Classification in Context

Adapting to changes in class and cost distribution

Peter.Flach@bristol.ac.uk Peter A. Flach www.cs.bris.ac.uk/~flach/

Department of Computer Science

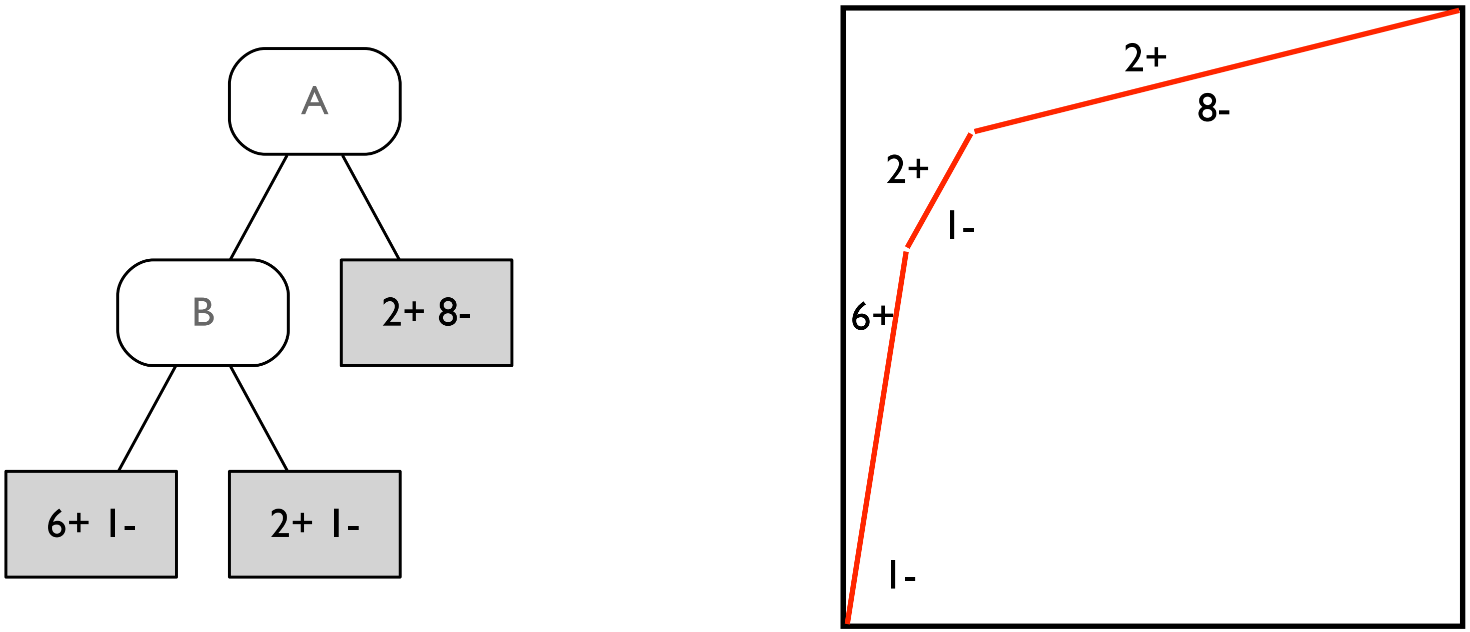
Intelligent Systems Laboratory

University of Bristol

Motivation and summary

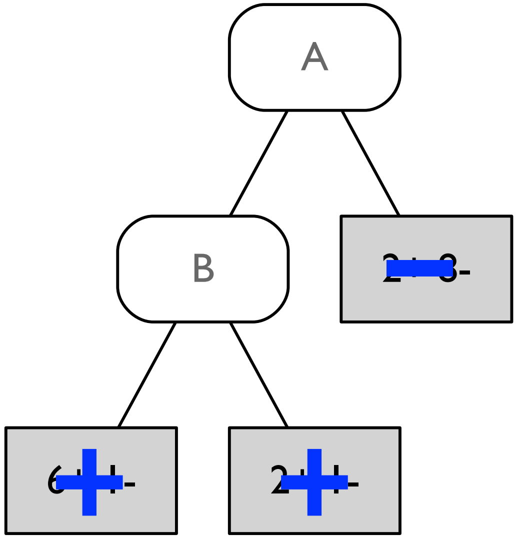
- In cost-sensitive classification, relative cost c (e.g., of misclassifying one positive relative to misclassifying one positive and one negative) provides a **context** in which loss is determined.
- This paper addresses two questions:
  - Q1: Can we reinterpret c in terms of class distributions?
    - A1: Yes, as **context change** from training to deployment. This allows us to recalibrate to such context changes.
  - Q2: What happens when we change the loss context more radically, by basing the loss on F-measure rather than accuracy?
    - A2: A lot. But we can still obtain scores calibrated to this new context.

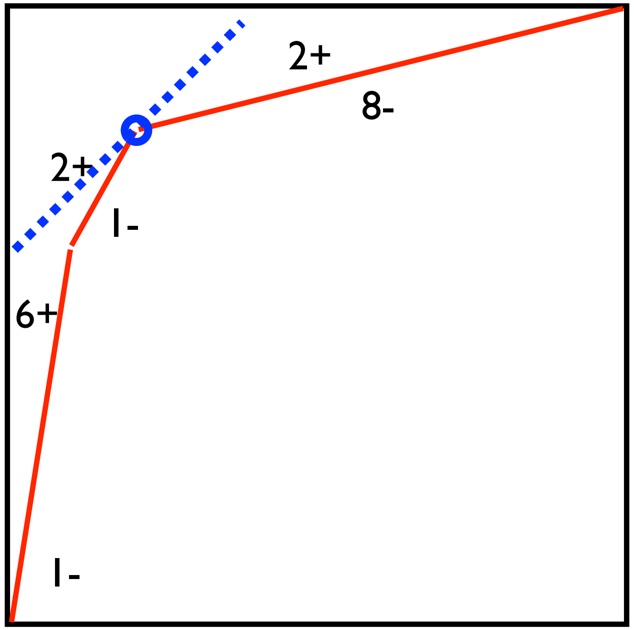
Decision tree and ROC curve



The ROC curve on the right has one segment for each leaf of the tree.
The curve is always convex on the training set.

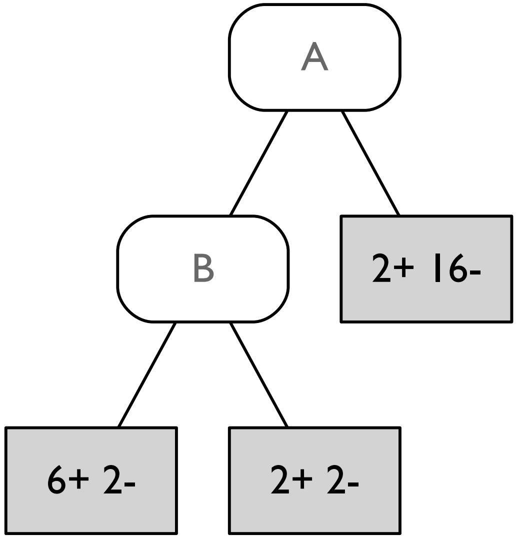
Majority class decision rule

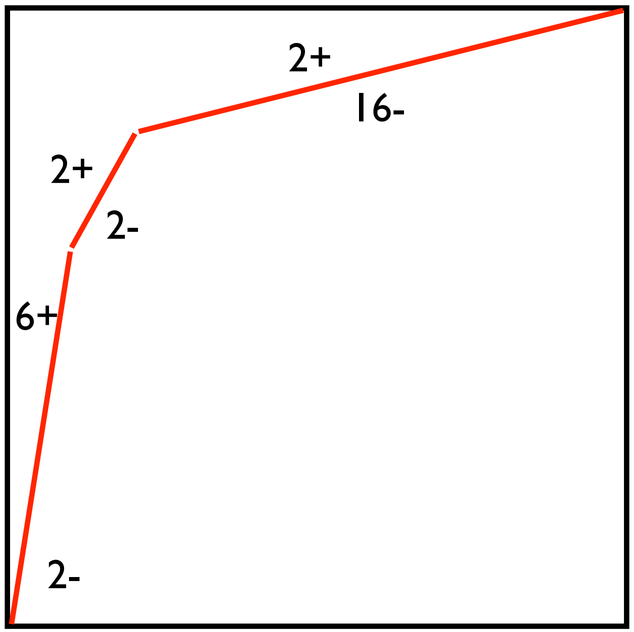




Labelling each leaf with its majority class achieves 80% accuracy.
This is optimal.

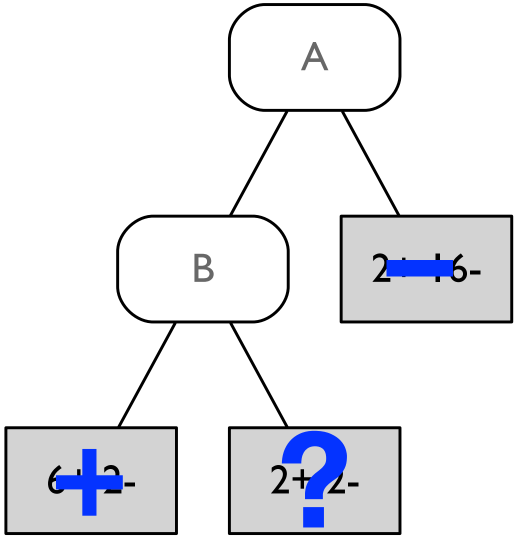
Adapting to deployment context (z=1/3)

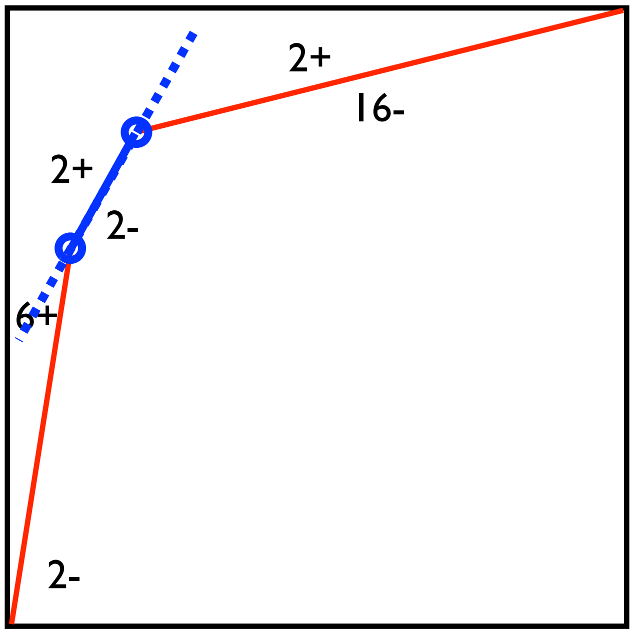




One way to represent the changed class distribution is by inserting another copy of each negative.
This does not change the shape of the ROC curve.

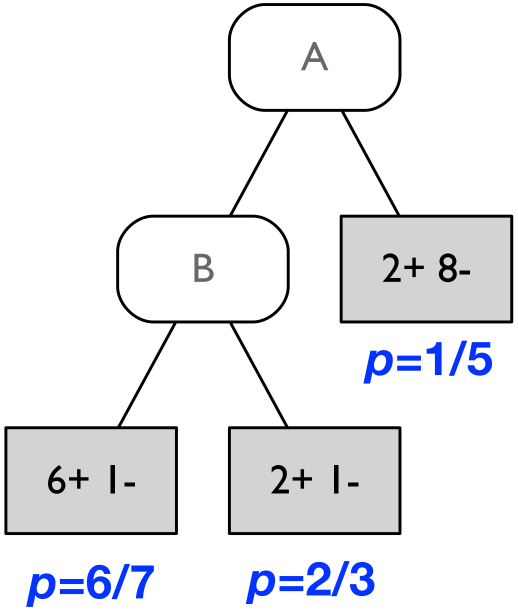
Sitting on the fence…

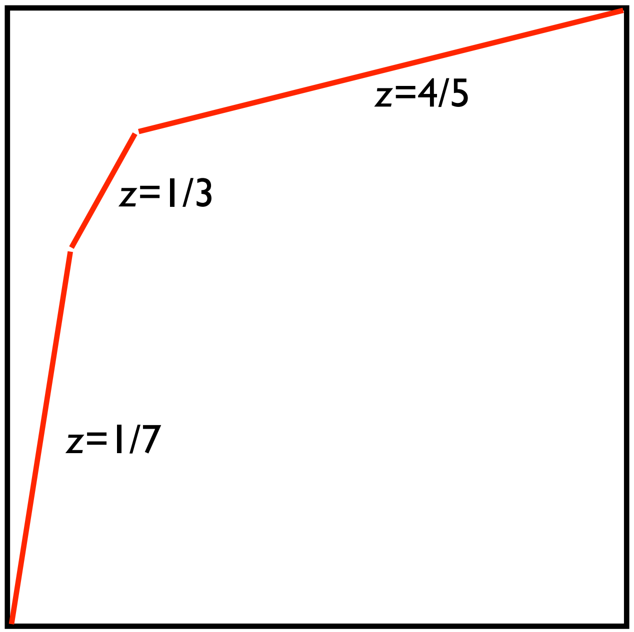




The middle leaf is now exactly on the decision boundary.
Regardless how we label it, we misclassify 6 of the 30 instances.

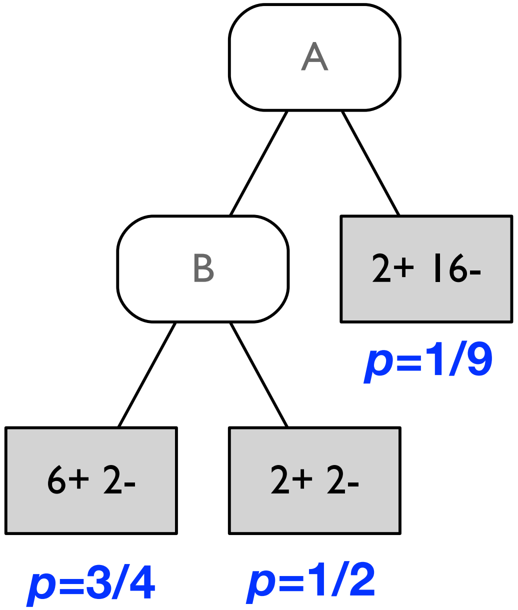
Calibration (for uniform prior)

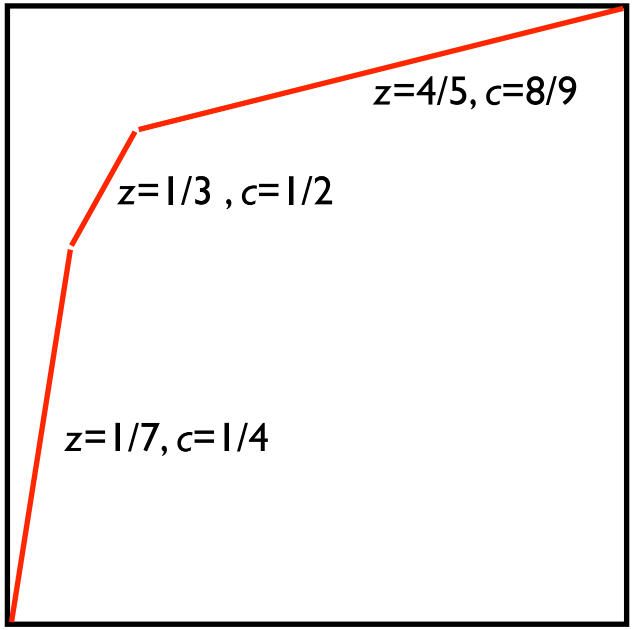




For each leaf, determine the value of z for which it is on the decision boundary and predict positive posterior p=(1-z).

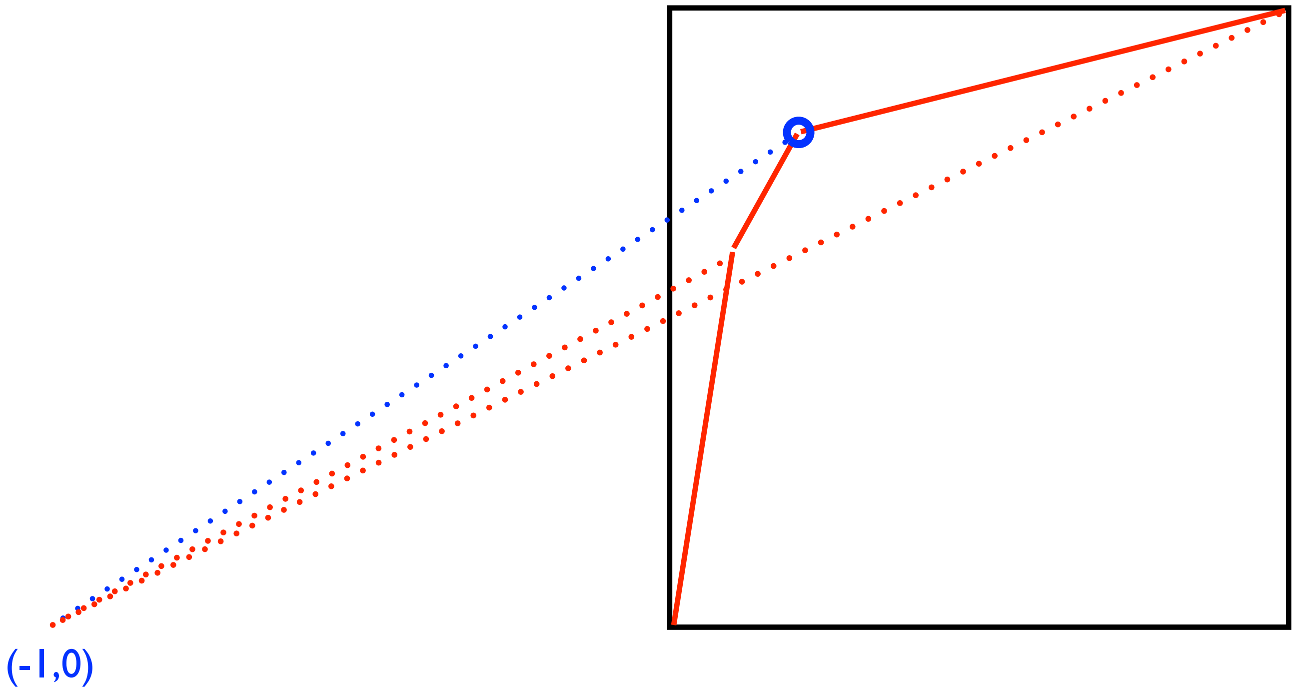
Calibration (for positive prior π=1/3)





For each leaf, determine the value of $c=(1-\pi)z/[\pi(1-z)+(1-\pi)z]$ for which it is on the decision boundary and predict positive posterior $p=(1-c)$.

What happens if we change loss to F-measure?



F-measure isometrics rotate around an imaginary point on the *x*-axis.
For uniform classes this point is (-1,0), which allows us to find the optimal point.
Is there a way to do this directly with scores, calibrated in some way?

Key insight: F-cost curves

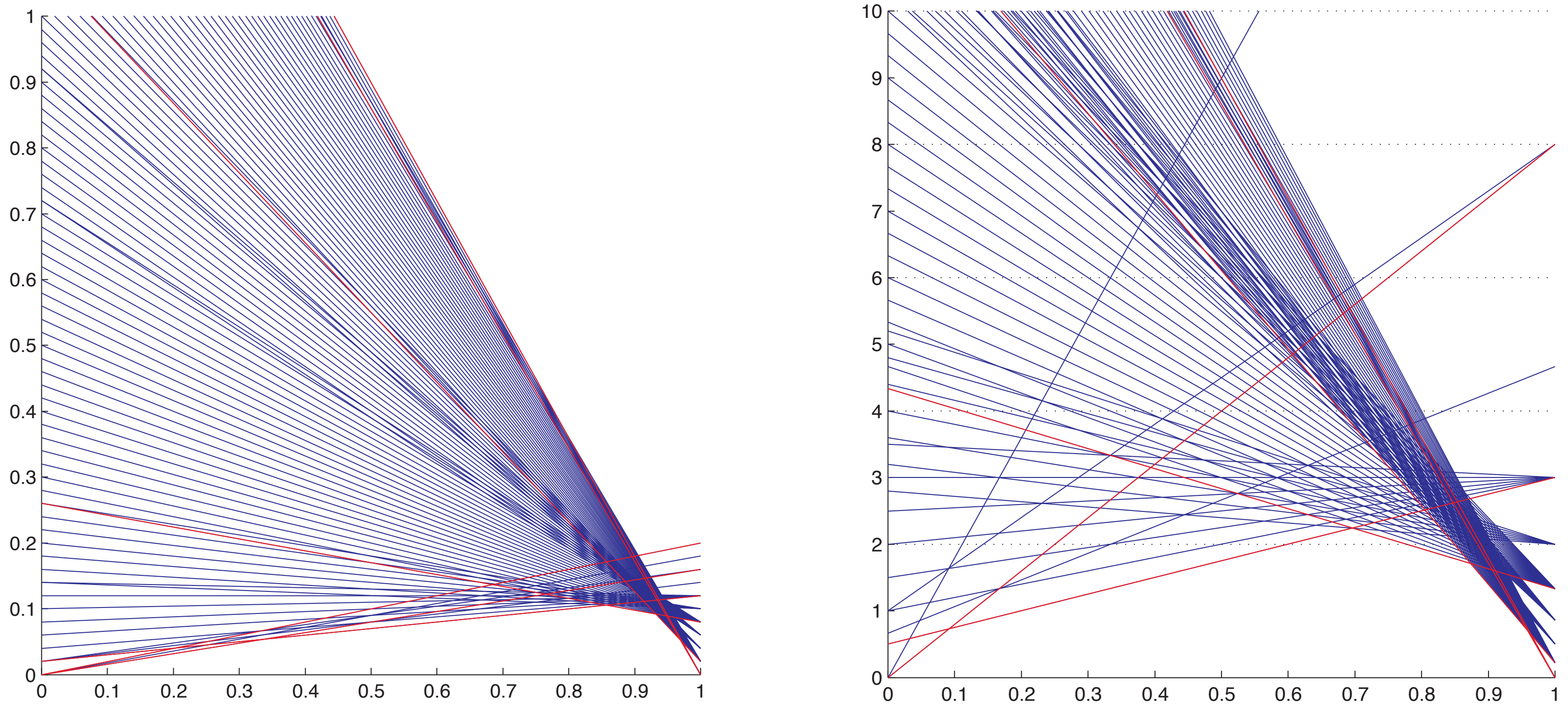
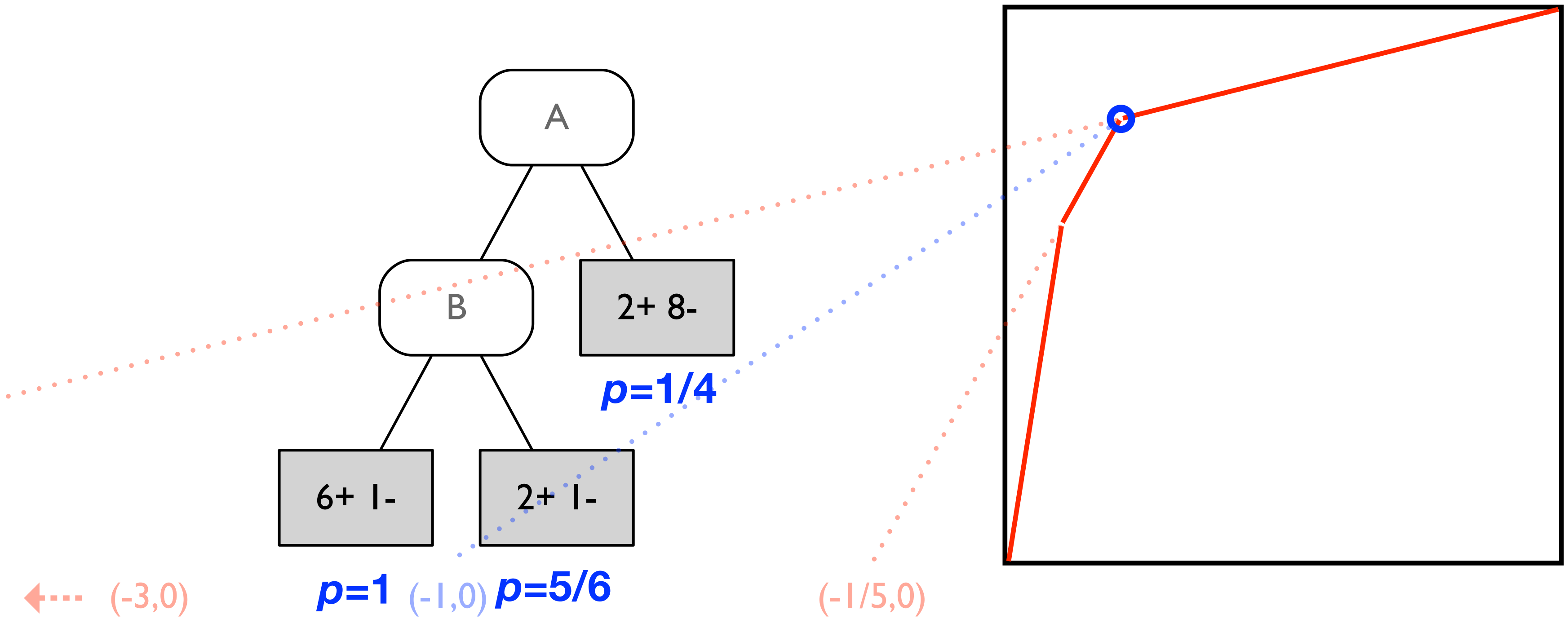


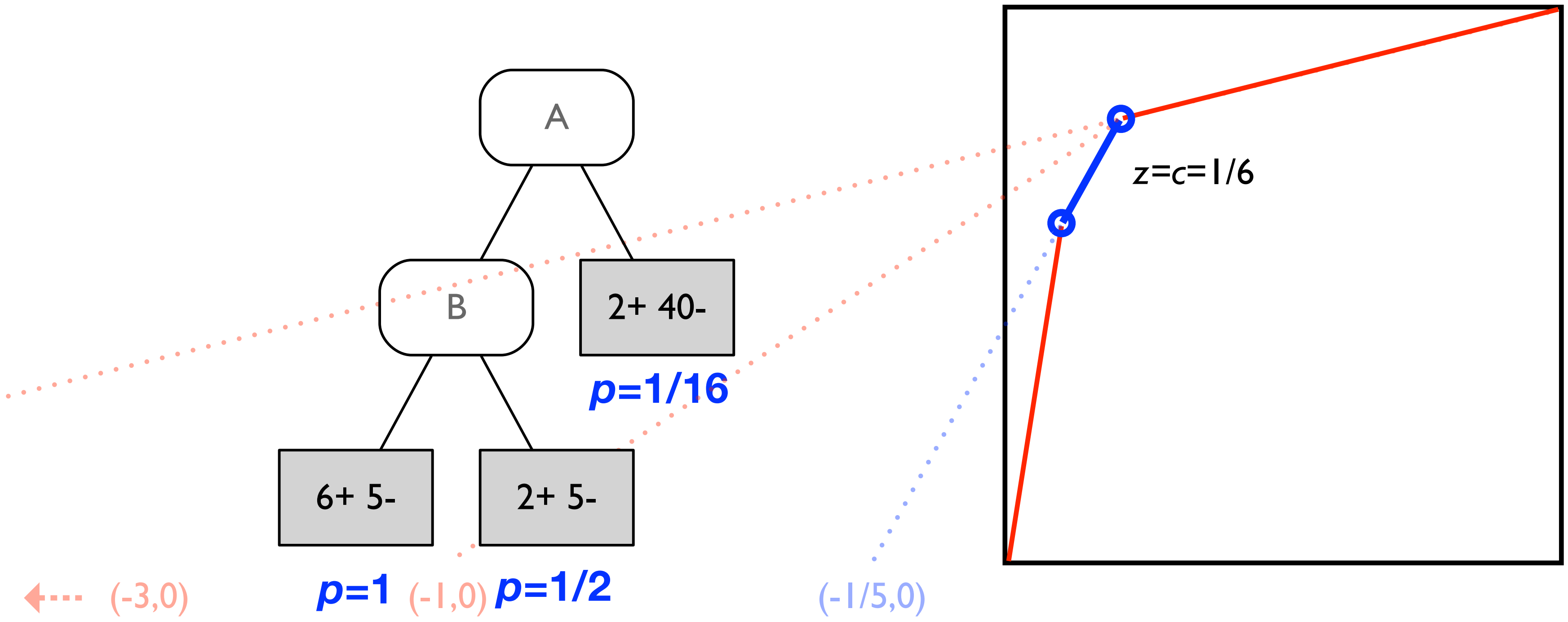
Fig. 3. (left) Accuracy-based cost lines. The x-axis shows c and the y-axis shows accuracy-based loss. The red cost lines correspond to operating points on the ROC convex hull. **(right)** Cost lines for F-measure loss. The y-axis shows $2FQ/(1-FQ)$. We can see that the optimal operating points are chosen for lower values of c, as expected.

Model calibrated for F-measure (1)



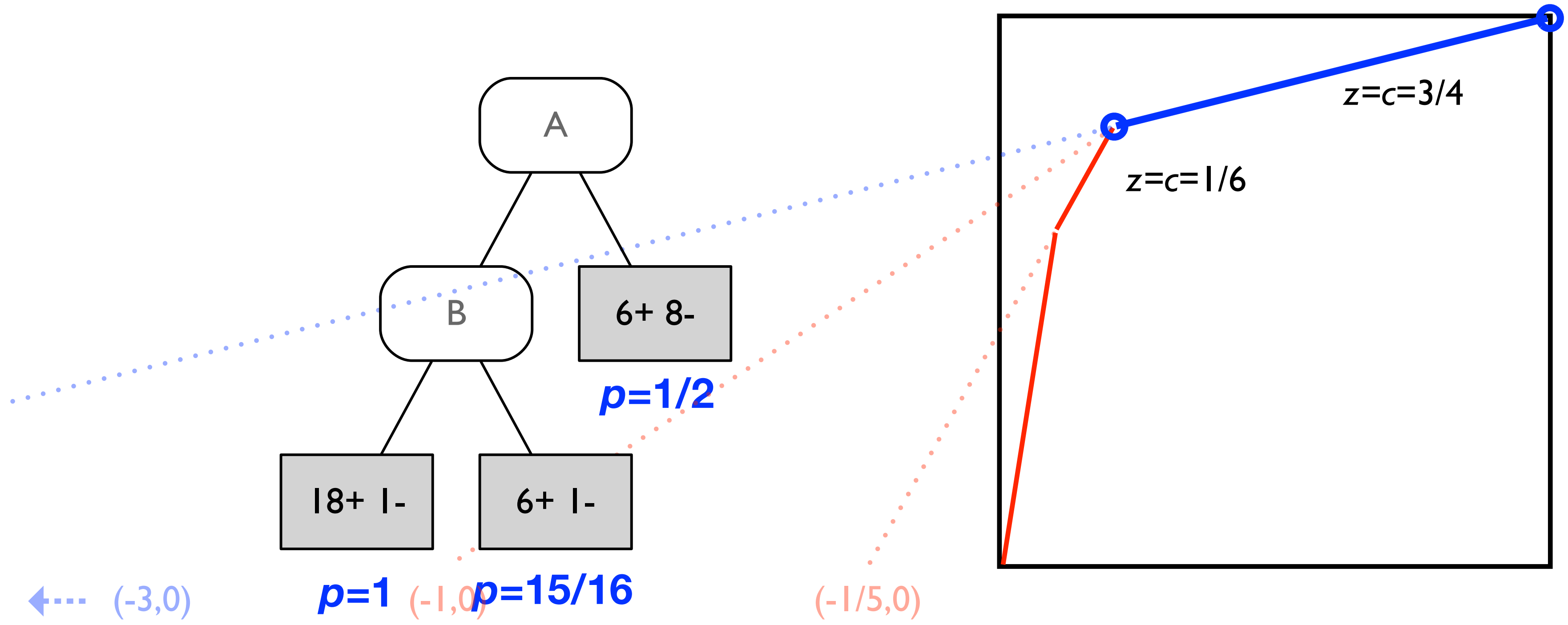
Calibrated F-measure scores for uniform classes.

Model calibrated for F-measure (2)



Decrease the positive-to-negative ratio to 1/5.

Model calibrated for F-measure (3)



Increase the positive-to-negative ratio to 3.

Concluding remarks

- One view of calibration is that a well-calibrated classifier aims to approximate the true posterior probabilities $p(Y \mid X)$.
  - Problem: $p(Y \mid X)$ is not useful in determining optimal thresholds for F-measure.
- The alternative view put forward here is that classifier scores are well-calibrated for loss measure Q and context C if
  - the threshold 1/2 is Q-optimal in context C;
  - more generally, the threshold (1-c) is Q-optimal in context C', where c is the context change.
- This naturally leads to a perspective where one classifier can output multiple scores for a single instance, each calibrated for a particular loss.
  - Q: does F-calibrated score have probabilistic interpretation?