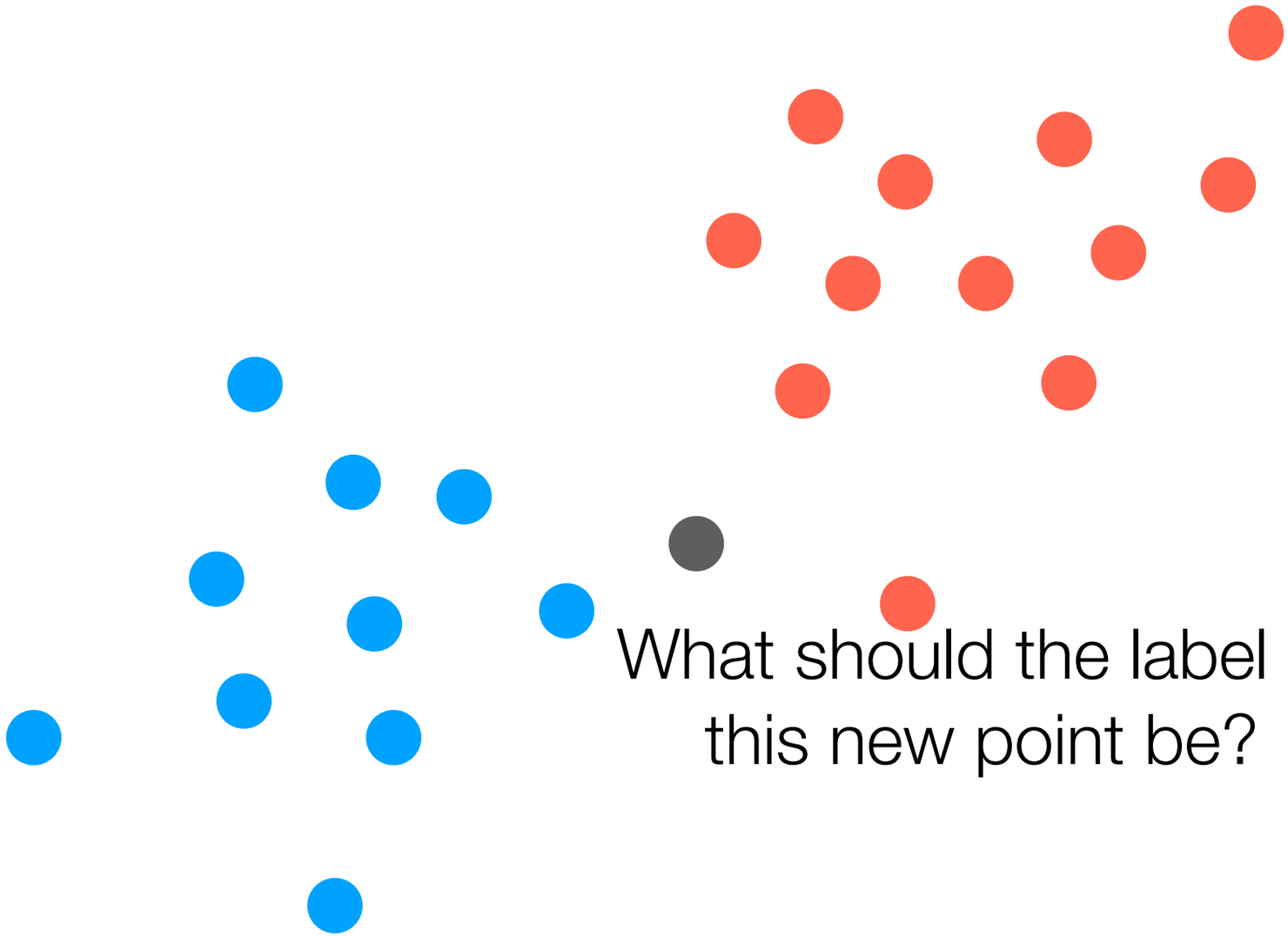
Carnegie Mellon University
Heinz College

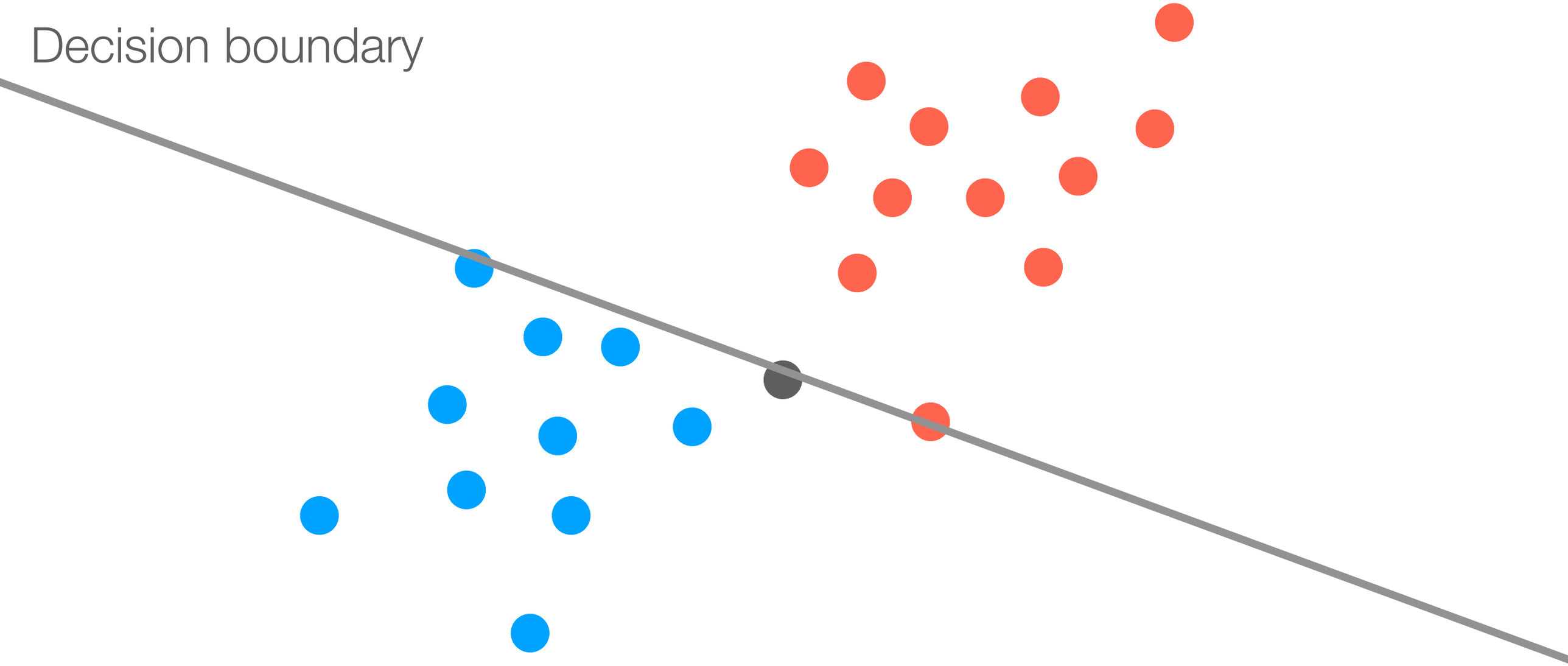
95-865: Support vector machines (another discriminative prediction method)

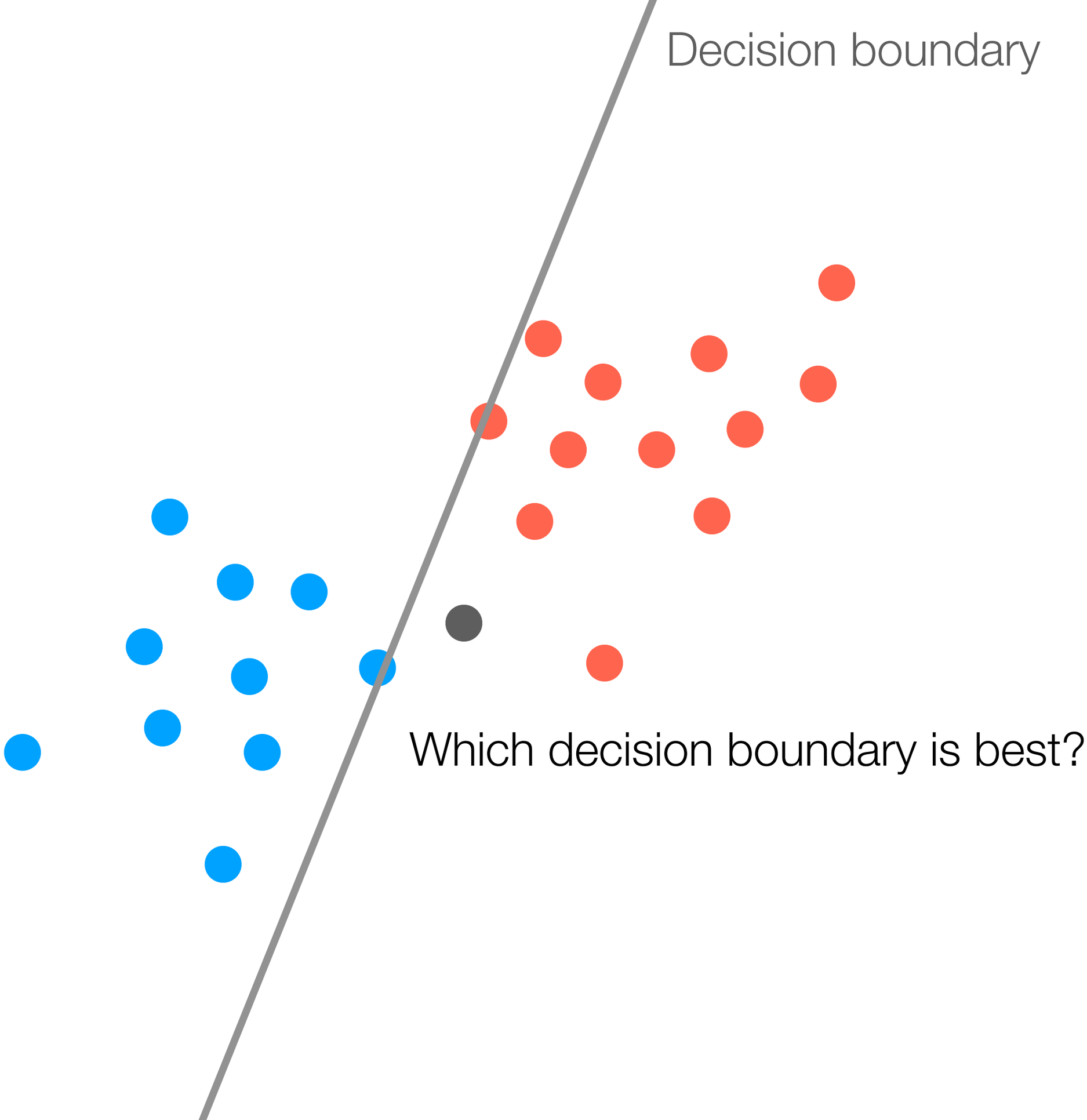
Slides by George Chen



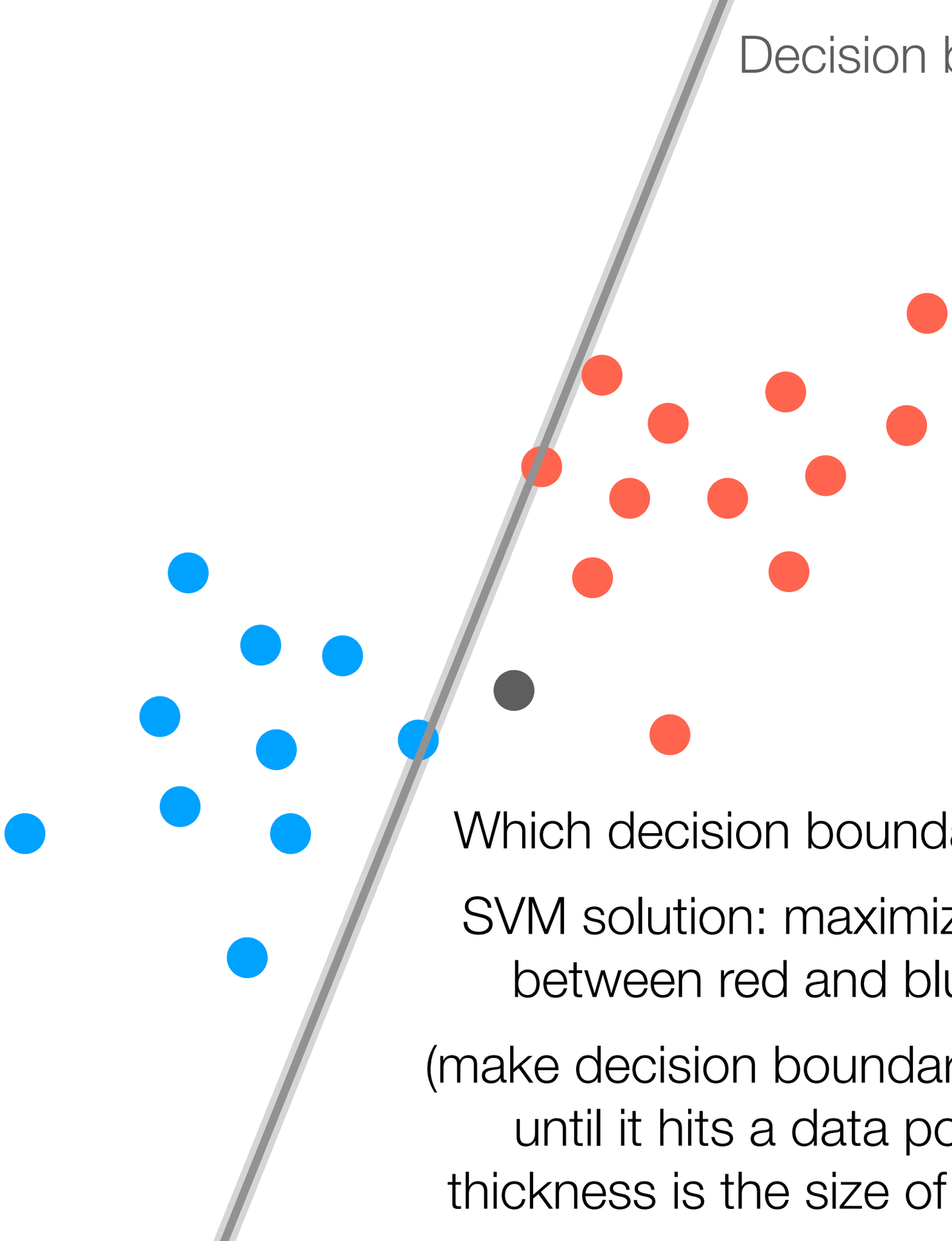
What should the label of this new point be?

Decision boundary





Which decision boundary is best?



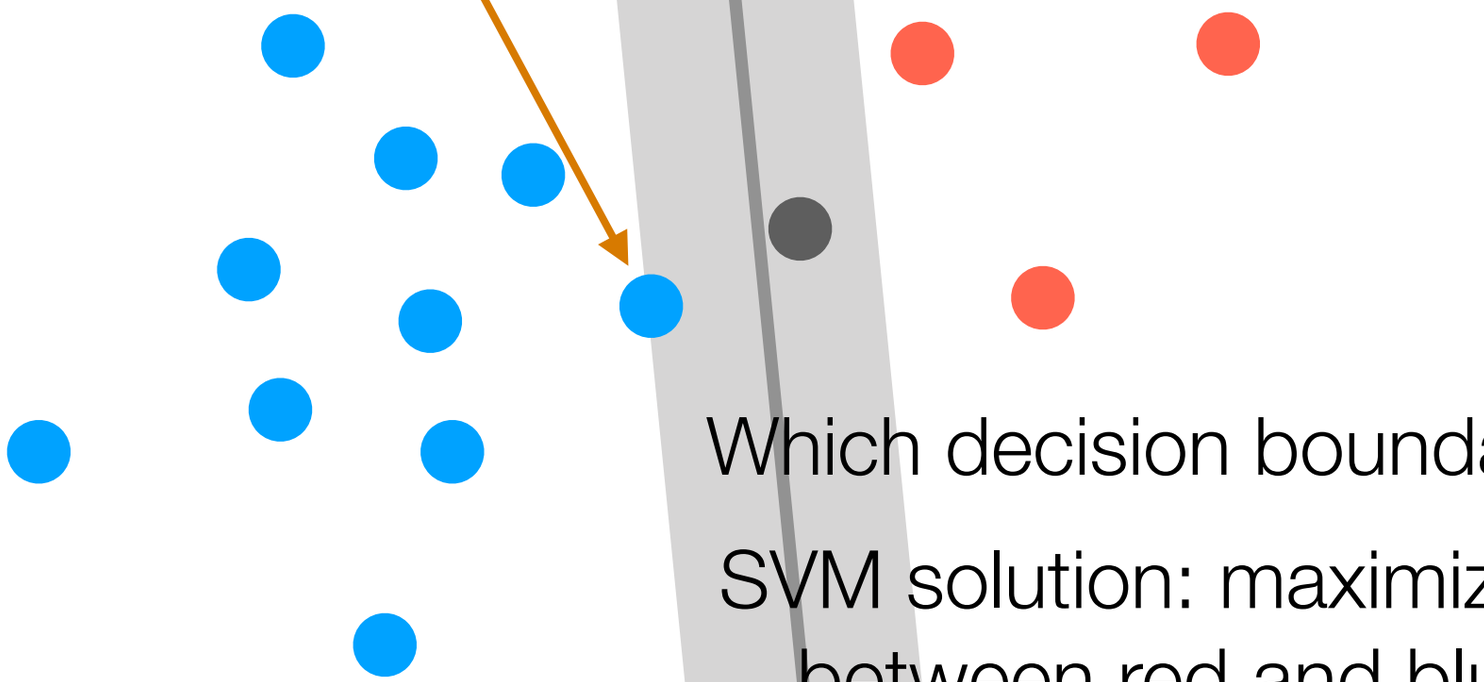
Which decision boundary is best?

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

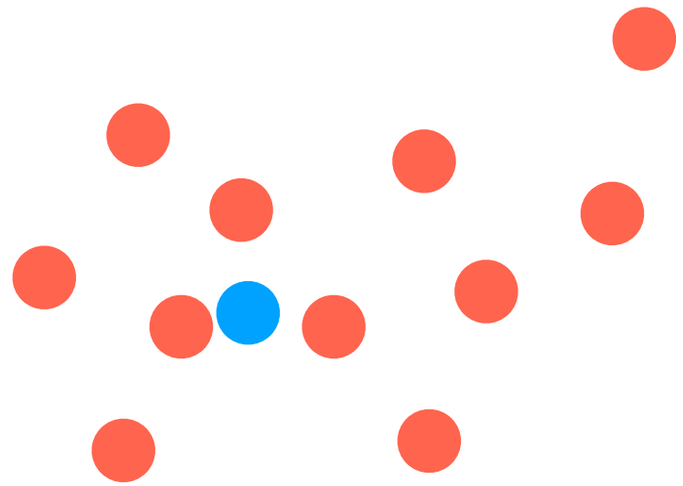
(make decision boundary line thicker until it hits a data point—this thickness is the size of the margin)

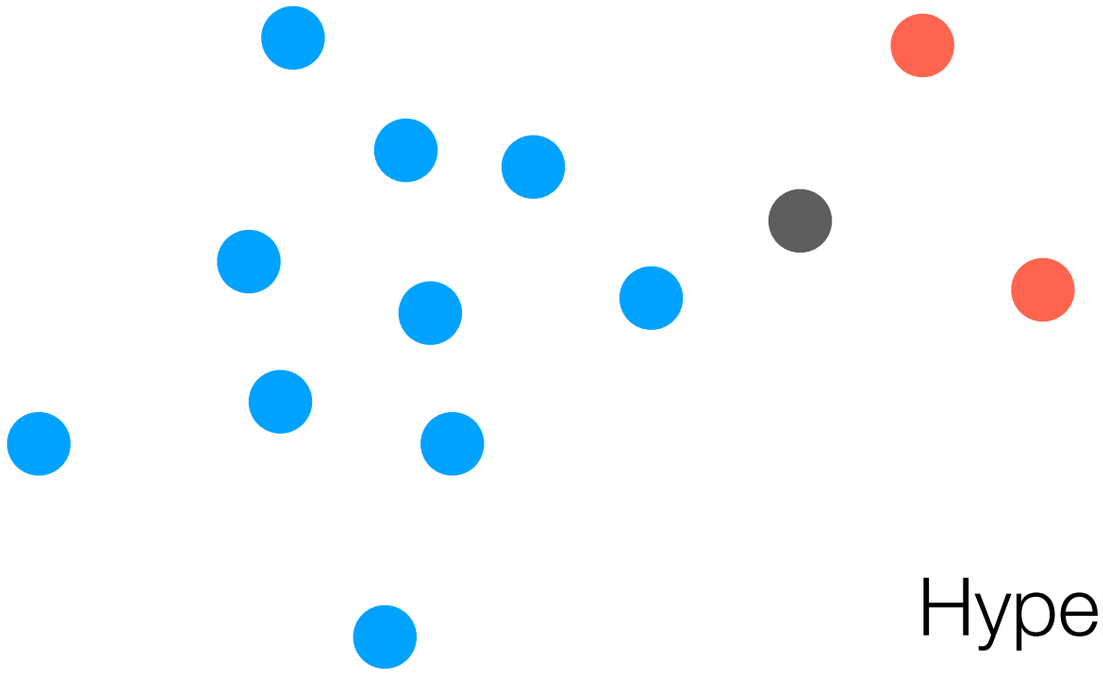
Decision boundary

The points that the margin hits are called support vectors

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

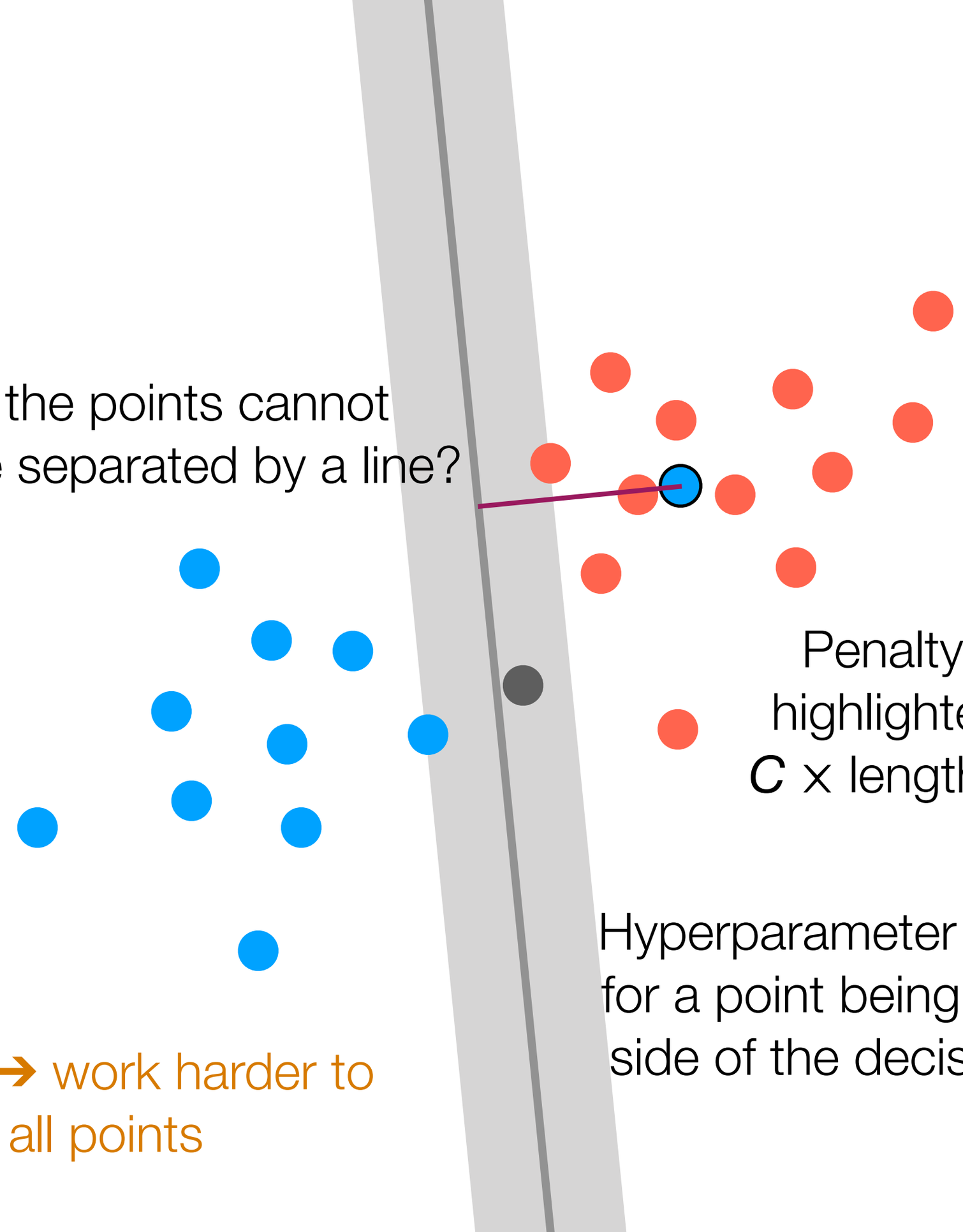
What if the points cannot actually be separated by a line?





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

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



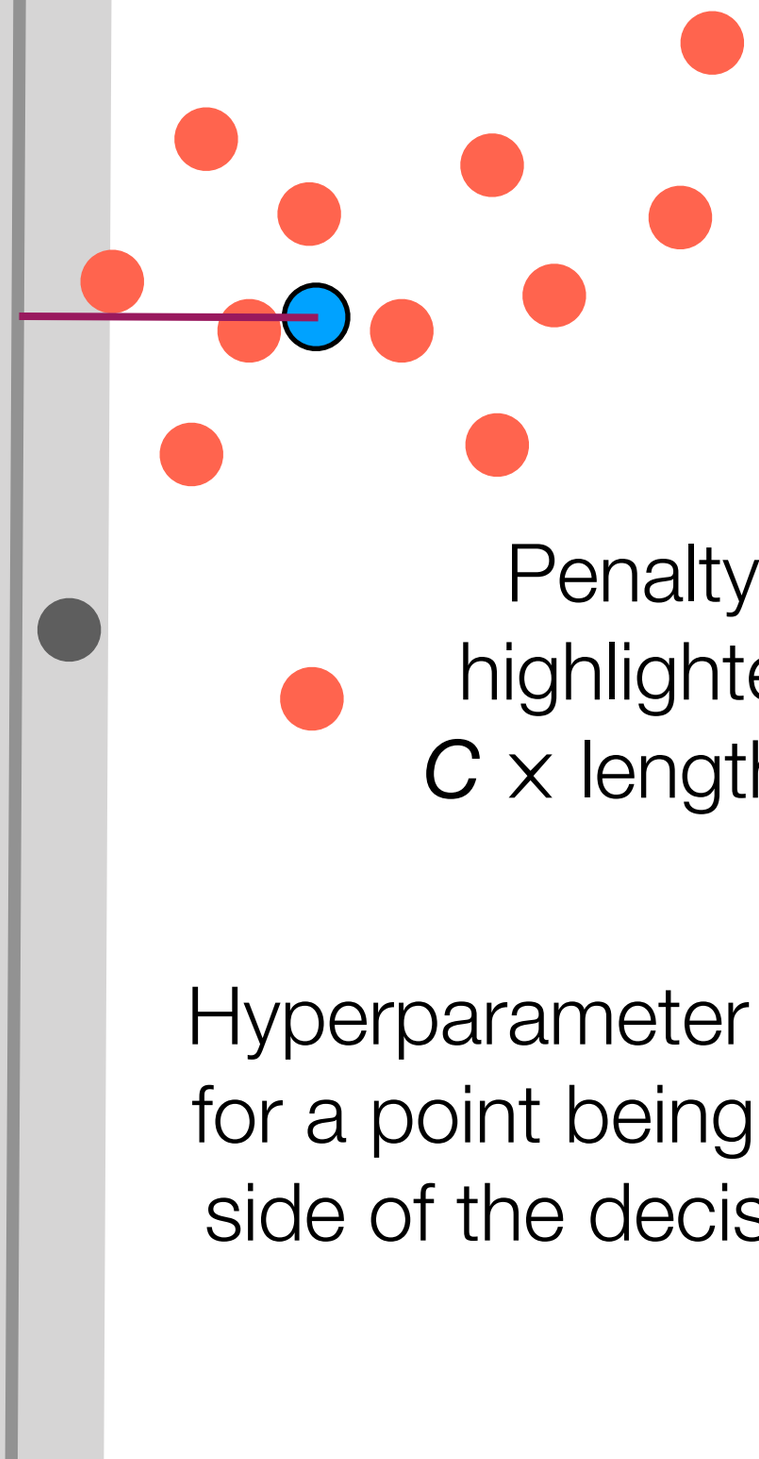
Penalty incurred for highlighted blue point: *C* x length of purple line

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

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
  - This is called **linear svm**

- Can instead use a different similarity function ("kernel" function) instead (popular choice: Gaussian kernel, also called "radial basis function" kernel)
  - This is called **kernel svm**

- Also: support vector *regression* (these are all in sklearn)