Discriminative Classifiers

Goals

- Introduce 2 new classifiers
- Understand how to fit them
- Think about their drawbacks
- Contrast with logistic regression
- Just an intro, not an exhaustive list(!), notable omissions:
  - SVMs
  - Generative classifiers

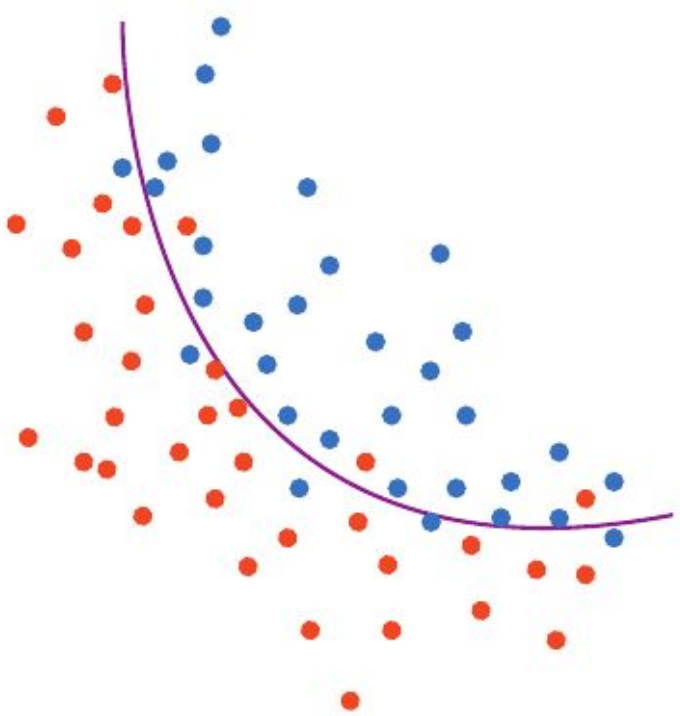
What are Discriminative Classifiers?

- Learn a function that directly maps from an input to a class

What are Discriminative Classifiers?

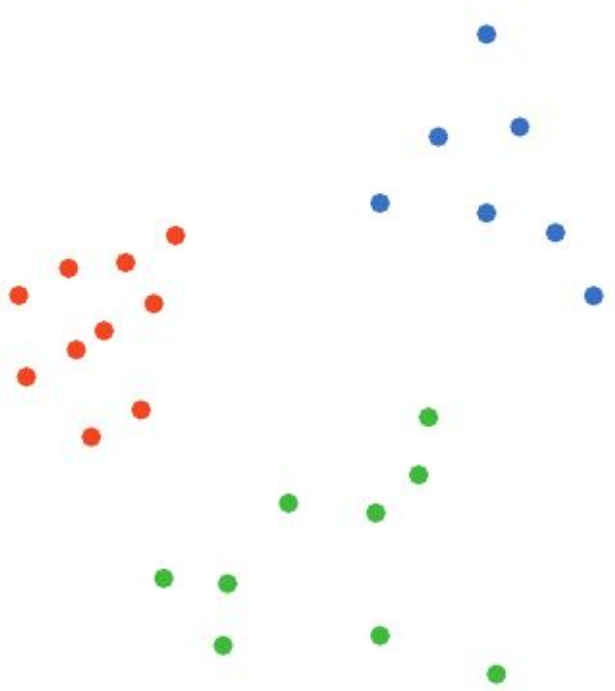
- Learn a function that directly maps from an input to a class
- This is equivalent to learning a boundary separating classes

What are Discriminative Classifiers?



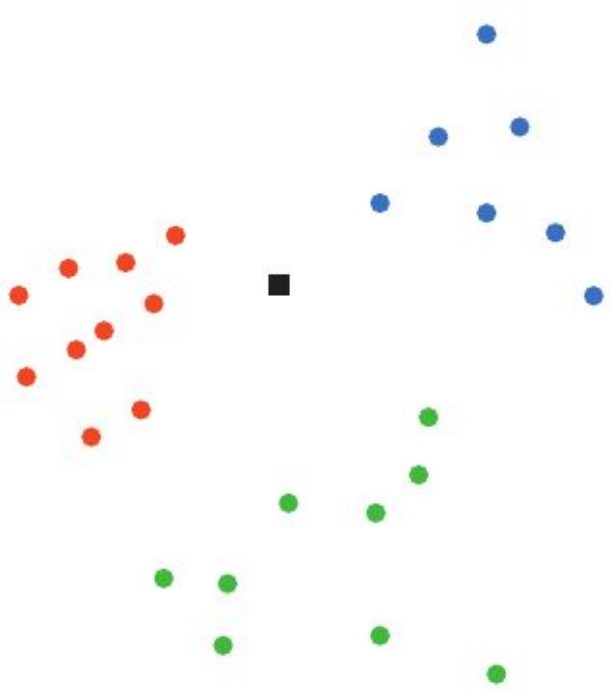
CAMBRIDGE SPARK

K-Nearest-Neighbours



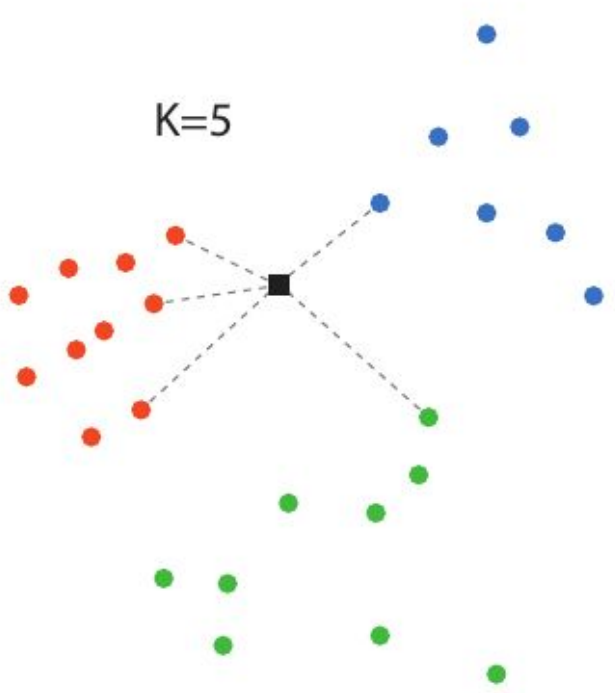
CAMBRIDGE SPARK

K-Nearest-Neighbours



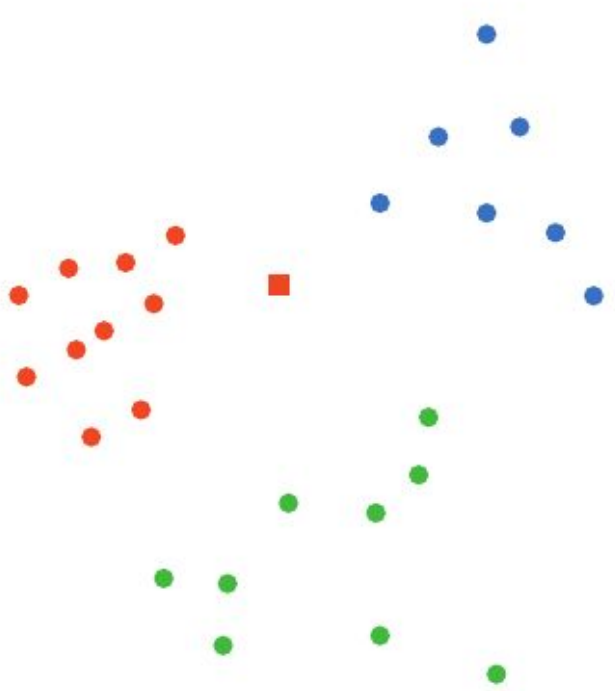
CAMBRIDGE SPARK

K-Nearest-Neighbours



CAMBRIDGE SPARK

K-Nearest-Neighbours



CAMBRIDGE SPARK

K-Nearest-Neighbours: Comments (1)

- Alternatively, neighbourhoods may be defined by a radius

K-Nearest-Neighbours: Comments (1)

- Alternatively, neighbourhoods may be defined by a radius
- In the binary classification case, keep K odd to avoid a tied vote

K-Nearest-Neighbours: Comments (1)

- Alternatively, neighbourhoods may be defined by a radius
- In the binary classification case, keep K odd to avoid a tied vote
- KNN is an example of instance-based learning

K-Nearest-Neighbours: Comments (2)

- The computational cost resides in the assignment of a class to a new point (req. computing the distance to all training points):

  O(TN(D + K))

  - **T**: number of test points to classify
  - **N**: number of training points
  - **D**: complexity associated with computing the distance bewteen two points
  - **K**: parameter of the KNN

Visualising KNN Decision Boundaries

- Rule of thumb:
  - **K↓** boundaries rougher → more local (risk overfitting)
  - **K↑** boundaries smoother → more global (risk underfitting)

CAMBRIDGE SPARK

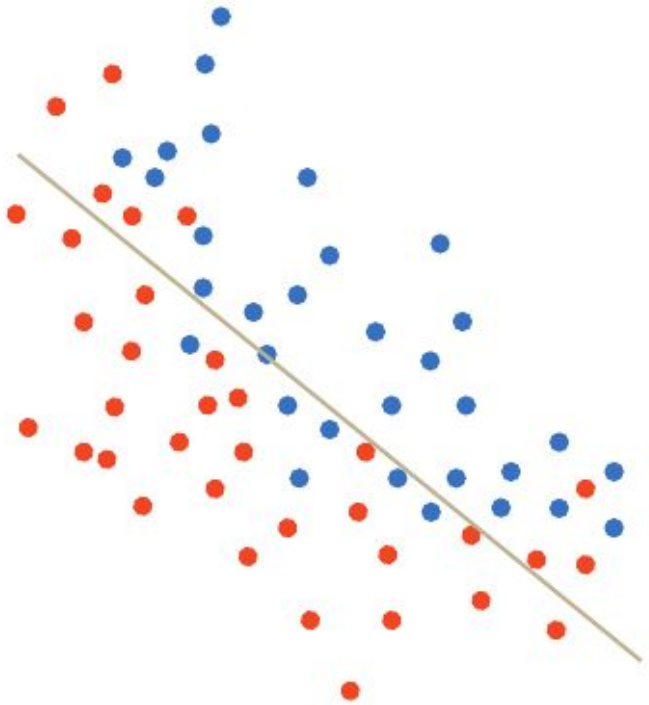
Visualising KNN Decision Boundaries

- Rule of thumb:
  - **K↓** boundaries rougher → more local (risk overfitting)
  - **K↑** boundaries smoother → more global (risk underfitting)

So… how can we pick a good K?

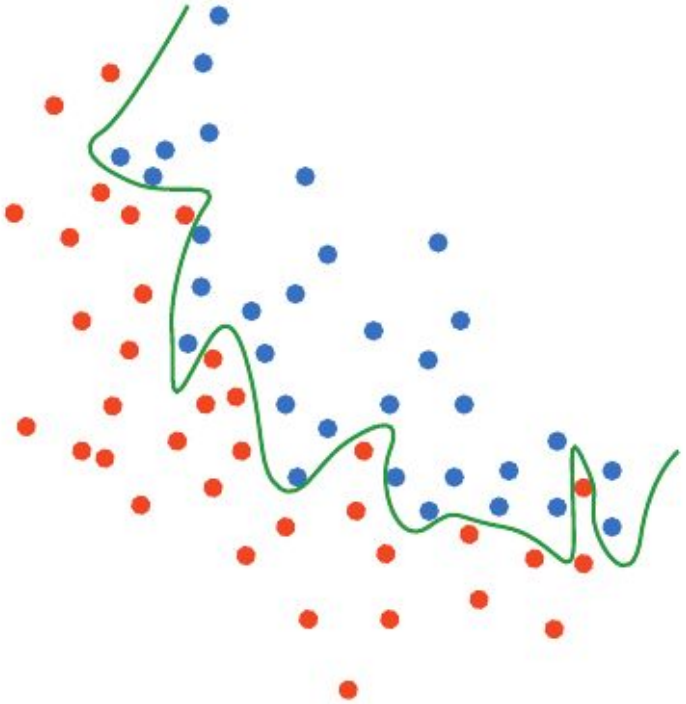
CAMBRIDGE SPARK

Overfitting & Underfitting



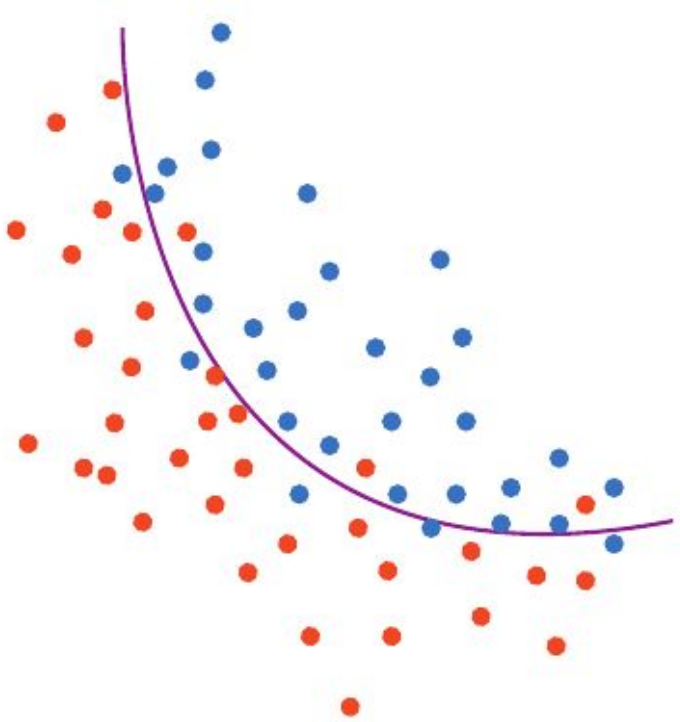
CAMBRIDGE SPARK

Overfitting & Underfitting



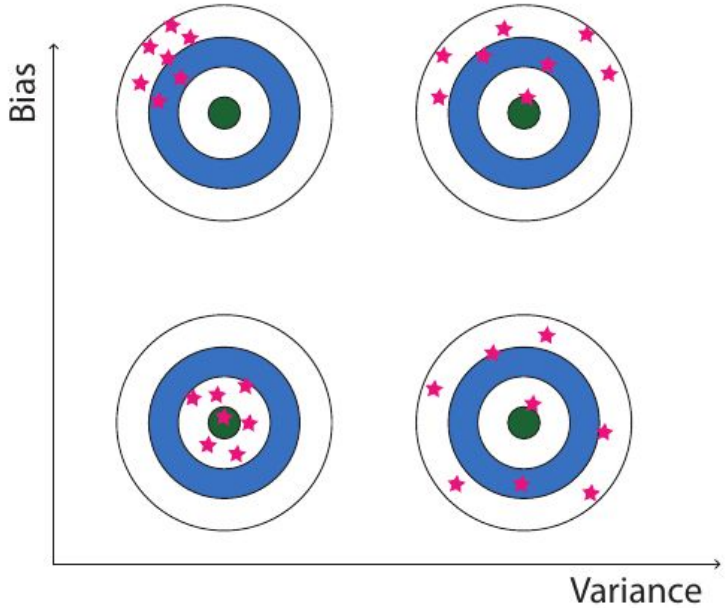
CAMBRIDGE SPARK

Overfitting & Underfitting



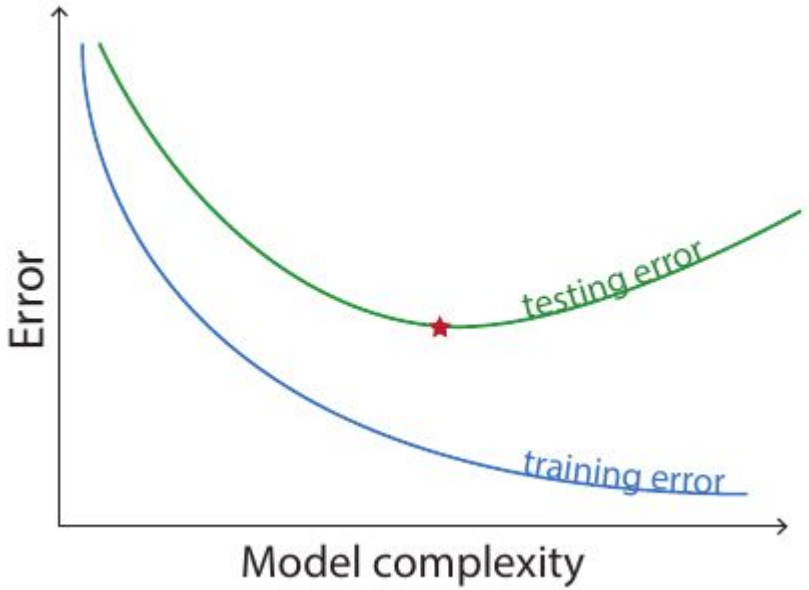
CAMBRIDGE SPARK

Bias & Variance



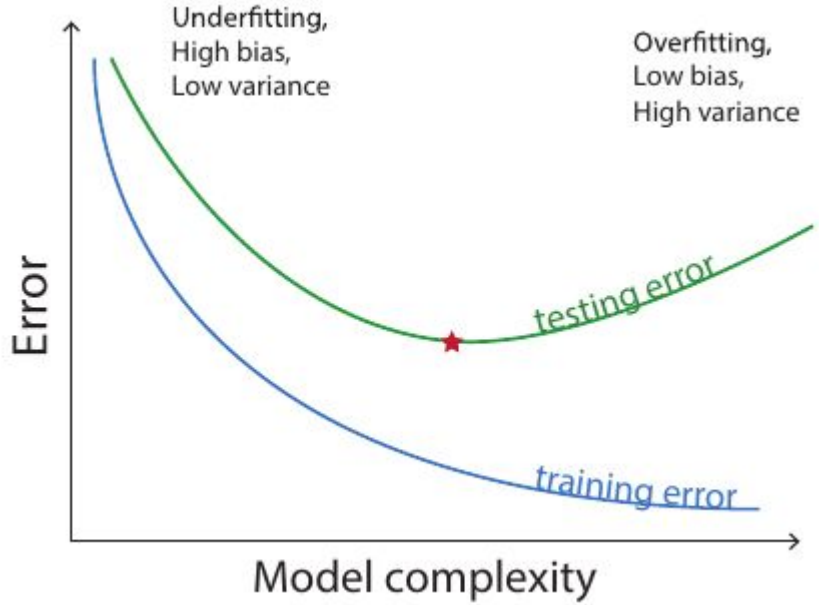
CAMBRIDGE SPARK

Bias & Variance



CAMBRIDGE SPARK

Bias & Variance



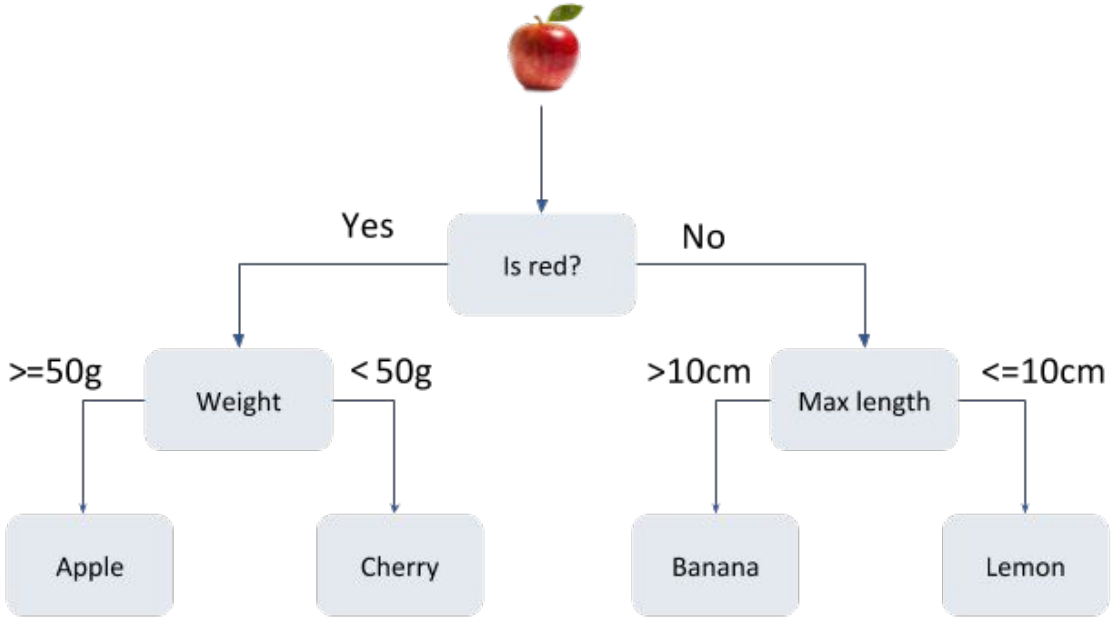
CAMBRIDGE SPARK

Hands-on session

knn-dtc.ipynb
(first half - 30 mins)

CAMBRIDGE SPARK

Decision Tree Classifiers (DTC)



CAMBRIDGE SPARK

Decision Tree Classifiers (DTC)

- A DTC is a model with a hierarchical structure. Each node corresponds to a feature and a split. To fit a DTC:

Decision Tree Classifiers (DTC)

- A DTC is a model with a hierarchical structure. Each node corresponds to a feature and a split. To fit a DTC:
    - At each level, consider all features and for each, the splitting point that best separates the classes

Decision Tree Classifiers (DTC)

- A DTC is a model with a hierarchical structure. Each node corresponds to a feature and a split. To fit a DTC:
  - At each level, consider all features and for each, the splitting point that best separates the classes
  - Take the feature that offers the best separation and continue

Decision Tree Classifiers (DTC)

- But what is separation?

Decision Tree Classifiers (DTC)

- But what is separation?
    - Different metrics can be used to measure separation

Decision Tree Classifiers (DTC)

- But what is separation?
    - Different metrics can be used to measure separation
    - Common ones are GINI impurity and information gain

GINI Impurity

- Assume there are k classes: 1, 2, …, k
- At a given node, P, there is a proportion p_1 of elements in class 1, p_1 of elements in class 2, etc

GINI Impurity

- Assume there are *k* classes: 1, 2, …, *k*
- At a given node, P, there is a proportion p_1 of elements in class 1, p_1 of elements in class 2, etc
- GINI Impurity at that node:

$$G(P) = 1 - \sum_{i=1}^{k} p_i^2$$

CAMBRIDGE SPARK

Fitting a tree with GINI Impurity

- At each node, the parent node *P* must be split into children nodes $c_1, \ldots, c_J$
- Each child node has n_j instances

Fitting a tree with GINI Impurity

- At each node, the parent node *P* must be split into children nodes $c_1, \ldots, c_J$
- Each child node has n_j instances
- So, the Gain for the split:

$$Gain = G(P) - \frac{\sum_j n_j G(c_j)}{\sum_j n_j} \sum_{i=1}^{k} p_i^2$$

CAMBRIDGE SPARK

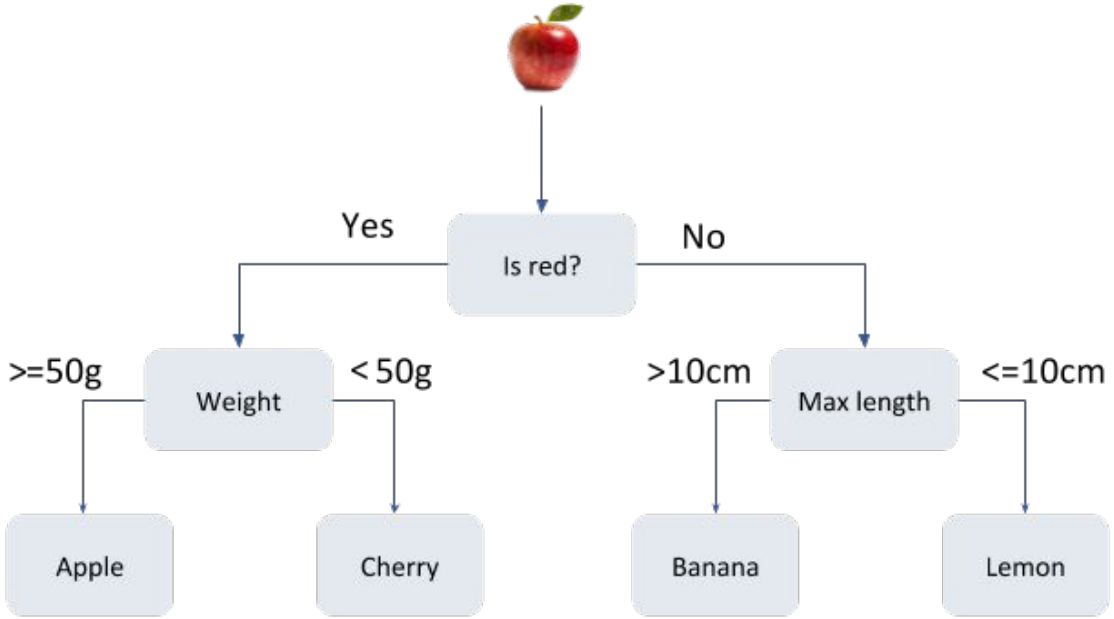
When to stop splitting?

- Pick a maximum depth

When to stop splitting?

- Pick a maximum depth
- Otherwise, we get one sample per “leaf node” → overfitting

Decision Tree Classifiers (DTC)



CAMBRIDGE SPARK

Hands-on session

knn-dtc.ipynb
(second half - 30 mins)

CAMBRIDGE SPARK