		6	Low memory footprint
boosted decision tree	●	○		6	Large memory footprint
neural network	●			9	Additional customization is possible
averaged perceptron	○	○	●	4	
support vector machine		○	●	5	Good for large feature sets
locally deep support vector machine	○			8	Good for large feature sets
Bayes' point machine		○	●	3	

Algorithm	Accuracy	Training time	Linearity	Parameters	Notes
Multi-class classification ● - shows excellent accuracy, fast training times, and the use of linearity ○ - shows good accuracy and moderate training times					
logistic regression		●	●	5	
decision forest	●	○		6	
decision jungle	●	○		6	Low memory footprint
neural network	●			9	Additional customization is possible
one-v-all	-	-	-	-	See properties of the two-class method selected

Algorithm	Accuracy	Training time	Linearity	Parameters	Notes
Regression ● - shows excellent accuracy, fast training times, and the use of linearity ○ - shows good accuracy and moderate training times					
linear		●	●	4	
Bayesian linear		○	●	2	
decision forest	●	○		6	
boosted decision tree	●	○		5	Large memory footprint
fast forest quantile	●	○		9	Distributions rather than point predictions
neural network	●			9	Additional customization is possible
Poisson			●	5	Technically log-linear. For predicting counts
ordinal				0	For predicting rank-ordering

Algorithm	Accuracy	Training time	Linearity	Parameters	Notes
Anomaly detection ● - shows excellent accuracy, fast training times, and the use of linearity ○ - shows good accuracy and moderate training times					
[support vector machine]()	○	○		2	Especially good for large feature sets
[PCA-based anomaly detection]()		○	●	3	
[K-means]()		○	●	4	A clustering algorithm

The ML Cheat Sheet

The Conclusion

- Define the problem. What problems do you want to solve?
- Start simple. Be familiar with the data and the baseline results.
- Then try simpler methods first and then something more complicated.
- For Big Data, let the data first pass through data warehous

Sources/Credits

- https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/
- https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-choice
- https://stackoverflow.com/questions/2595176/which-machine-learning-classifier-to-choose-in-general