Carnegie Mellon University
Heinz College

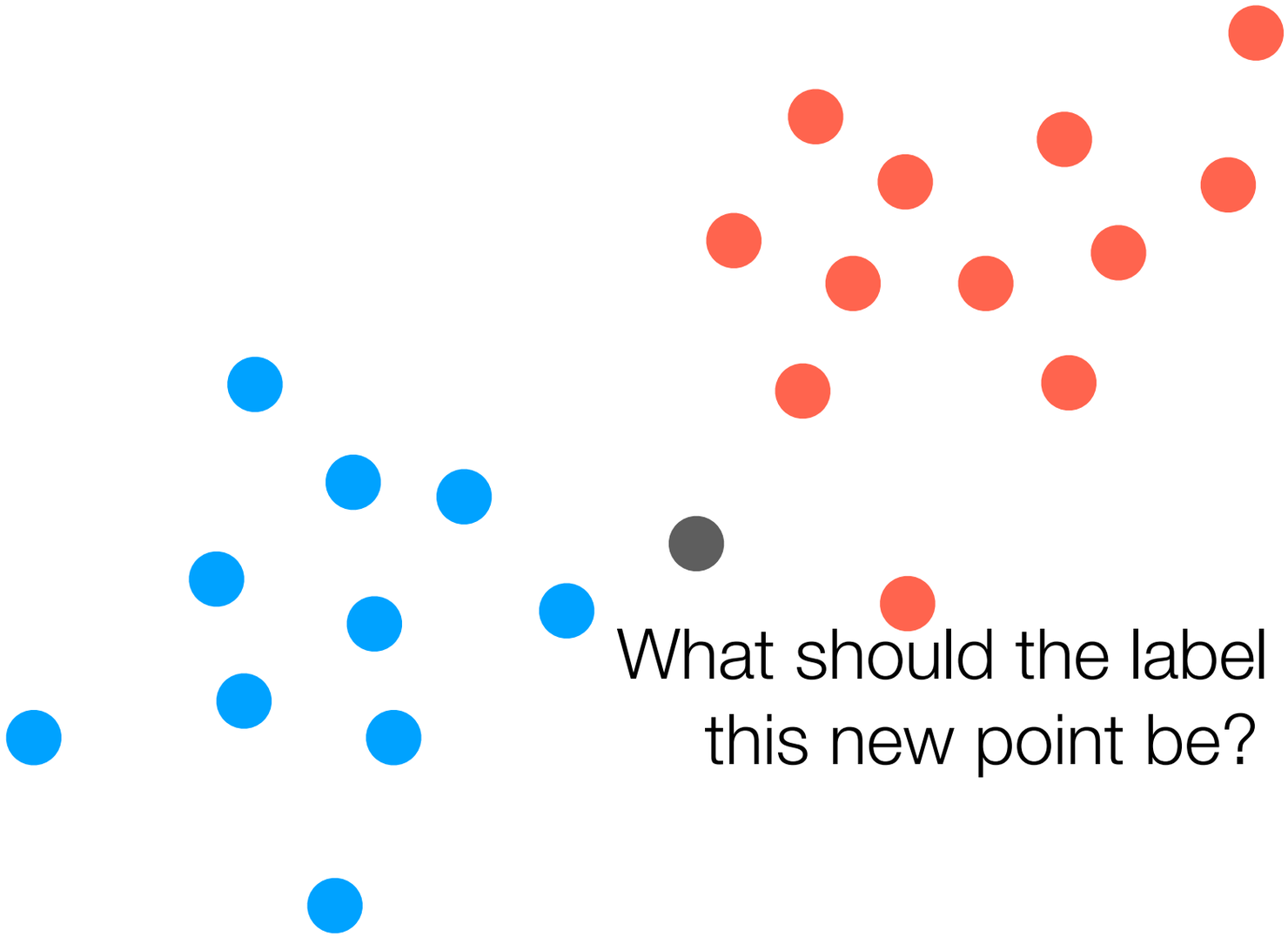
From Classical to Modern Classification Approaches

SVMs, decision trees and forests,
intro to neural nets and deep learning

George Chen
(some neural net & deep learning slides are by Phillip Isola)

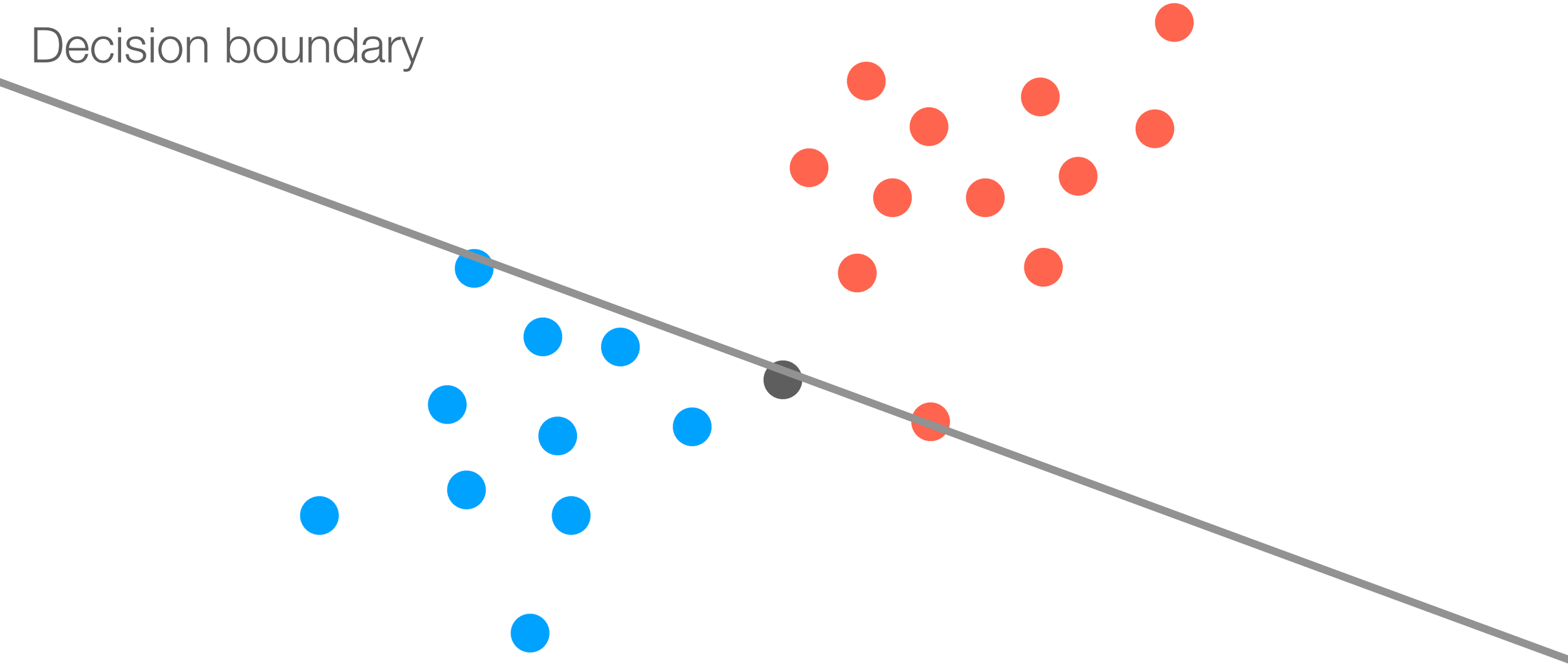
CMU 95-865 Spring 2018

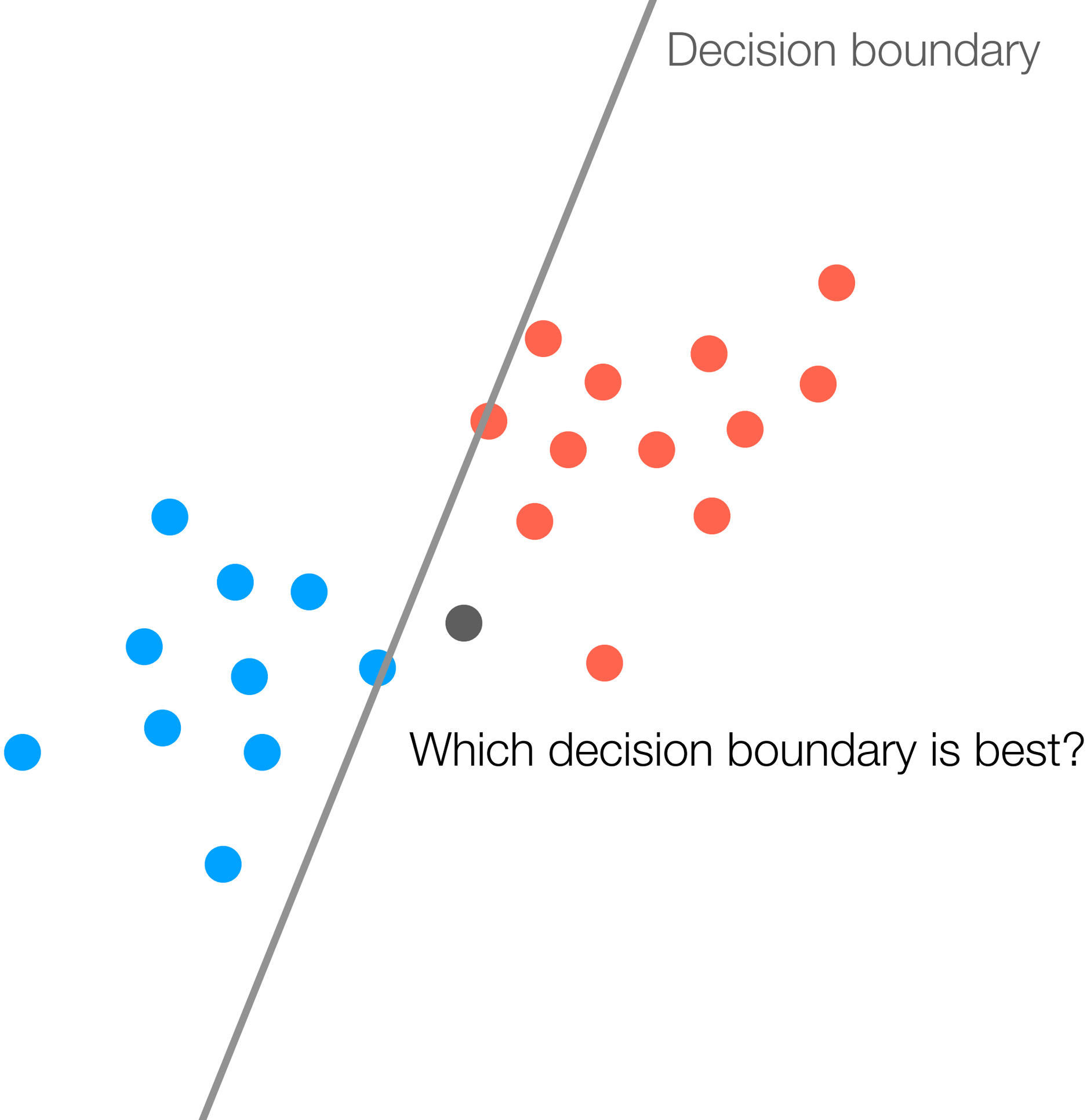
Support Vector Machines



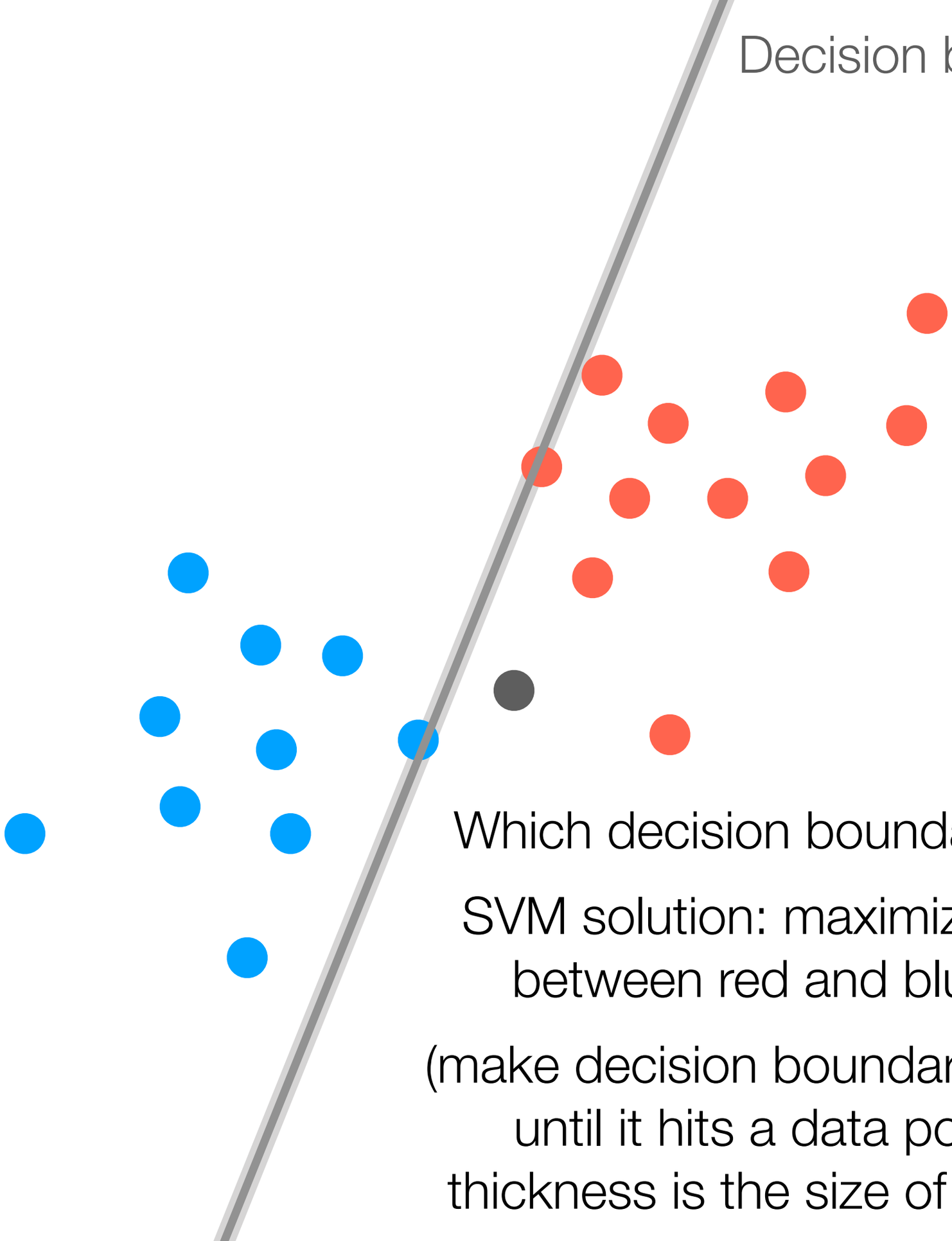
What should the label of this new point be?

Decision boundary





Which decision boundary is best?



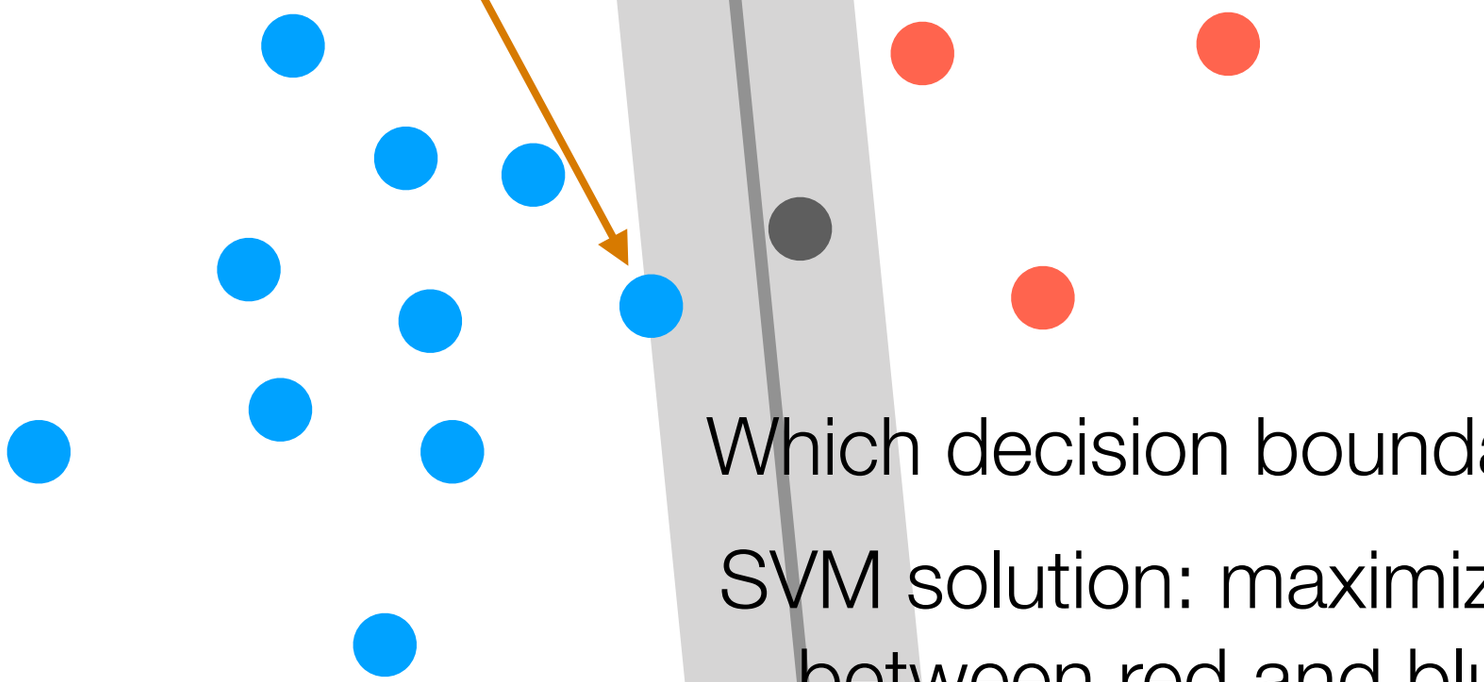
Which decision boundary is best?

SVM solution: maximize “margin” between red and blue points

(make decision boundary line thicker until it hits a data point—this thickness is the size of the margin)

Decision boundary

The points that the margin hits are called support vectors



Which decision boundary is best?

SVM solution: maximize “margin” between red and blue points

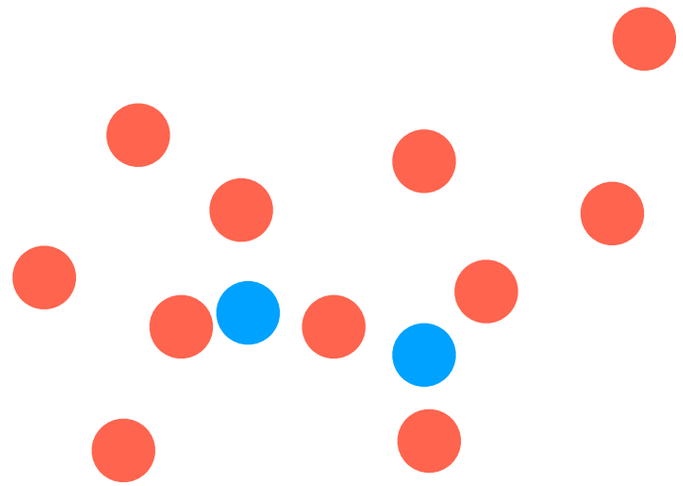
(make decision boundary line thicker until it hits a data point—this thickness is the size of the margin)

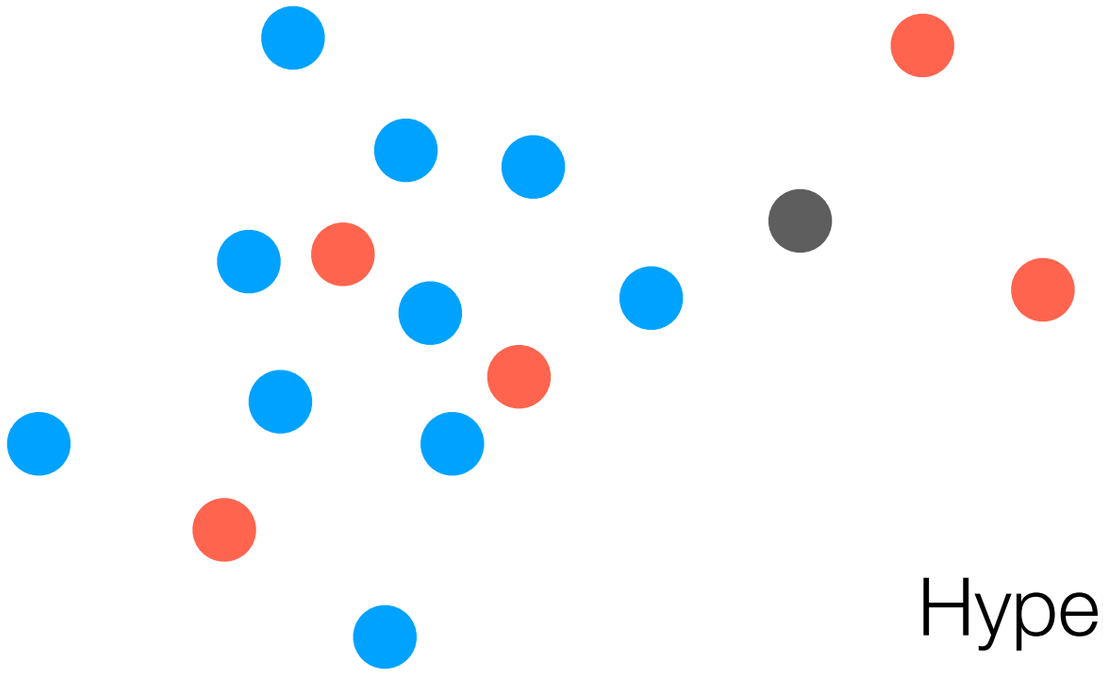
Decision boundary

The points that the margin hits are called support vectors

The decision boundary that the SVM outputs only depends on the support vectors

What if the points cannot actually be separated by a line?

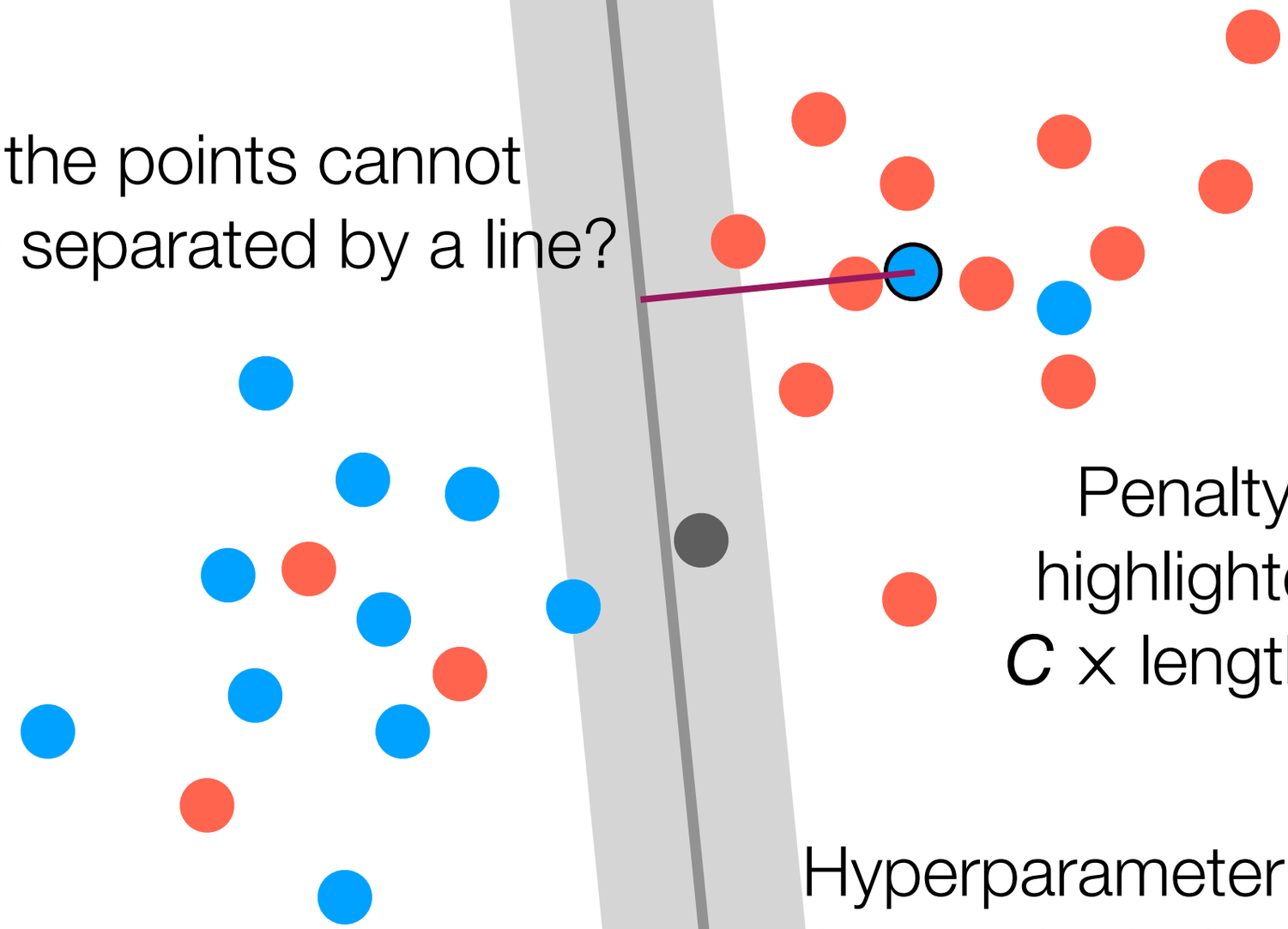




Hyperparameter *C* is a penalty for a point being on the wrong side of the decision boundary

C-Support Vector Classification

What if the points cannot actually be separated by a line?



Penalty incurred for highlighted blue point: *C* x length of purple line

Hyperparameter *C* is a penalty for a point being on the wrong side of the decision boundary

Larger *C* → work harder to fit all points

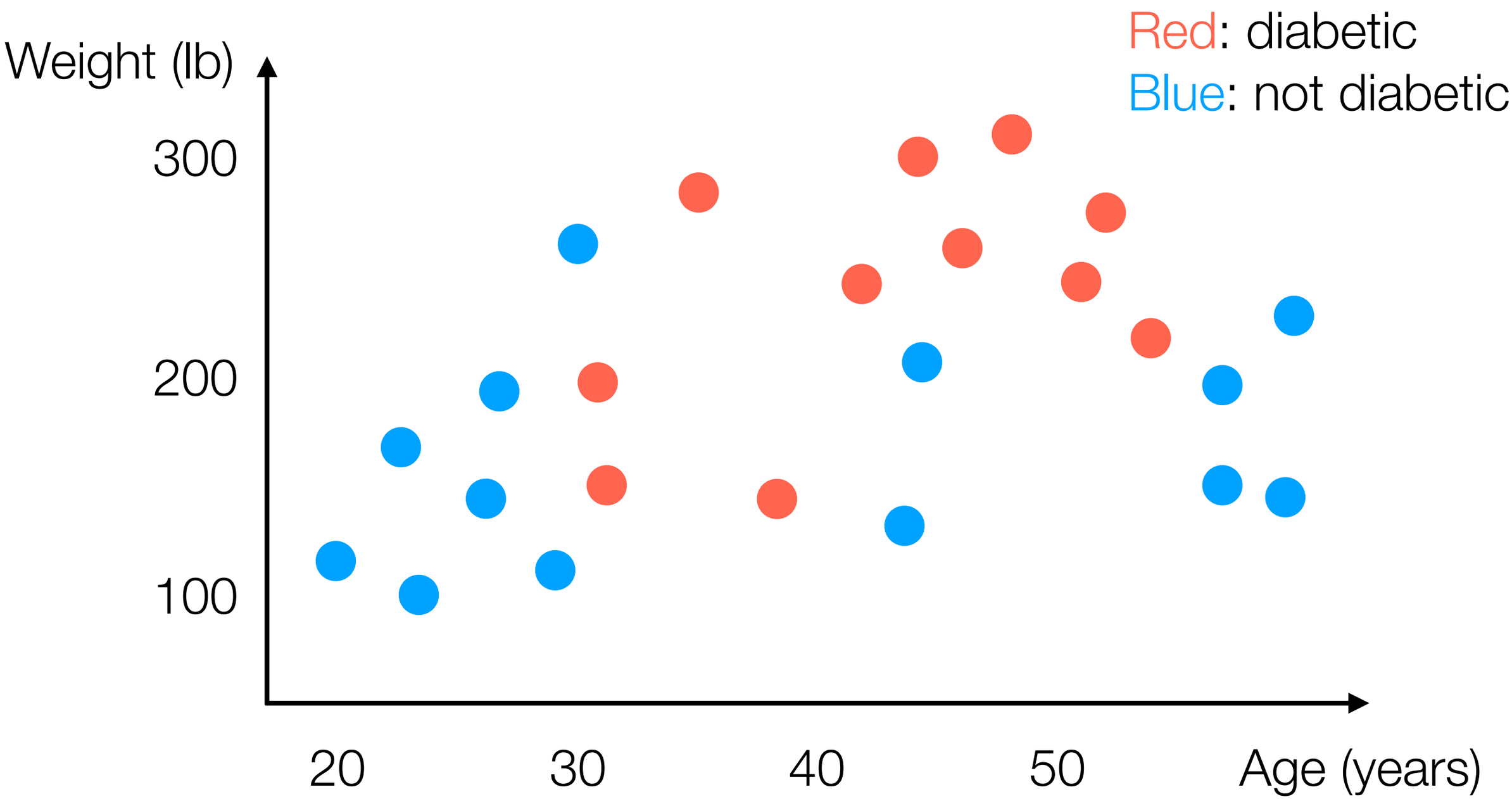
C-Support Vector Classification

- Basic version measures distance using Euclidean distance
  - Turns out to correspond to measuring similarity between two points by taking their dot product

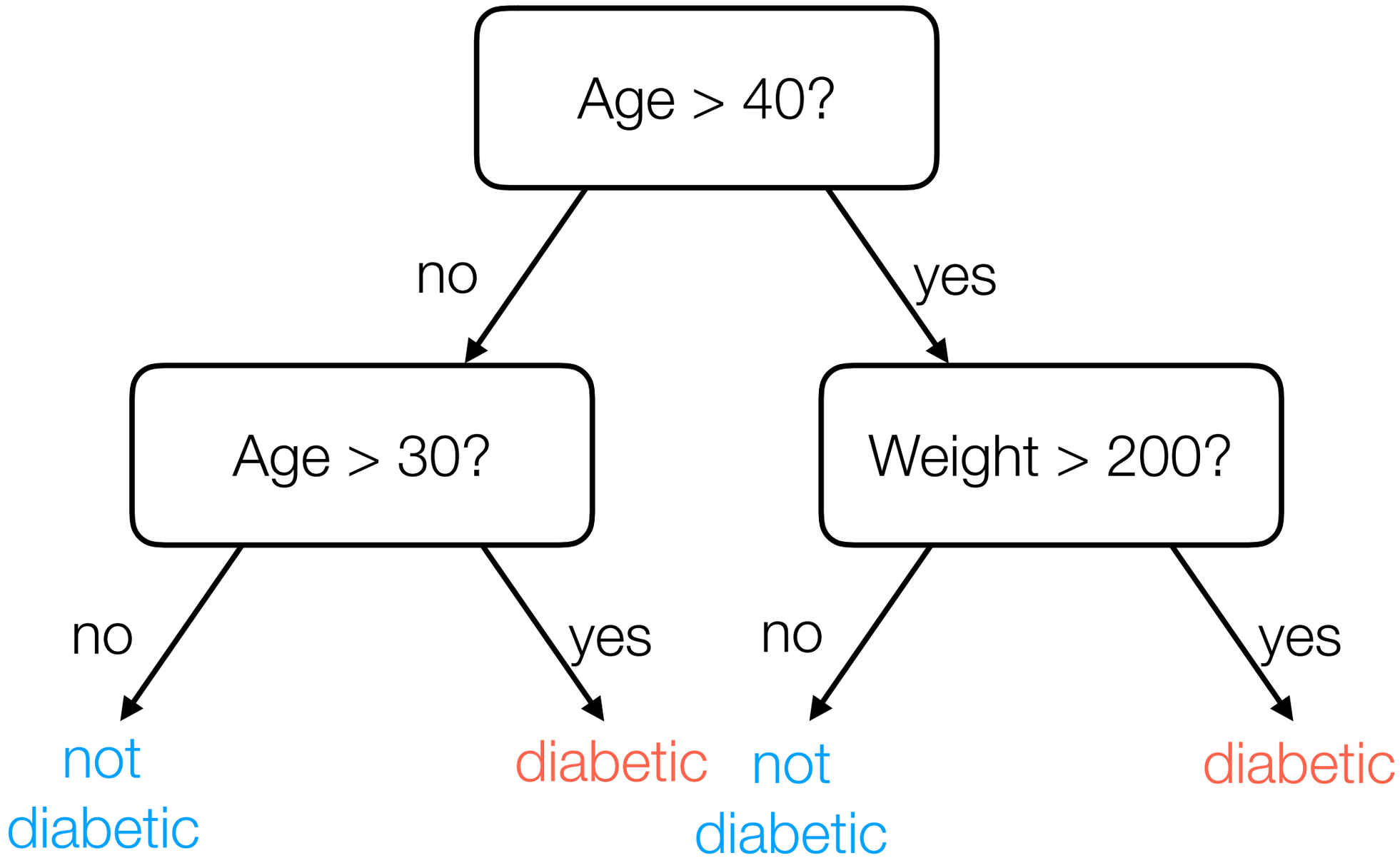
- Can instead use a different similarity function (“kernel” function) instead (popular choice: Gaussian kernel, also called “radial basis function” kernel)

Decision Trees

Example Made-Up Data



Example Decision Tree

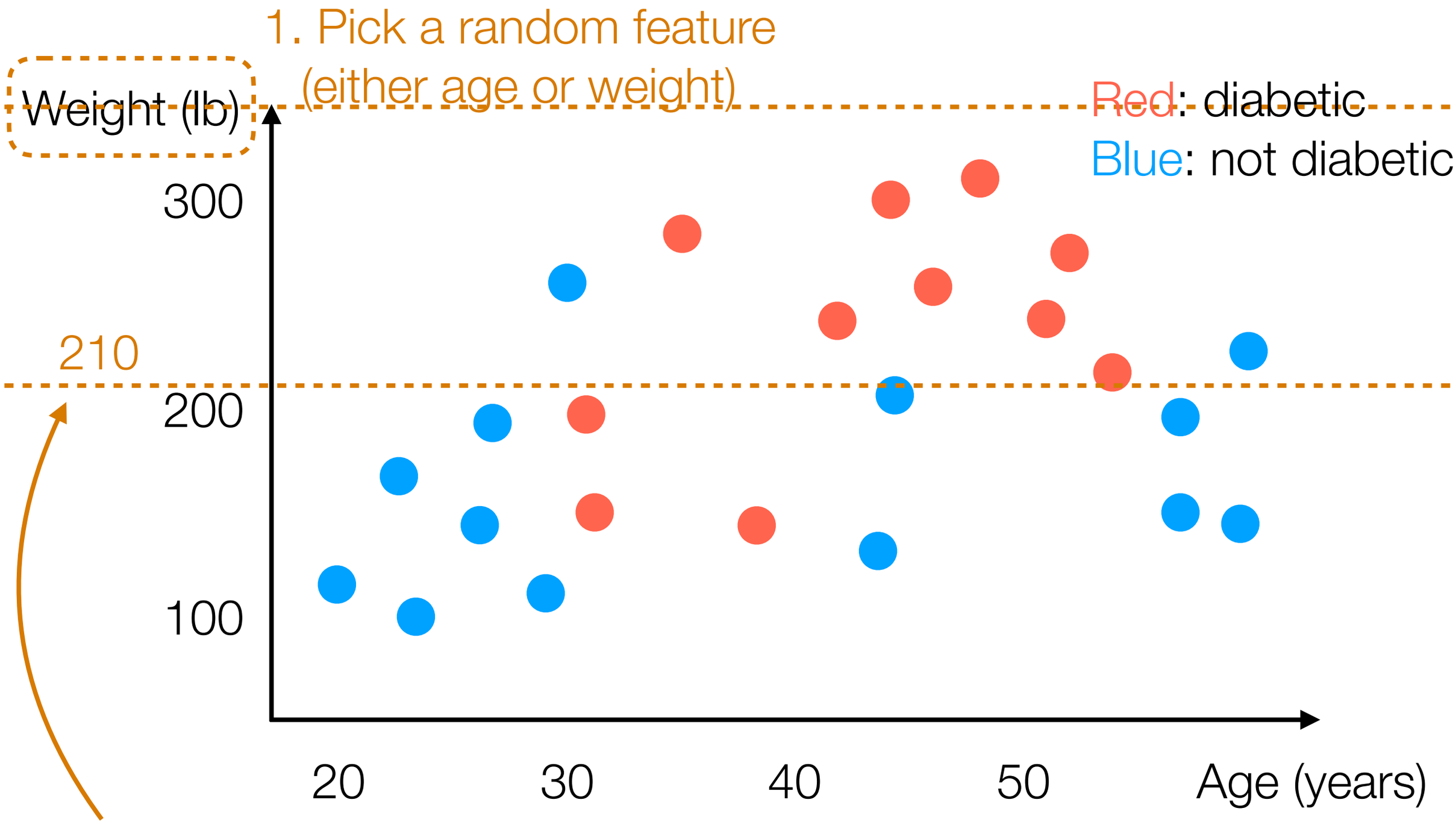


Learning a Decision Tree

- Many ways: general approach actually looks a lot like divisive clustering *but accounts for label information*

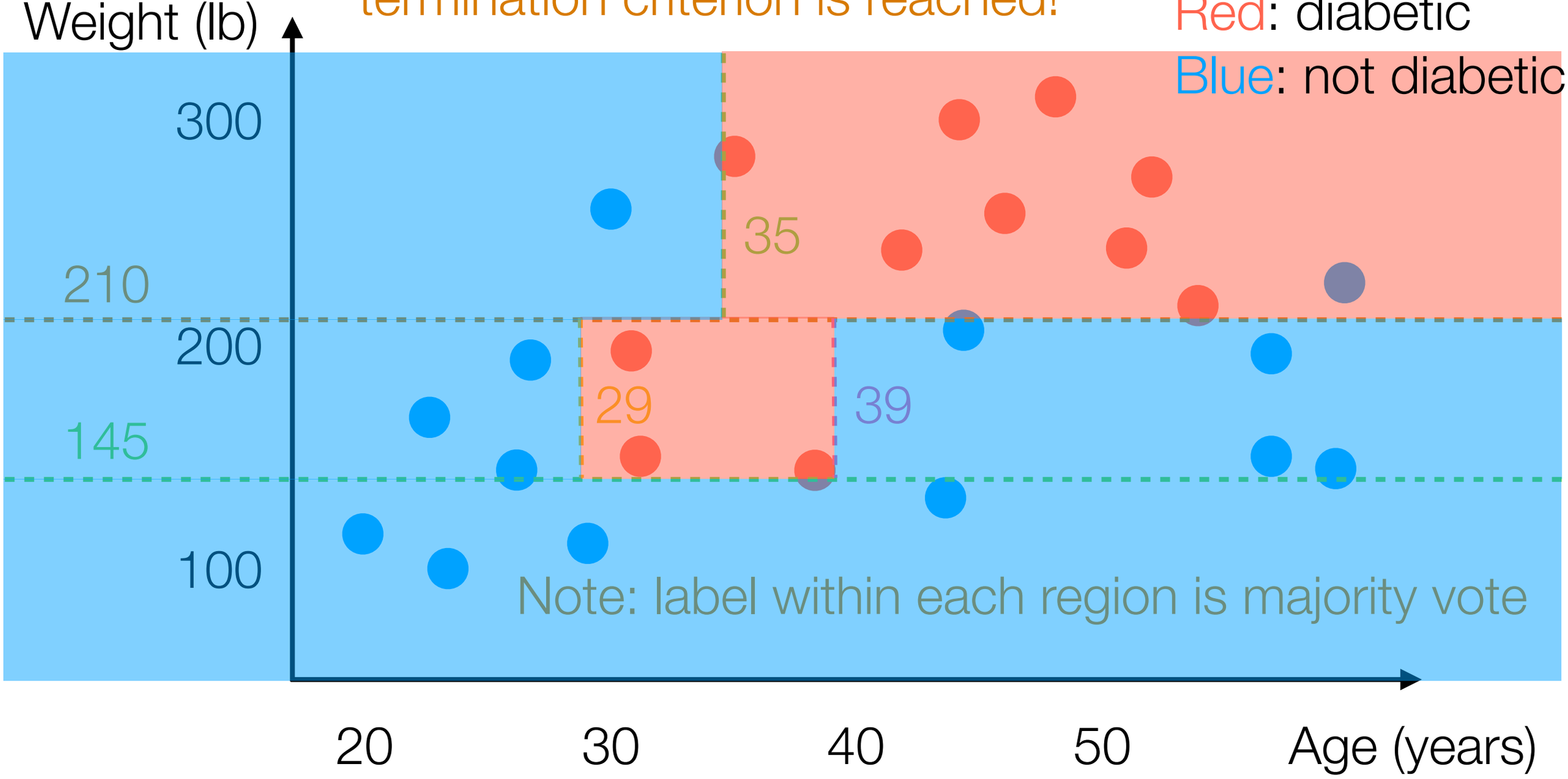
- I'll show one way (that nobody actually uses in practice) but it's easy to explain

Learning a Decision Tree



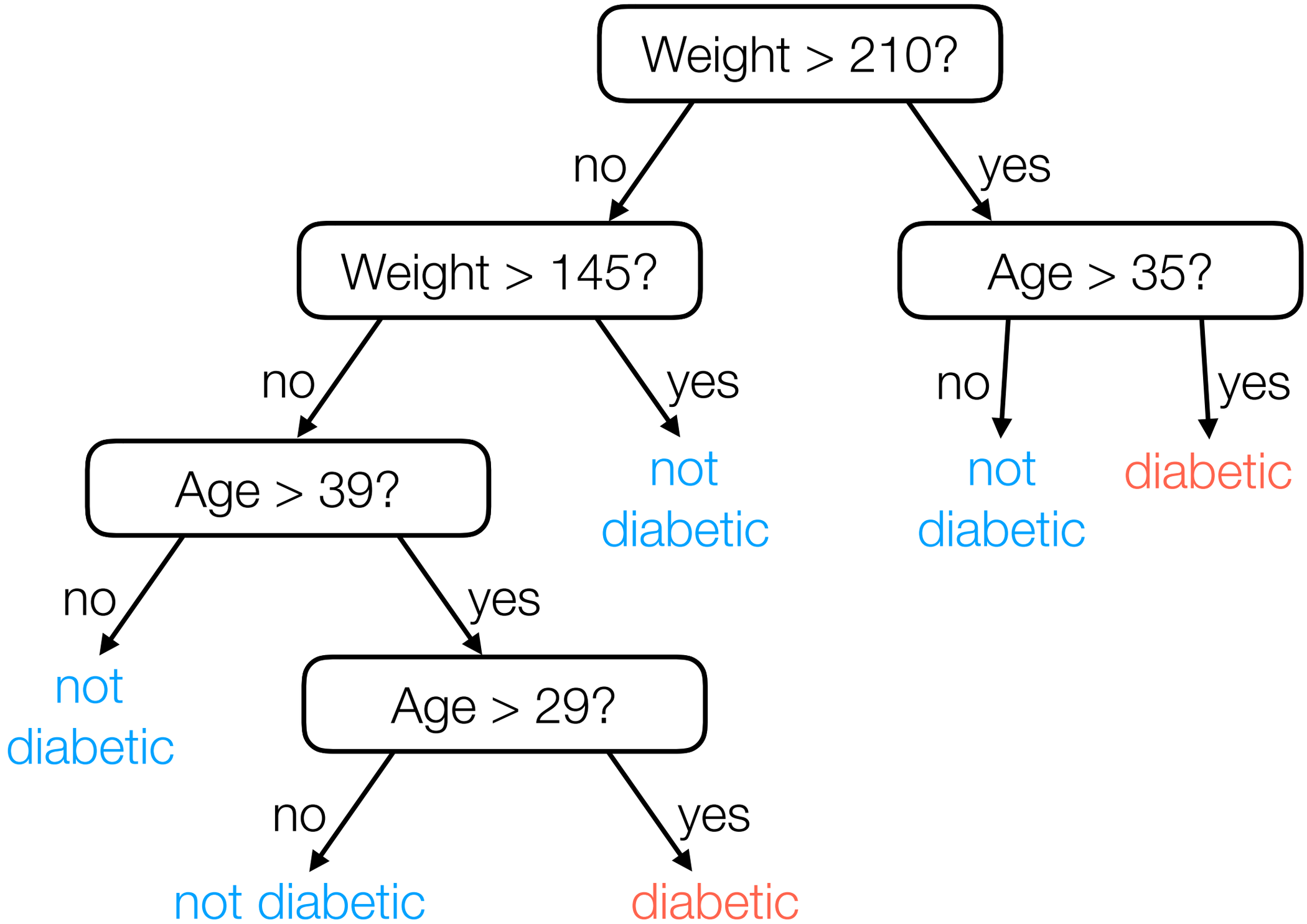
Learning a Decision Tree

Within each side, recurse until a termination criterion is reached!



Example termination criteria: ≥90% points within region has same label, number of points within region is <5

Decision Tree Learned



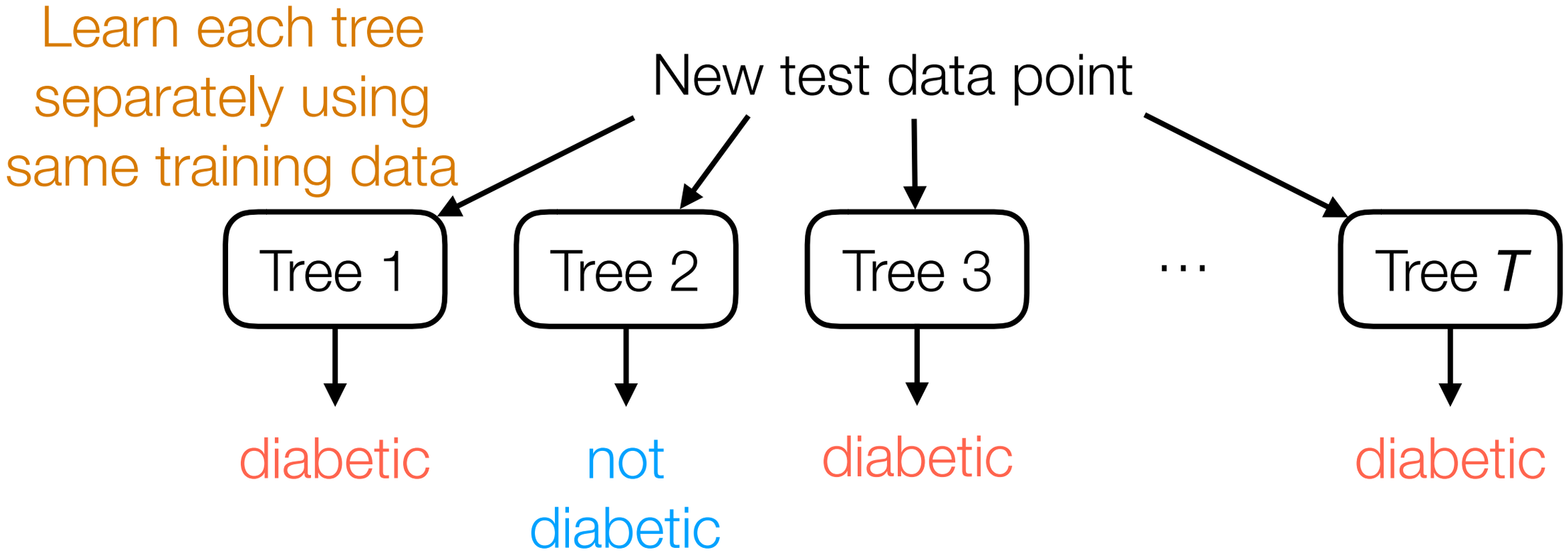
For a new person with feature vector (age, weight), easy to predict!

Decision Forest for Classification

- Typically, a decision tree is learned with randomness (e.g., we randomly chose which feature to threshold)

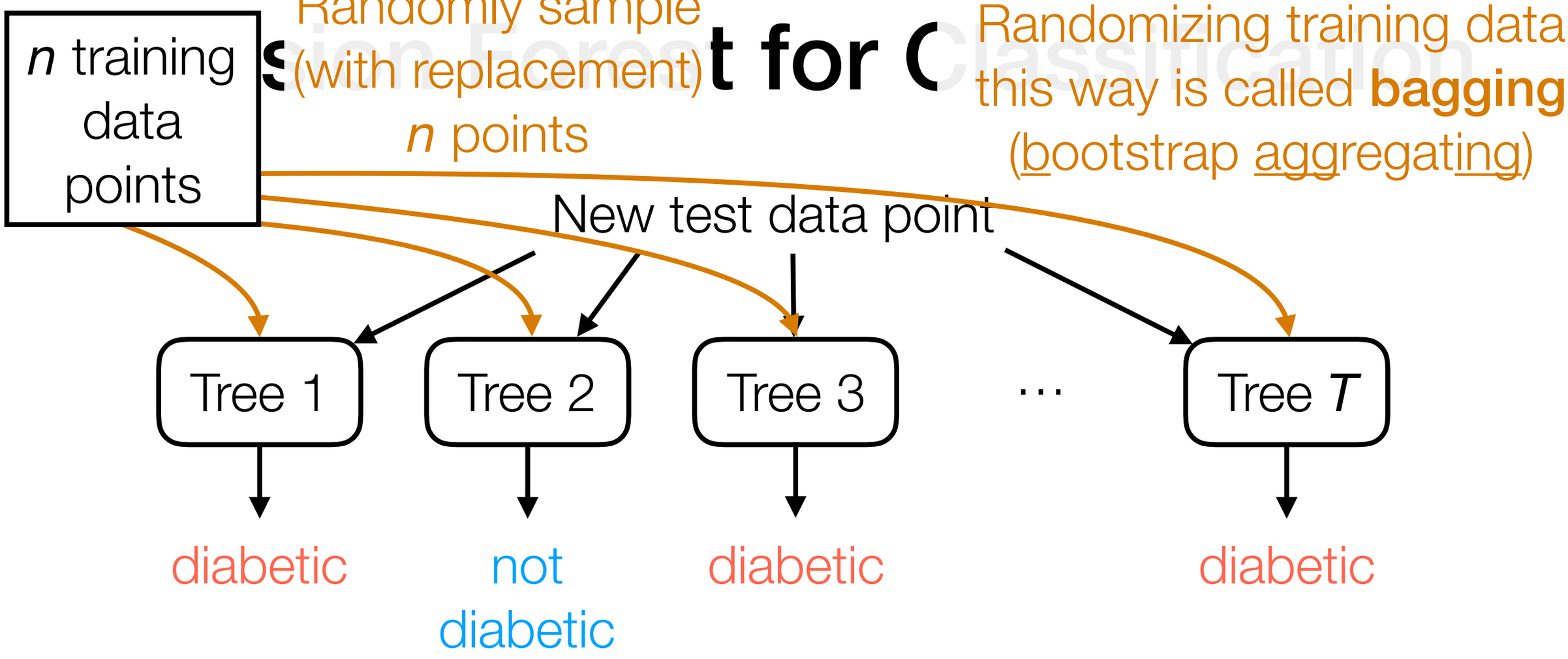
  ➔ by re-running the same learning procedure, we can get different decision trees that make different predictions!

- For a more stable prediction, use many decision trees



Final prediction: majority vote of the different trees' predictions

Random Forest for Classification



Question: What happens if all the trees are the same?

Adding randomness can make trees more different!

- **Random Forest:** in addition to randomly choosing features to threshold, also randomize training data used for each tree
- **Extremely randomized trees:** further randomize thresholds rather than trying to pick clever thresholds

Intro to Neural Nets and Deep Learning

IMAGENET

Over 10 million images, 1000 object classes

2011: Traditional computer vision achieves accuracy ~74%

2012: Initial deep neural network approach accuracy ~84%

2015 onwards: Deep learning achieves accuracy 96%+

Russakovsky et al. ImageNet Large Scale Visual Recognition Challenge. IJCV 2015.

Deep Learning Takeover

Academia:

- Top computer vision conferences (CVPR, ICCV, ECCV) are now nearly all about deep learning
- Top machine learning conferences (ICML, NIPS) have *heavily* been taken over by deep learning

Heavily dominated by industry now!

Extremely useful in practice:

- Near human level image classification (including handwritten digit recognition)
- Near human level speech recognition
- Improvements in machine translation, text-to-speech
- Self-driving cars
- *Better* than humans at playing Go

Google facebook amazon

Google DeepMind’s AlphaGo vs Lee Sedol, 2016

Is it all hype?

Fooling Neural Networks in the Physical World with 3D Adversarial Objects

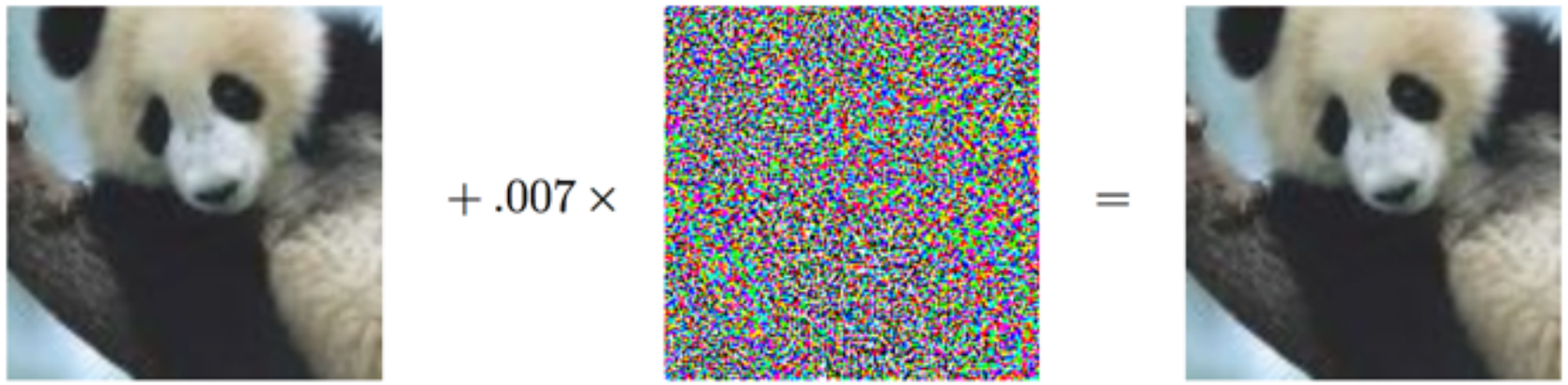
31 Oct 2017 · 3 min read — shared on Hacker News, Lobsters, Reddit, Twitter

We've developed an approach to generate *3D adversarial objects* that reliably fool neural networks in the real world, no matter how the objects are looked at.

Neural network based classifiers reach near-human performance in many tasks, and they're used in high risk, real world systems. Yet, these same neural networks are particularly vulnerable to *adversarial examples*, carefully perturbed inputs that cause

Source: labsix

Source: Gizmodo article “This Neural Network's Hilariously Bad Image Descriptions Are Still Advanced AI”. September 16, 2015. (They’re using the NeuralTalk image-to-caption software.)



panda
~58% confidence

adversarial
noise

gibbon
~99% confidence

Source: Goodfellow, Shlens, and Szegedy. Explaining and Harnessing Adversarial Examples. ICLR 2015.

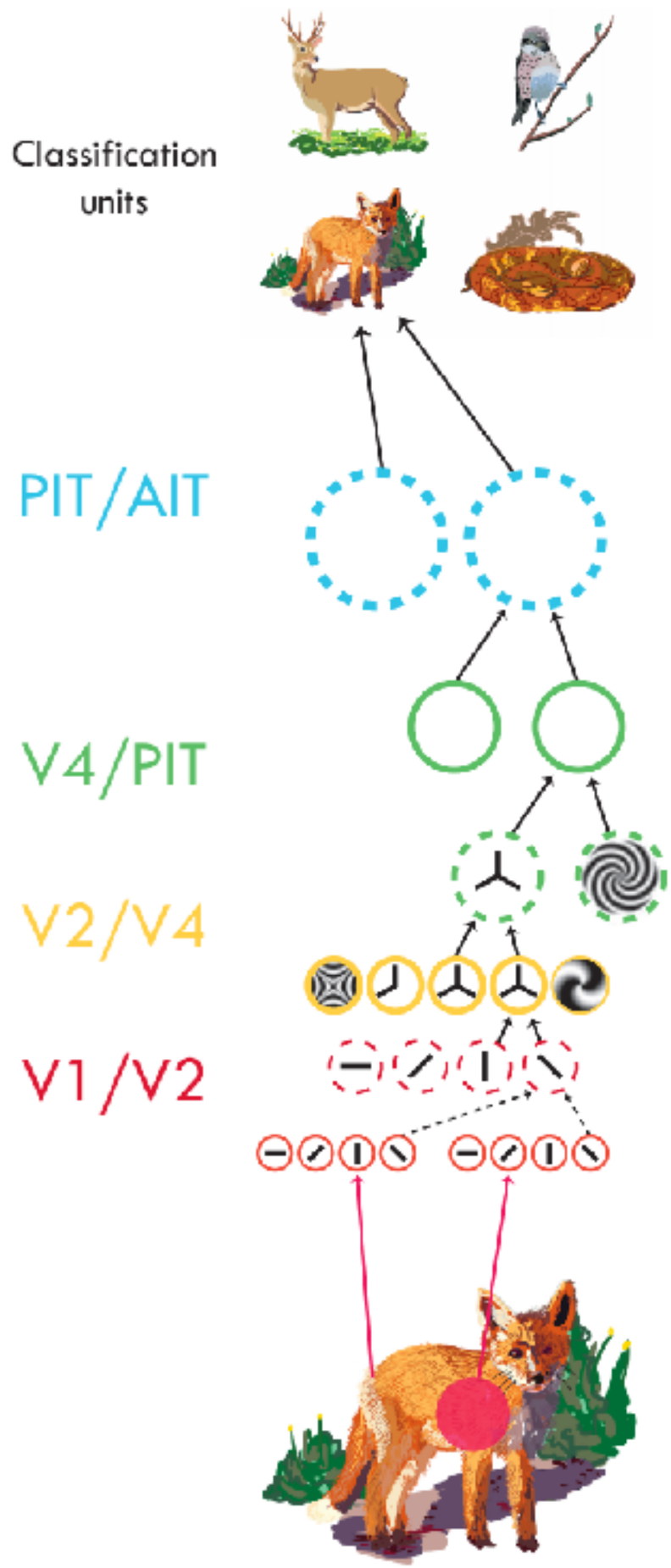
Another AI Winter?

~1970's: First AI winter over symbolic AI

~1980's: Second AI winter over "expert systems"

Every time: Lots of hype, explosion in funding, then bubble bursts

What is deep learning?



Slide by Phillip Isola

Serre, 2014

Basic Idea

Brain/Machine

“clown fish”

Slide by Phillip Isola

Object Recognition

Edges

Texture

Colors

Segments

Parts

“clown fish”

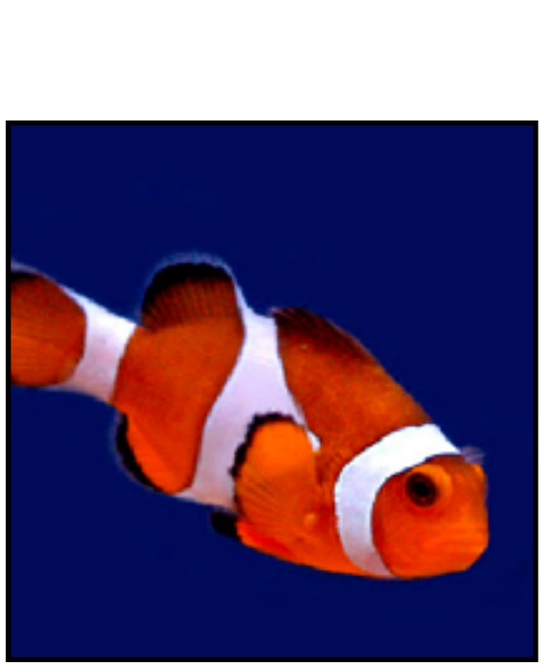
Feature extractors

Classifier

Slide by Phillip Isola

Object Recognition

Learned



Edges

Texture

Colors

Segments

Parts

“clown fish”

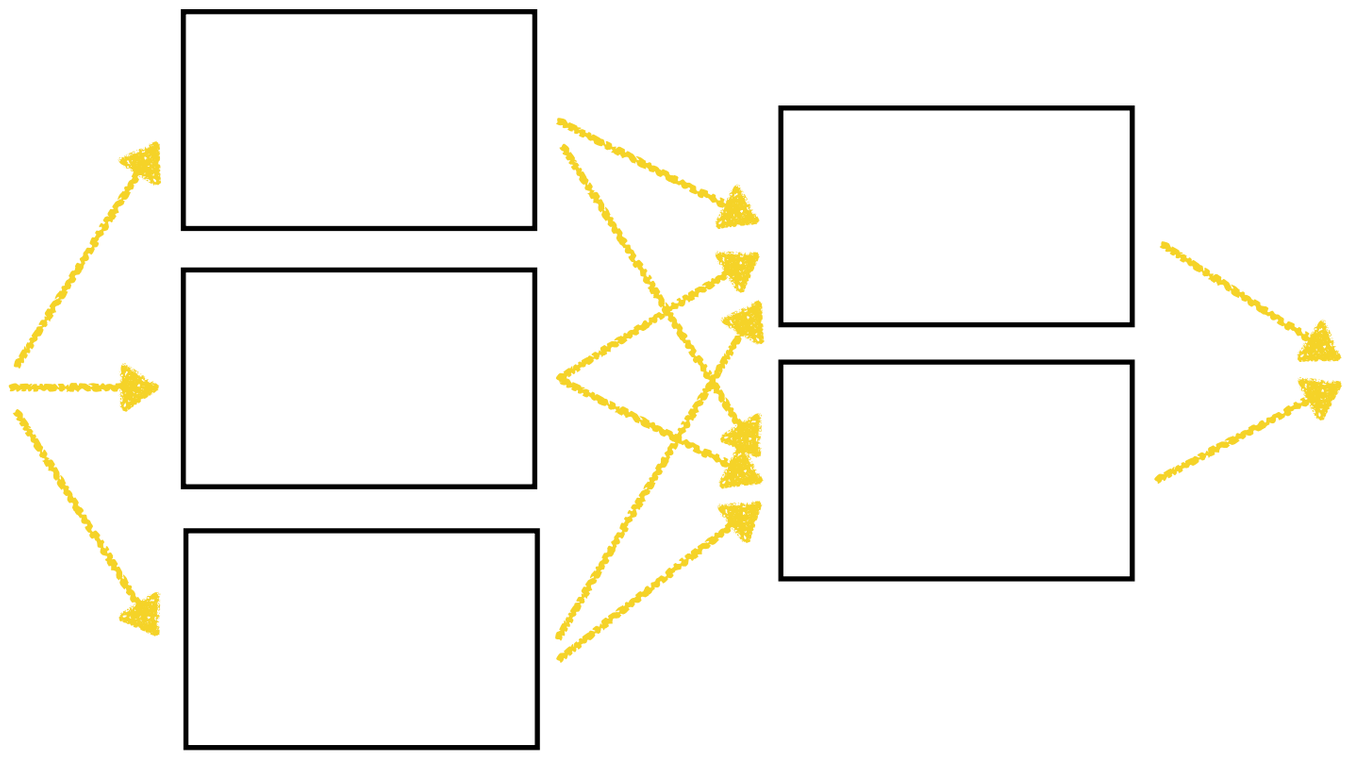
Feature extractors

Classifier

Slide by Phillip Isola

Neural Network

Learned

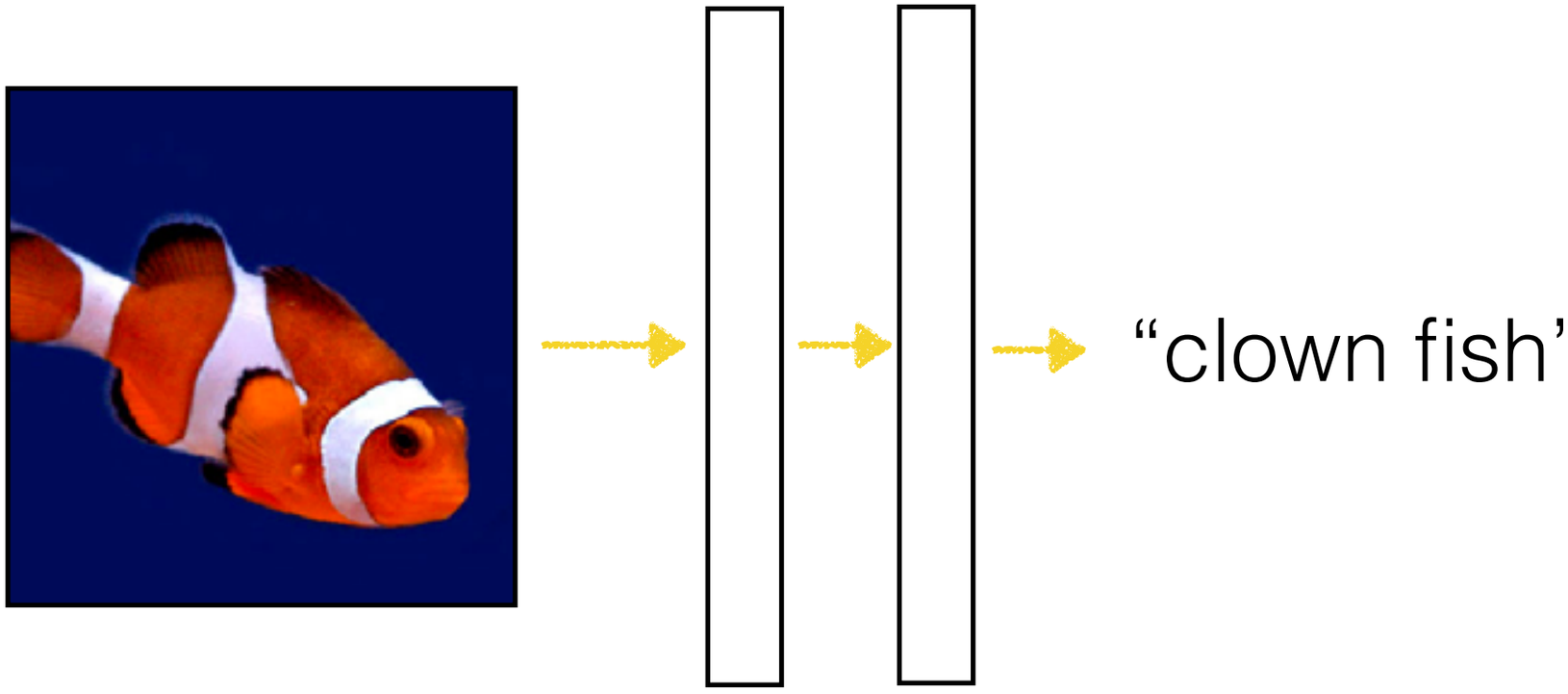


“clown fish”

Slide by Phillip Isola

Neural Network

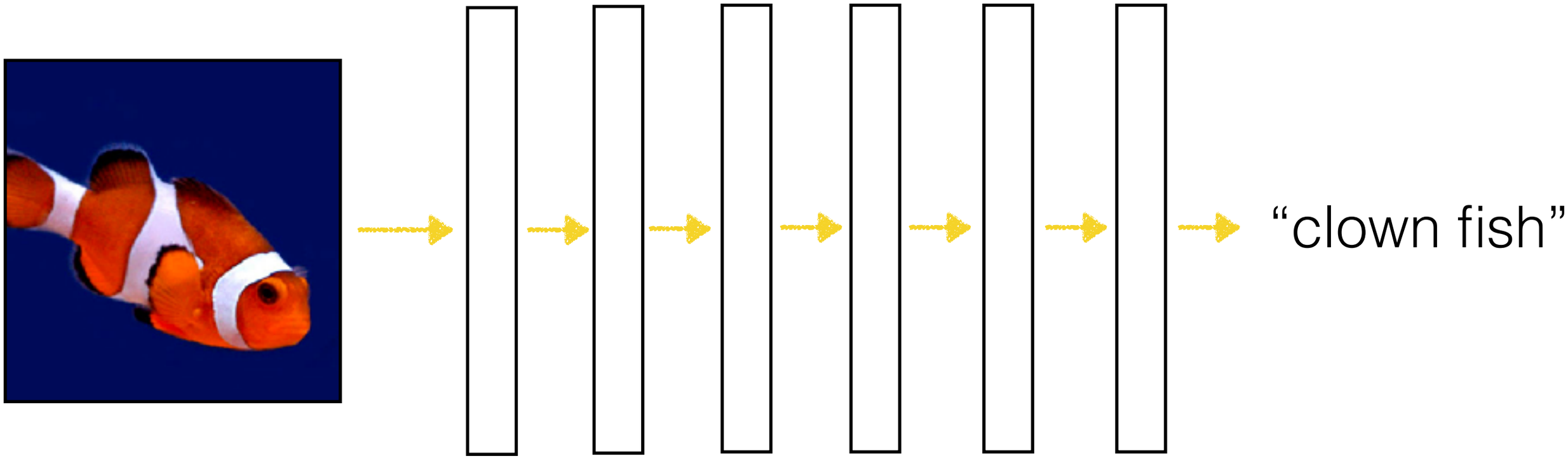
Learned



Slide by Phillip Isola

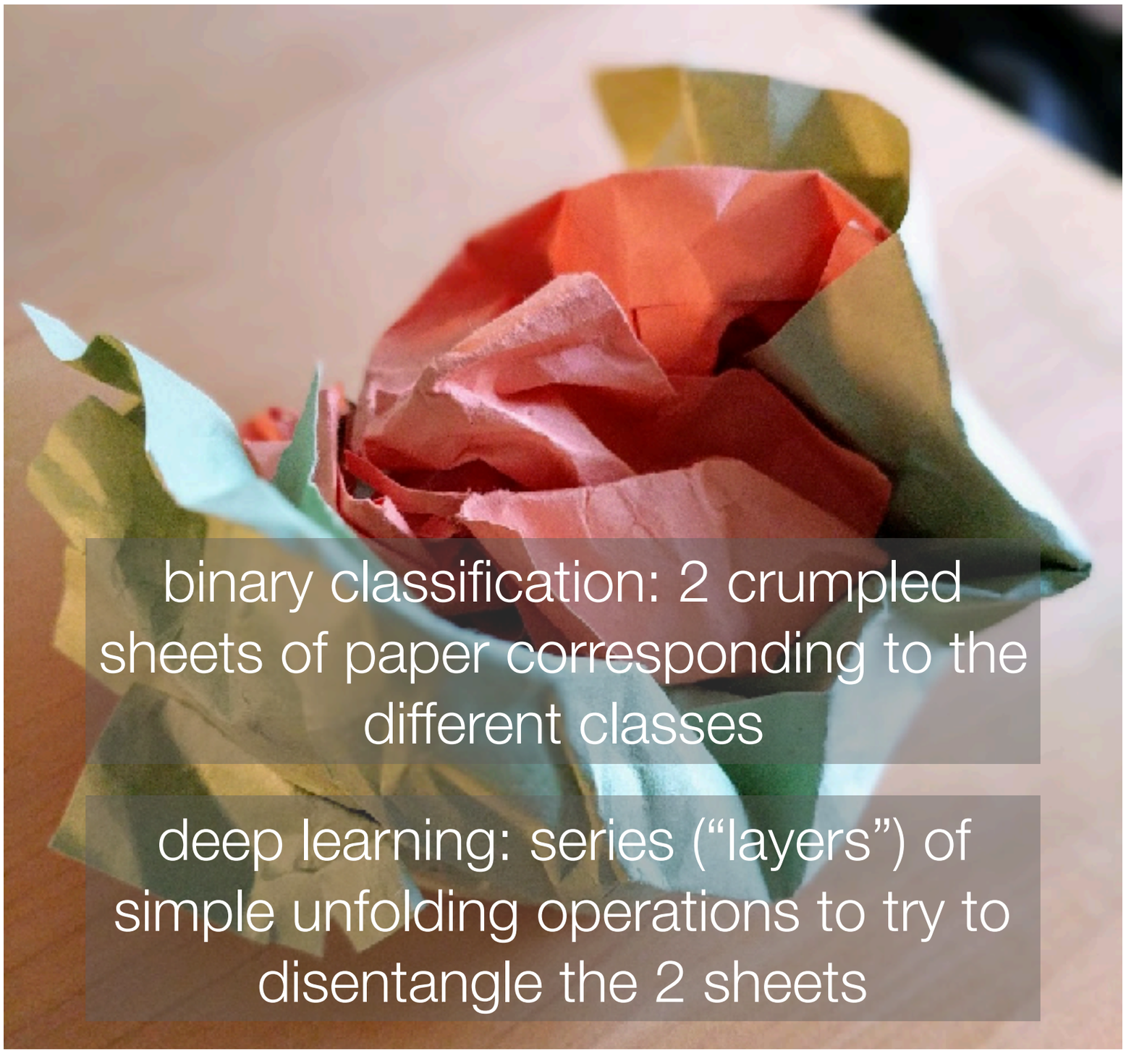
Deep Neural Network

Learned



Slide by Phillip Isola

Crumpled Paper Analogy



Analogy: Francois Chollet, photo: George Chen