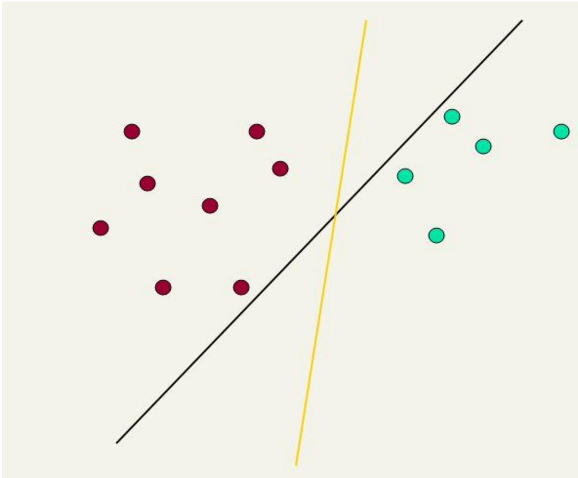
Support Vector Machine

Linear classifiers: Which Hyperplane?

- Lots of possible solutions for *a, b, c.*
- Some methods find a separating hyperplane, but not the optimal one [according to some criterion of expected goodness]
    - E.g., perceptron
- Support Vector Machine (SVM) finds an optimal* solution.
    - Maximizes the distance between the hyperplane and the "difficult points" close to decision boundary
    - One intuition: if there are no points near the decision surface, then there are no very uncertain classification decisions

This line represents the decision boundary:
$ax + by - c = 0$



Another intuition

- If you have to place a fat separator between classes, you have less choices, and so the capacity of the model has been decreased

Support Vector Machine (SVM)

- SVMs maximize the margin around the separating hyperplane.
    - A.k.a. large margin classifiers
- The decision function is fully specified by a subset of training samples, the support vectors.
- Solving SVMs is a quadratic programming problem
- Seen by many as the most successful current text classification method*

Maximum Margin: Formalization

- w: decision hyperplane normal vector
- $\mathbf{x}_i$: data point i
- y_i: class of data point i (+1 or −1)   NB: Not 1/0
- Classifier is: $f(\mathbf{x}_i) = \text{sign}(\mathbf{w}^T\mathbf{x}_i + b)$
- Functional margin of $\mathbf{x}_i$ is: $y_i (\mathbf{w}^T\mathbf{x}_i + b)$
  - But note that we can increase this margin simply by scaling w, b….
- Functional margin of dataset is twice the minimum functional margin for any point
  - The factor of 2 comes from measuring the whole width of the margin