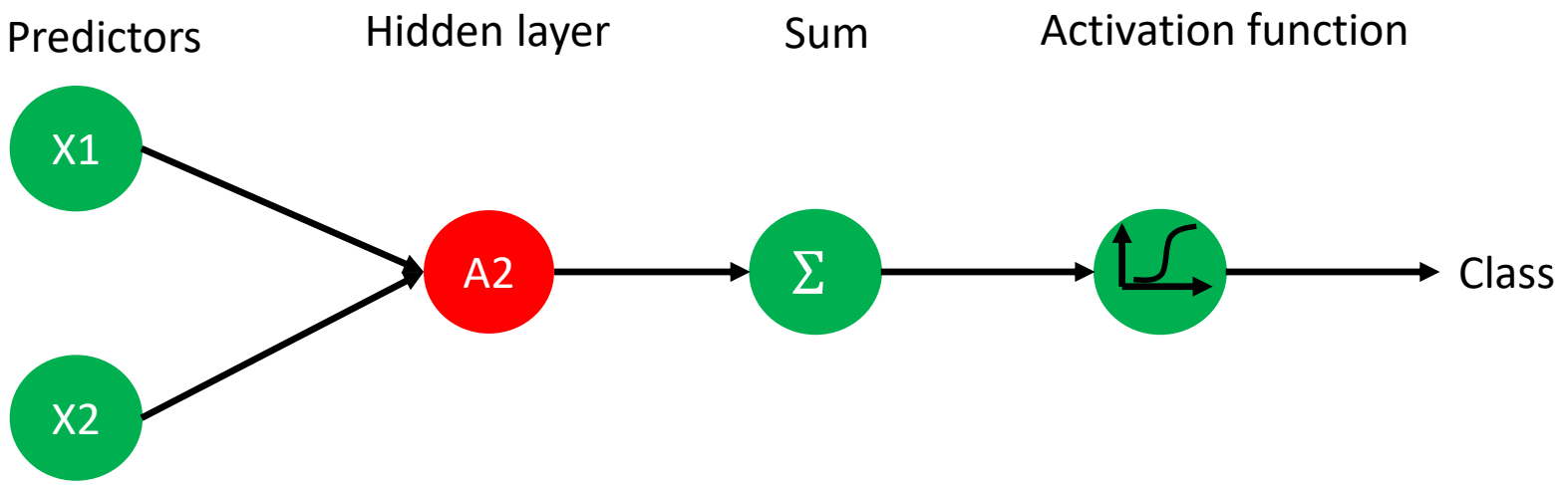
Logistic regression

Logistic regression

- A special case of a neural network with a single hidden node. A node is formed by the sum of the weighted combination of the predictors. An activation function is applied to produce the output



Lasso regression

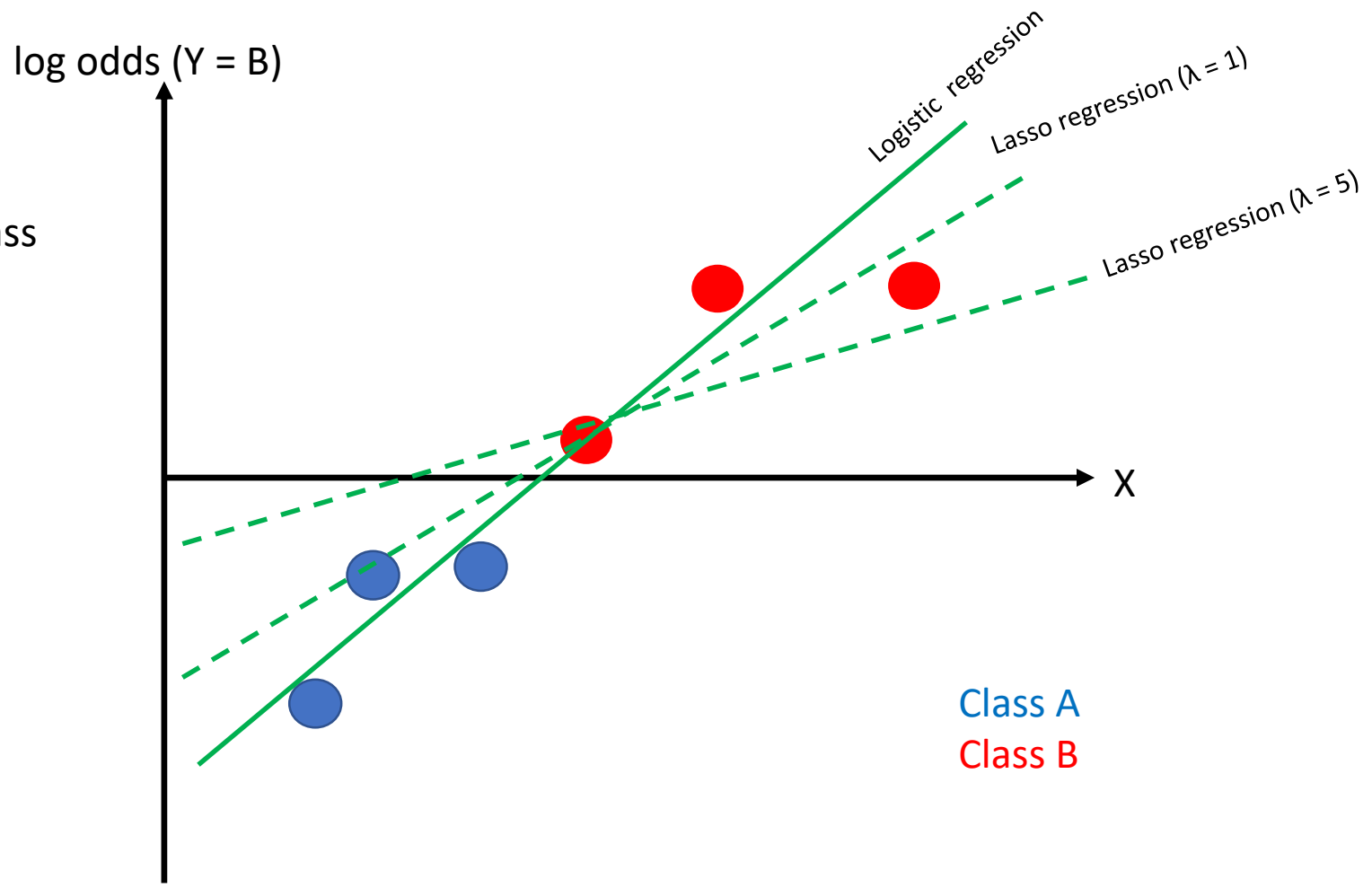
Logistic regression

Minimise likelihood of incorrect class prediction

Lasso regression

Minimise (likelihood + λ * |slope|)

- When λ = 0, it is equal to a logistic regression
- A higher λ results in a gentler slope, till it eventually becomes horizontal. A horizontal slope indicates a null relationship, and the predictor drops out
- Many λ values are repeatedly used to find the optimal that maximises prediction accuracy and minimise the variability of prediction

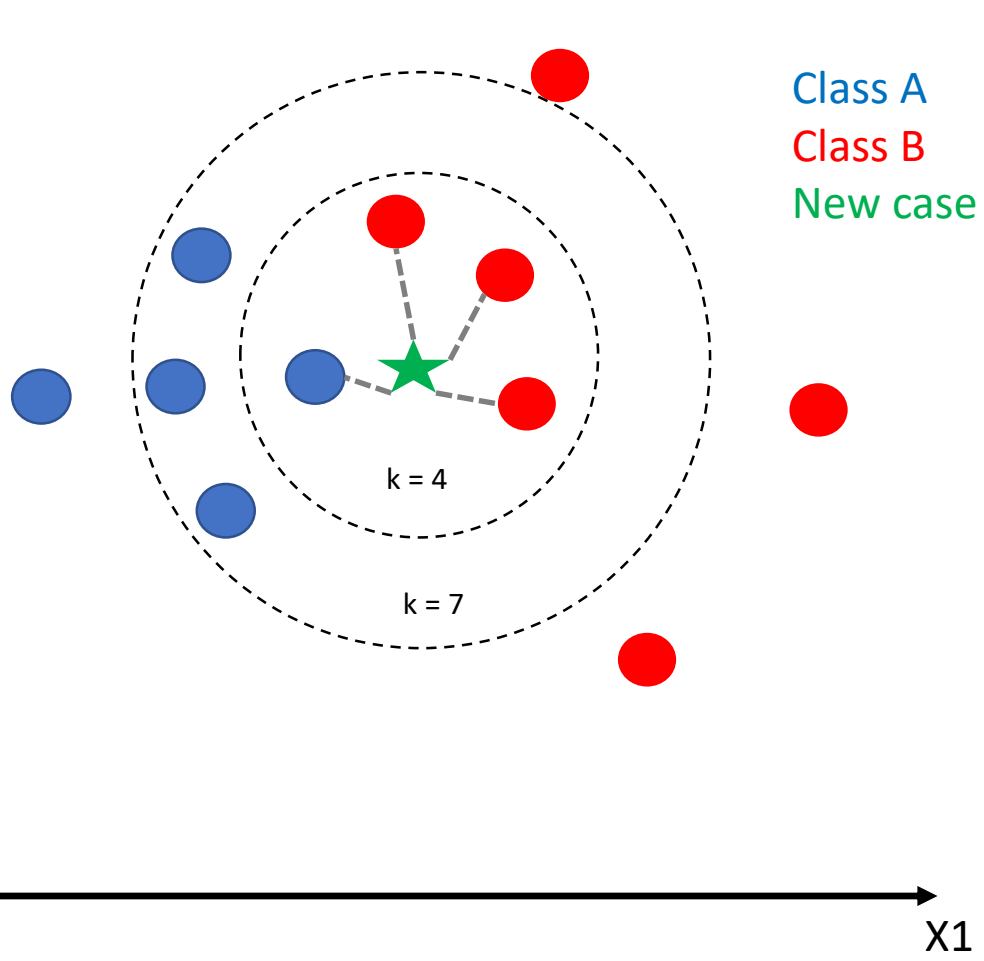


K nearest neighbour

K nearest neighbour

Predictions made for a new case by selecting the class of the majority observations to which it is closest

- Many k values are repeatedly used to find the optimal that maximises prediction accuracy and minimise the variability of prediction

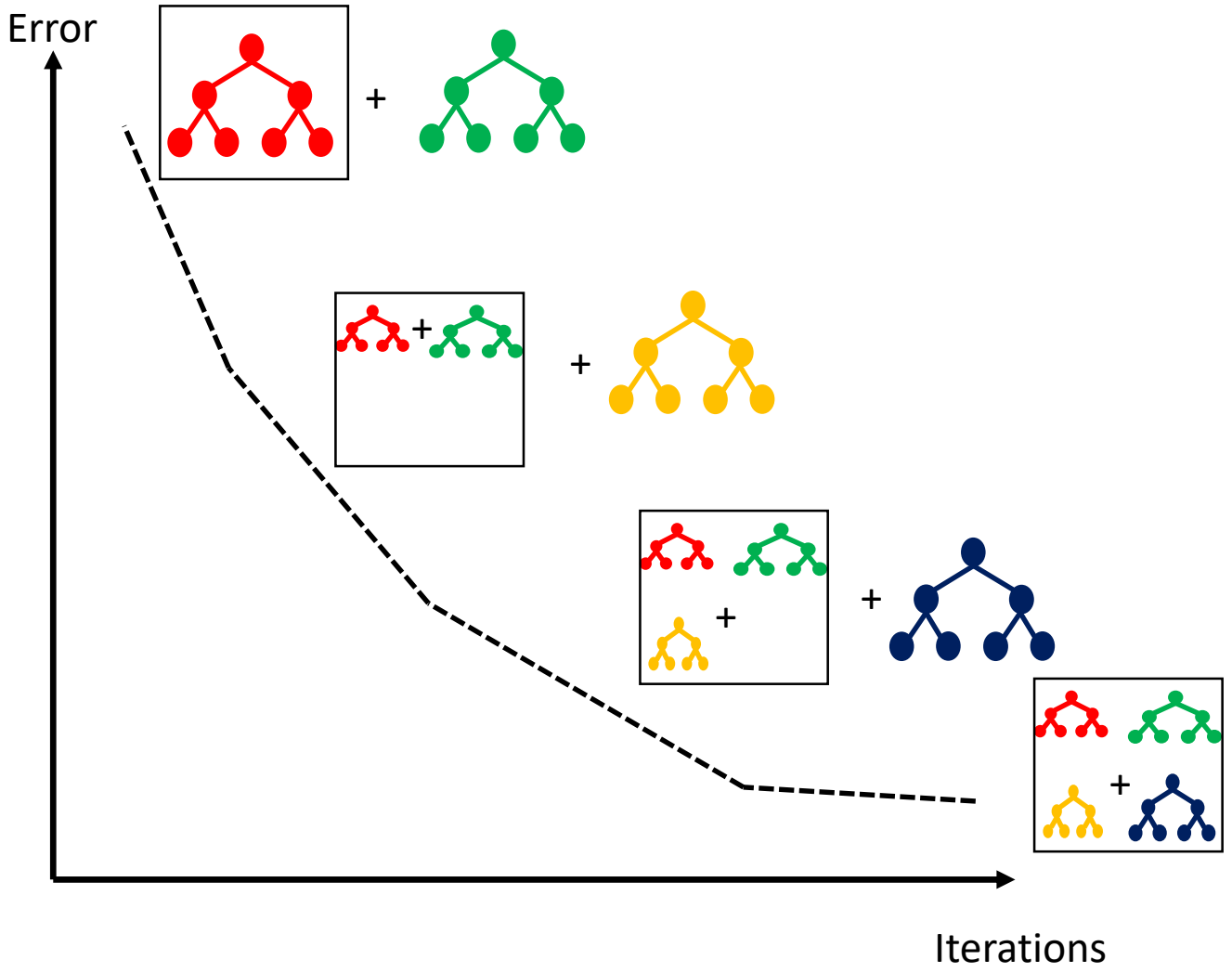


Gradient boosting

Gradient boosting machines (GBM)

Boosting iteratively adds a relatively simple tree model to compensate for the errors made by the previous models

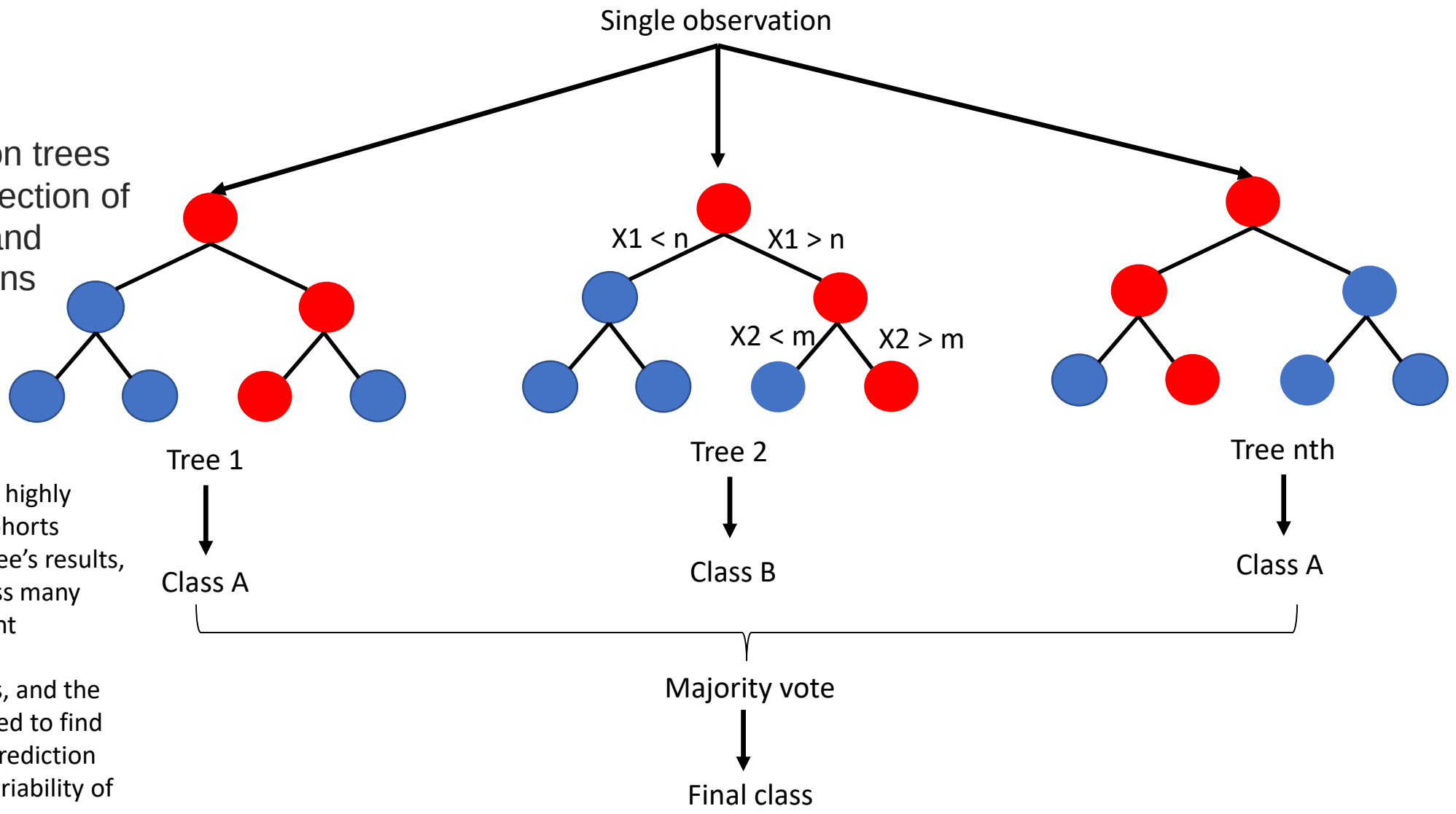
- At each iteration, a tree can include one or more predictors. Subsequent trees can include previously included predictors, but with different weights.
- Typically, the number of iterations (hence, number of trees) are tuned to find the optimal that maximises prediction accuracy and minimise the variability of prediction



Random forest

Random forest (RF)

RF builds many decision trees based on a random selection of cases and predictors, and aggregate the predictions



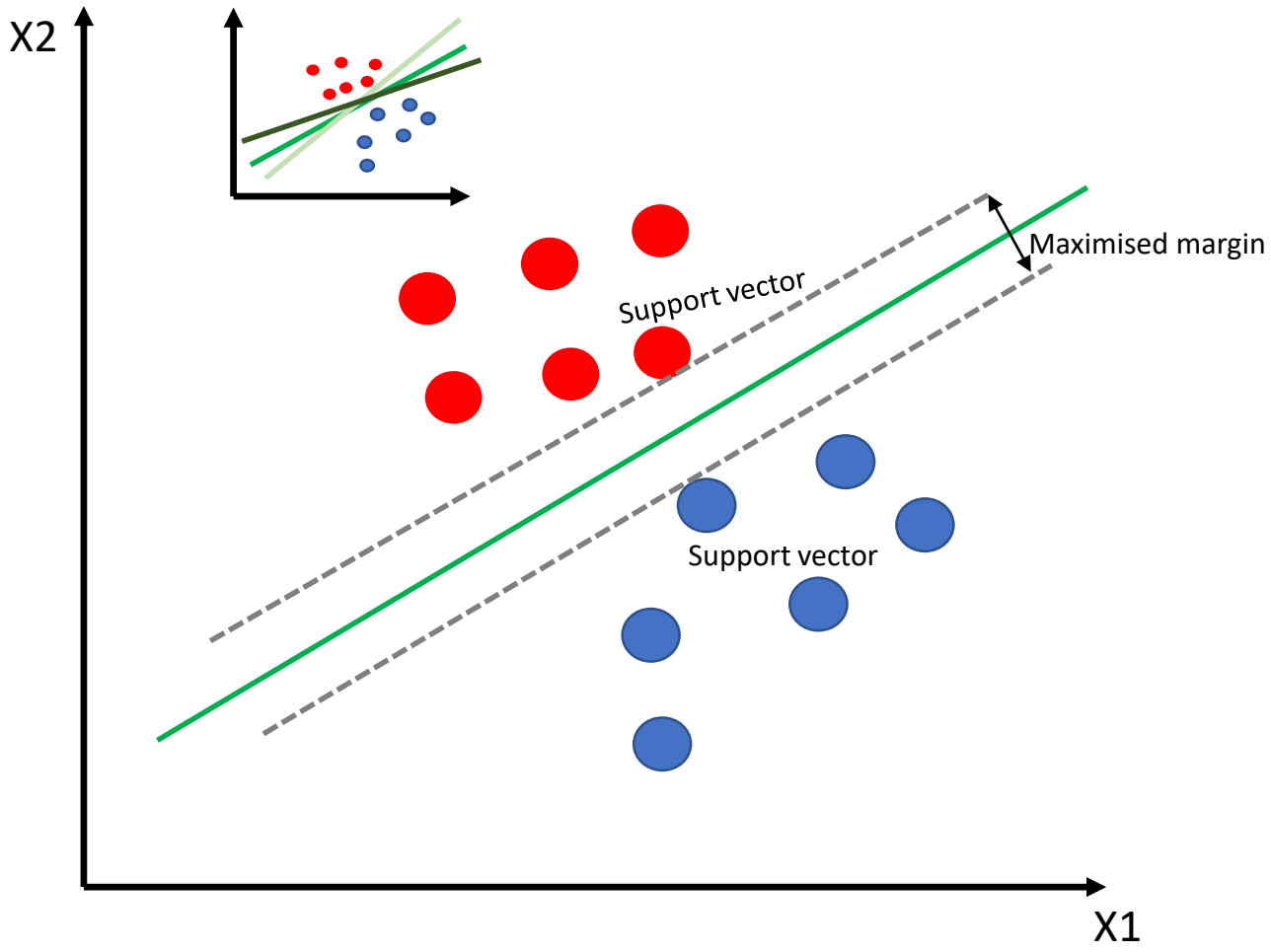
- A decision tree's prediction is highly variable between different cohorts
- Rather than relying on one tree's results, an "average consensus" across many trees provides more consistent predictions
- Typically, the number of trees, and the number of predictors are tuned to find the optimal that maximises prediction accuracy and minimise the variability of prediction

Support vector machine

Support vector machine (SVM)

SVM finds a separating line (or hyperplane) between data of two classes

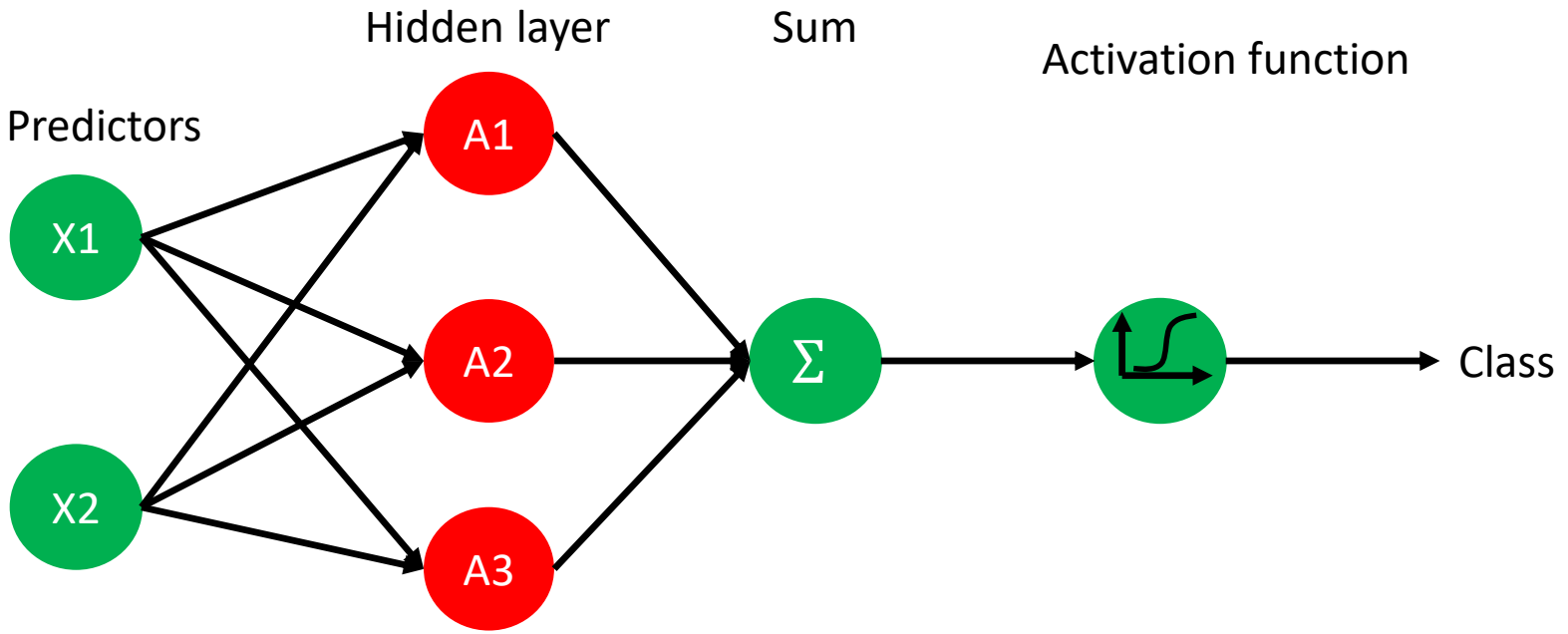
- Multiple lines can separate data of two classes (see insert).
- SVM finds the points closest to the line from both classes. These points are called support vectors. The distance between the line and the support vectors is computed, which is called the margin. SVM goal is to maximize the margin.
- The curvature of the line and the width of the margin are tuned to find the optimal that maximises prediction accuracy and minimise the variability of prediction



Neural networks

Neural network

A non-linear aggregate extension of simpler regression methods



- Each circle within the hidden layer is called a “node”. A node is formed by the sum of the nonlinear weighted combination of the predictors.
- All the nodes are summed, and an activation function applied to produce the output
- Typically, the number of nodes is tuned to find the optimal that maximises prediction accuracy and minimise the variability of prediction