Data Mining with Weka

Classification boundaries

Ian H. Witten

Classification boundaries

Weka's Boundary Visualizer for OneR

- ❖ Open iris.2D.arff, a 2D dataset
  - – *(could create it yourself by removing sepallength and sepalwidth attributes)*
- ❖ Weka GUI Chooser: Visualization>BoundaryVisualizer
  - – *open iris.2D.arff*
  - – *Note: petallength on X, petalwidth on Y*
  - – *choose rules>OneR*
  - – *check Plot training data*
  - – *click Start*
  - – *in the Explorer, examine OneR's rule*

Classification boundaries

Visualize boundaries for other schemes

❖ Choose lazy>IBk

- – *Plot training data*; click *Start*
- – *k = 5, 20; note mixed colors*

❖ Choose bayes>NaiveBayes

- – *set useSupervisedDiscretization to true*

❖ Choose trees>J48

- – *relate the plot to the Explorer output*
- – *experiment with minNumbObj = 5 and 10: controls leaf size*

Classification boundaries

❖ Classifiers create boundaries in instance space

❖ Different classifiers have different biases

❖ Looked at OneR, IBk, NaiveBayes, J48

❖ Visualization restricted to numeric attributes, and 2D plots