Classification trees: summary

How to build a classification tree:

**Recursive binary splitting**

1. Choose a stopping criterion. There are mainly two possibilities:

* Fix a minimum number of different labels (classes) for the subset of data given by the splitting. E.g. if with the i-th split we obtain that in the left branch the data can be divided only in two categories then do not split that branch.
* Fix a maximum depth for the tree.

Once the stopping criterion is reached, a final leave of the tree is obtained.

1. If the stopping criterion is true: stop

Else: going top-down, choose at every split the best variable to consider and the best way to split its levels (possible values it can get). Given the matrix of observations X and the set of labels **y**, determine the couple (i\*, t\*) such that:

(i⋆,t⋆) ← argmini,t E(y|xi≥t) + E(y|xi<t)

where E(y) is the classification error, that is the number of observation with label different form the most common, divided by the number of classes.

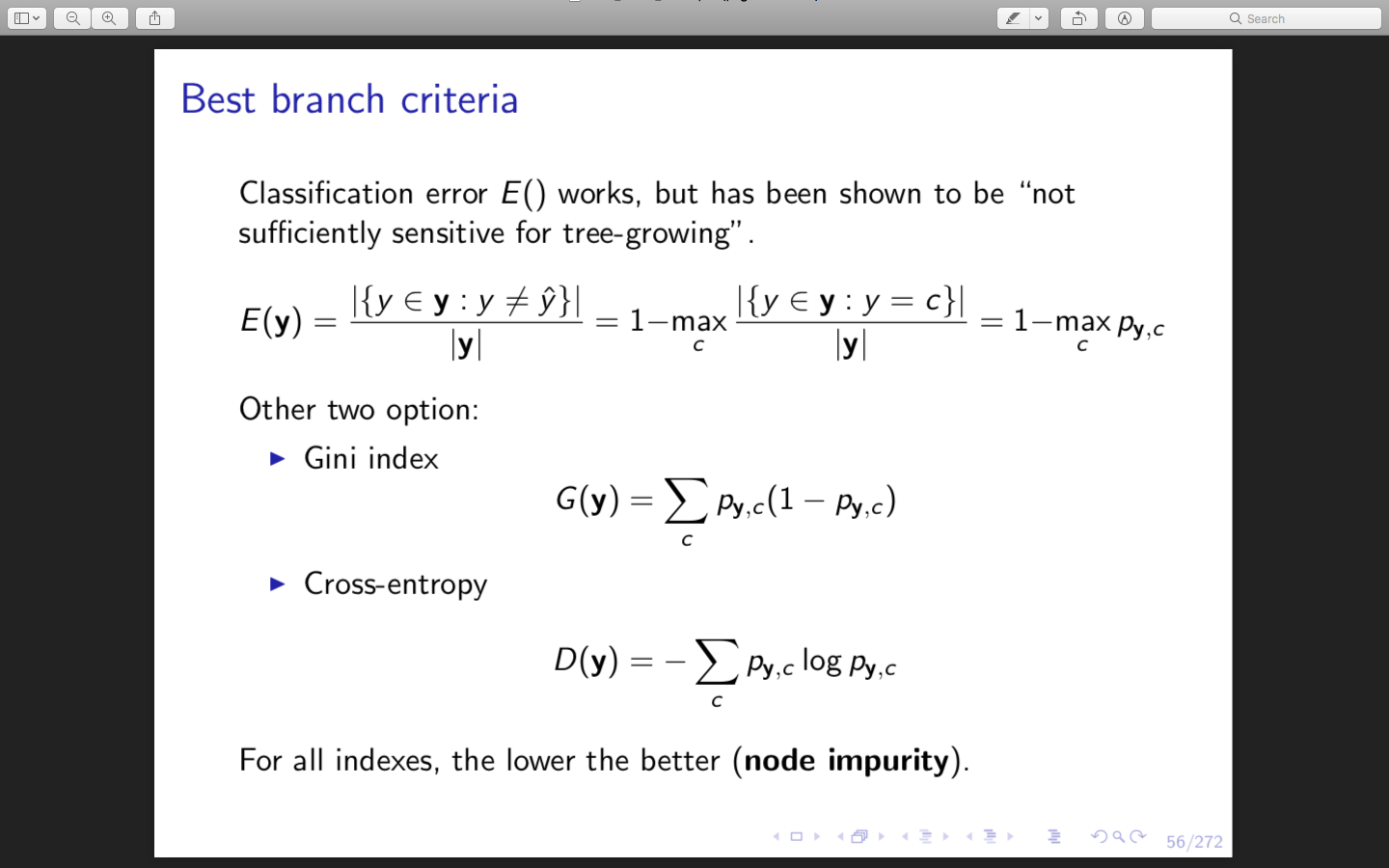
E(y) = |{y ∈**y** : y !=yˆ} / |**y**|

With yˆ = the most common class in y

1. On each of the branches obtained, repeat step 2 and 3.

This method is recursive and **greedy**, because at each step it is made to minimize the error in that step and not in the whole tree.

Actually, there are criteria for choosing the best branch that are better than the classification error:



(book, p.312)

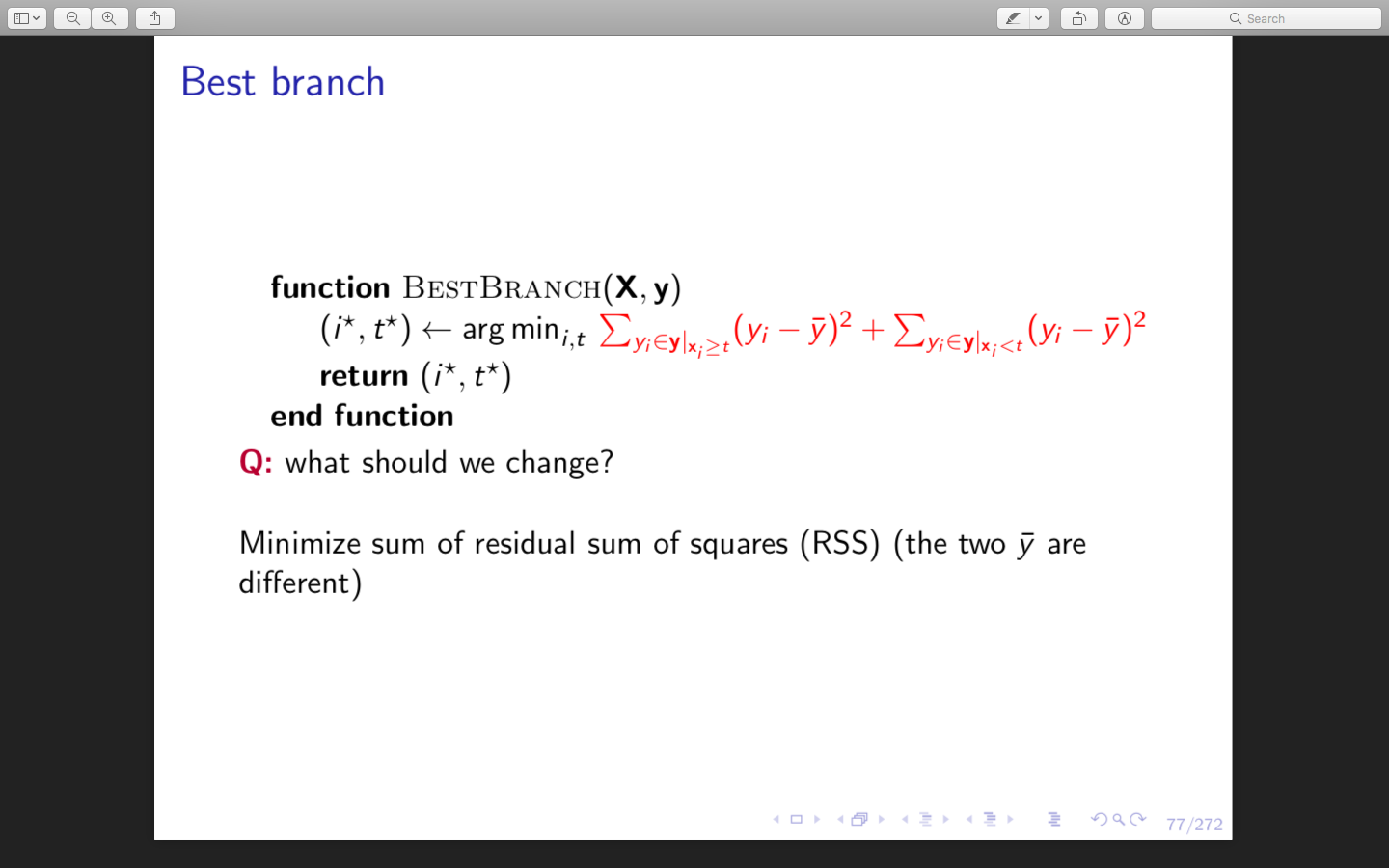
it is important to choose a good stopping criterion in order to get the best balance between bias and complexity. If the criterion is too strong we will have a very poor tree, not enough complex. But if it is too soft, we will have a very deep tree with high complexity which brings to overfitting.

To avoid overfitting, we have to choose a quite large Kmin, limit depth and do not split if low overall impurity decrease.

After building we have to **prune**:

First learn a full tree, then choose a sequence of subtrees such that every tree is a subtree of the previous and it is less complex. Then choose the tree in the sequence with the minimum k-fold cross validation error.

To build a regression tree the steps are the same but we have to exchange:

* The ‘most common class’ y^ with the mean response value
* The classification error in step 2 with the RSS: 
* The stopping criterion may be a minimum number observation per region or, again, the depth of the tree.