Data Mining with Weka

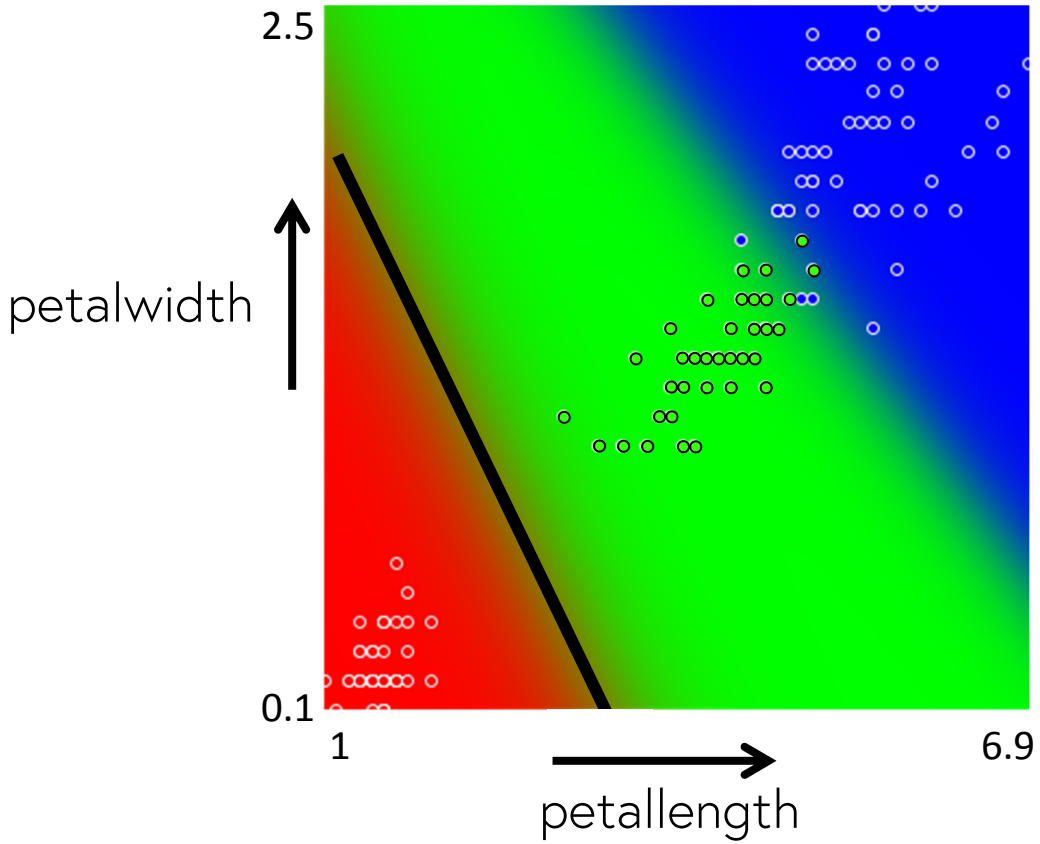
Support vector machines

Ian H. Witten

Support vector machines

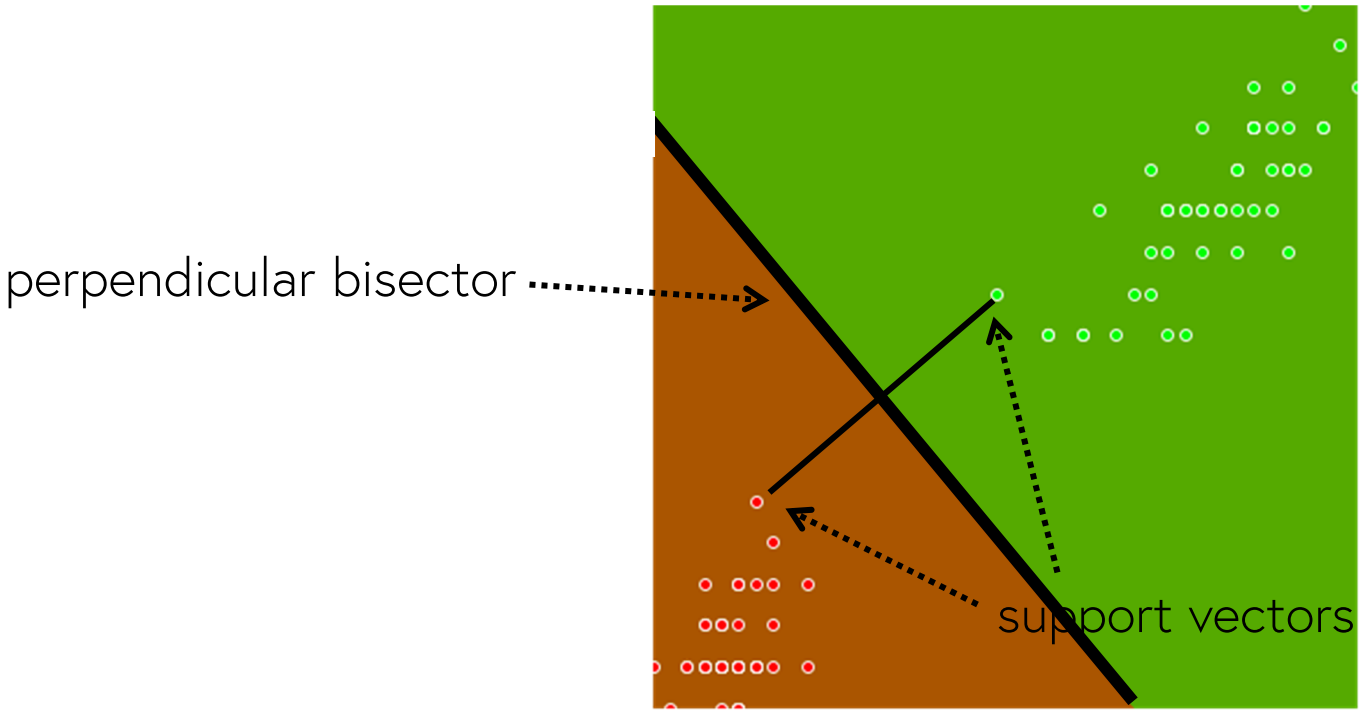
Logistic regression ⇒ linear boundaries

- Weka's boundary visualizer
  - *from the activity following "Nearest neighbor" lesson*



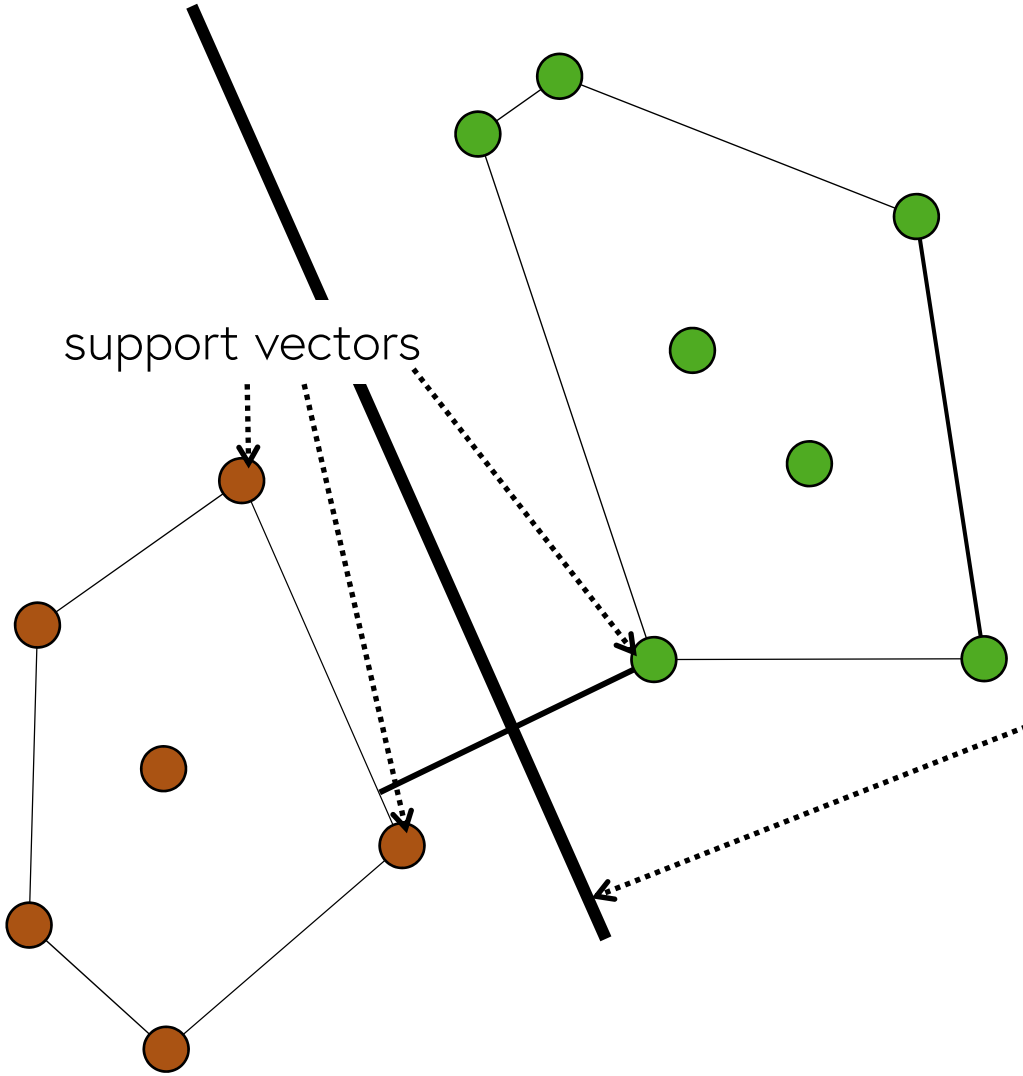
Support vector machines

Support vector geometry



Support vector machines

Maximum margin hyperplane



Support vectors define the boundary

All other instances can be deleted without changing it!

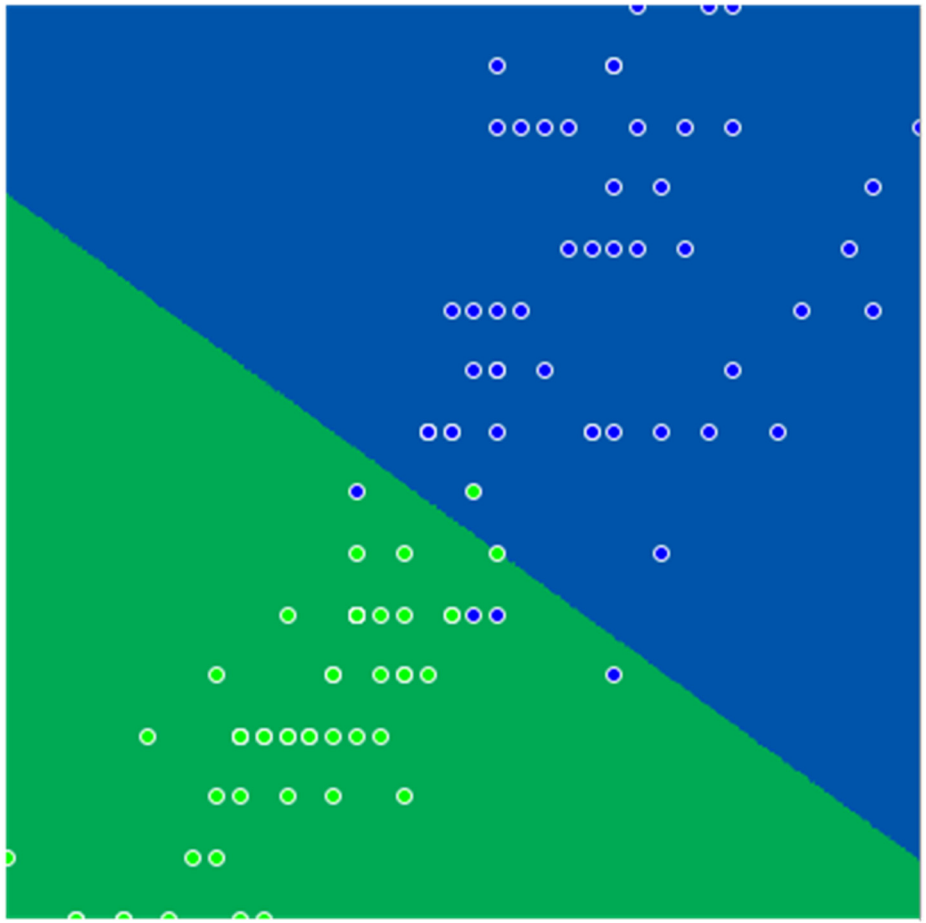
$$x = b + \sum \alpha_i y_i \mathbf{a(i)} \bullet \mathbf{a}$$

sum over support vectors **a(i)**

Support vector machines

Classes that are not linearly separable

(more complex)



Support vector machines

- Linear decision boundary
  - but can get more complex boundaries with the "Kernel trick"
- Very resilient to overfitting
  - boundary depends on a very few points
- Weka: functions>SMO
  - restricted to two classes
  - so use Multiresponse linear regression … or Pairwise linear regression
- Weka: functions>LibSVM
  - External library for support vector machines
  - faster than SMO, more sophisticated options