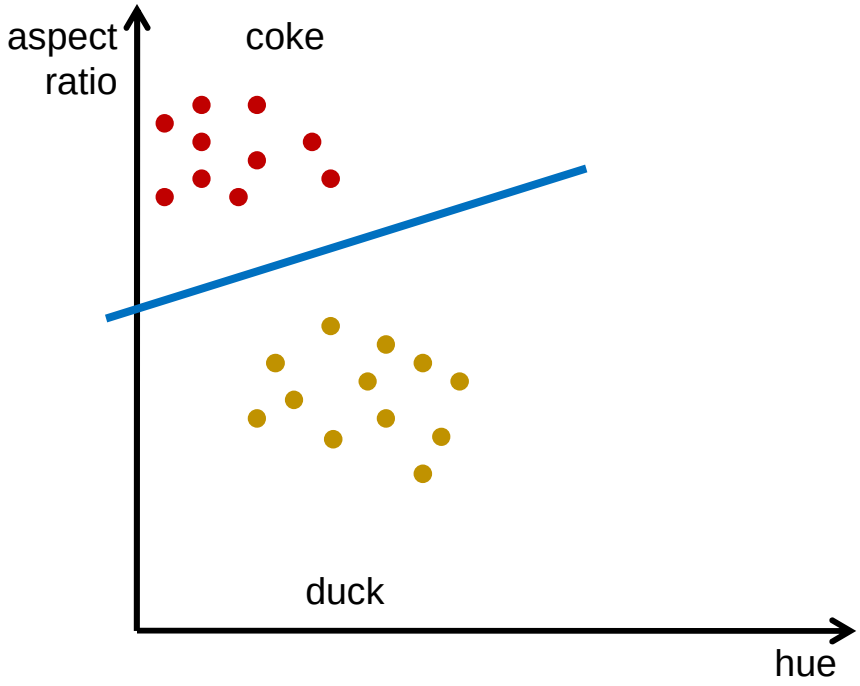
Slide 6

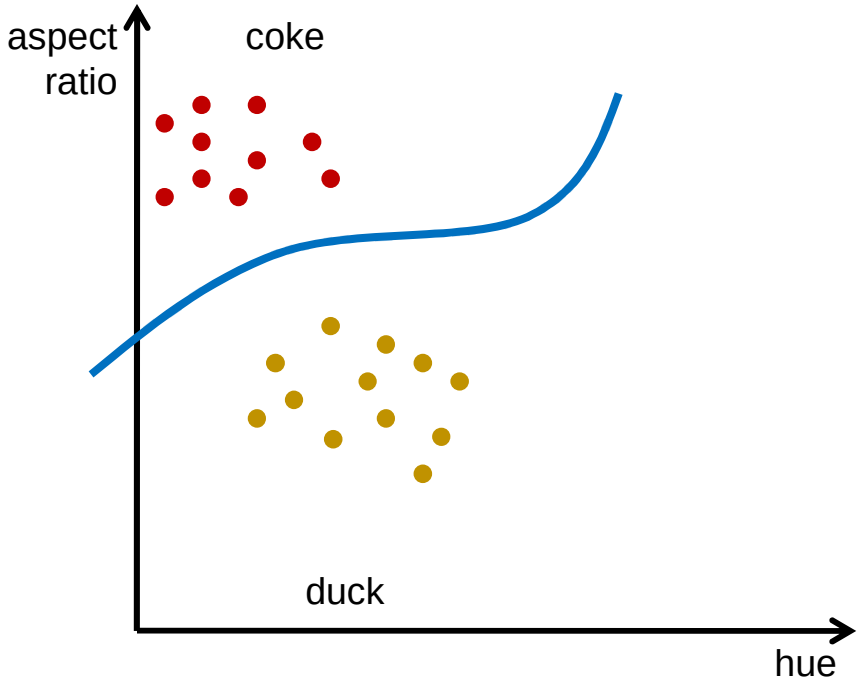
Classification cont.

- many approaches for decision rules and learning
  - linear classification



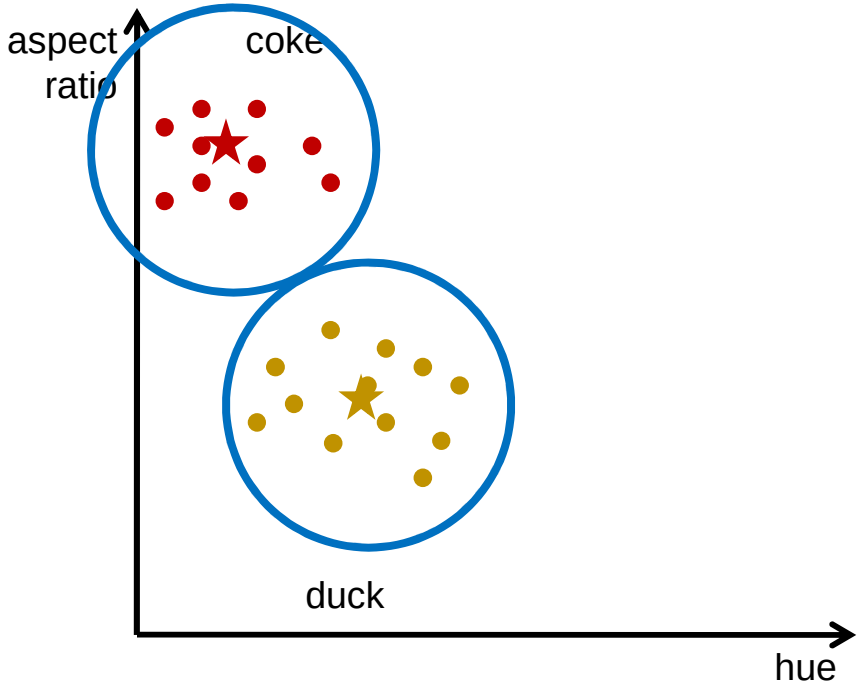
Classification cont.

- many approaches for decision rules and learning
  - linear classification
  - artificial neural networks/ deep learning



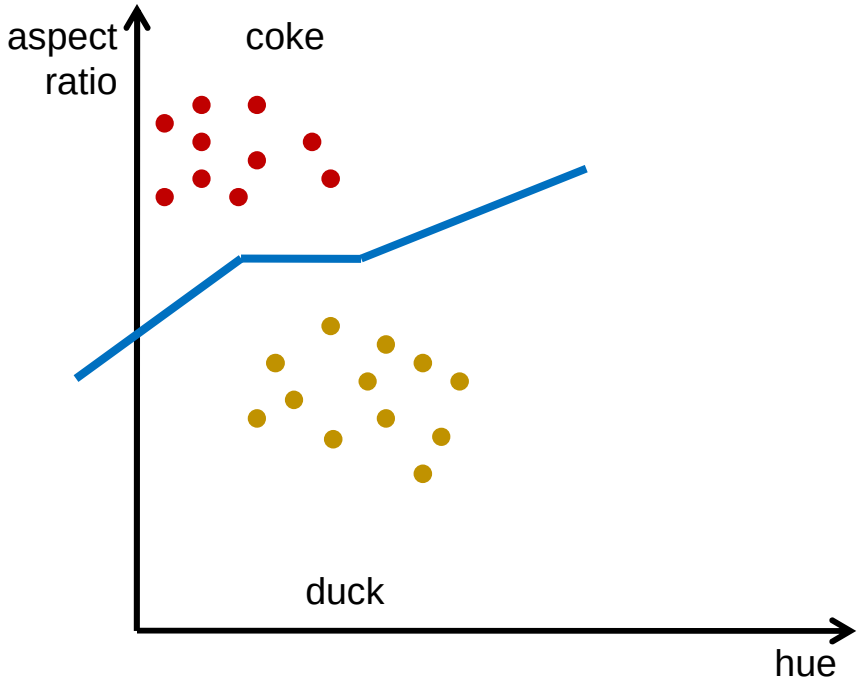
Classification cont.

- many approaches for decision rules and learning
  - linear classification
  - artificial neural networks/ deep learning
  - prototype-based methods



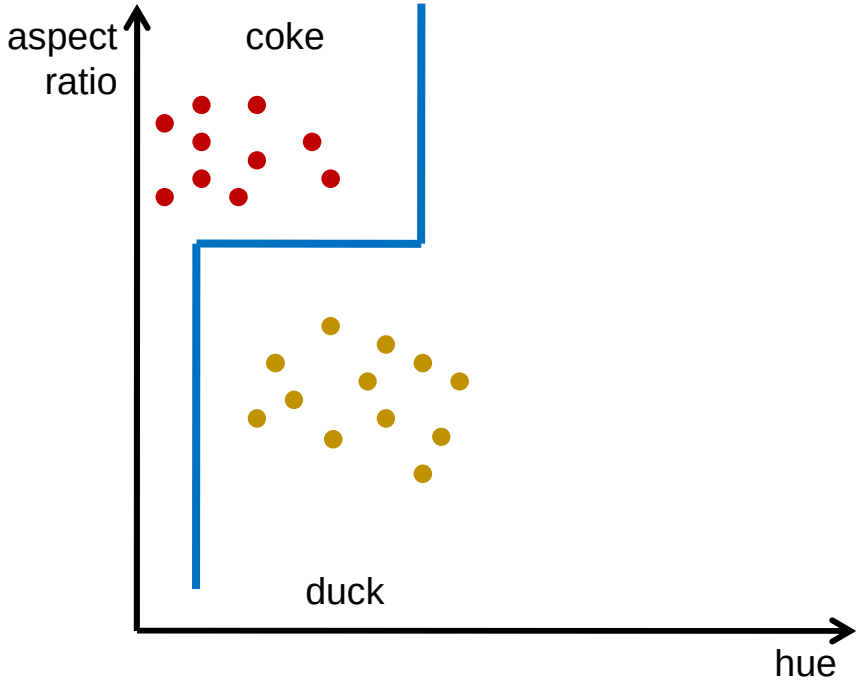
Classification cont.

- many approaches for decision rules and learning
  - linear classification
  - artificial neural networks/ deep learning
  - prototype-based methods
  - case based reasoning



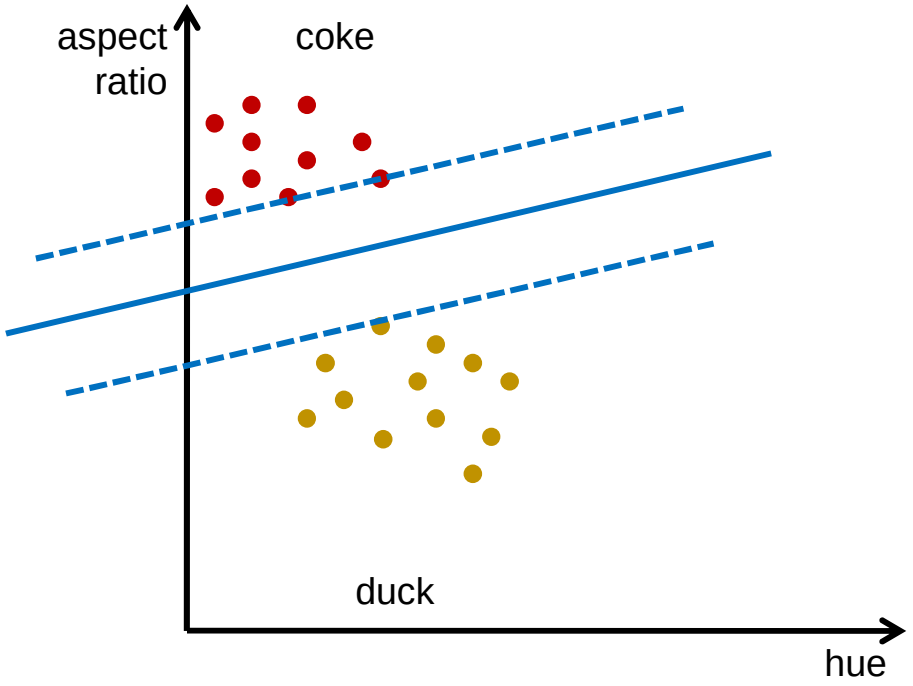
Classification cont.

- many approaches for decision rules and learning
  - linear classification
  - artificial neural networks/ deep learning
  - prototype-based methods
  - case based reasoning
  - decision trees



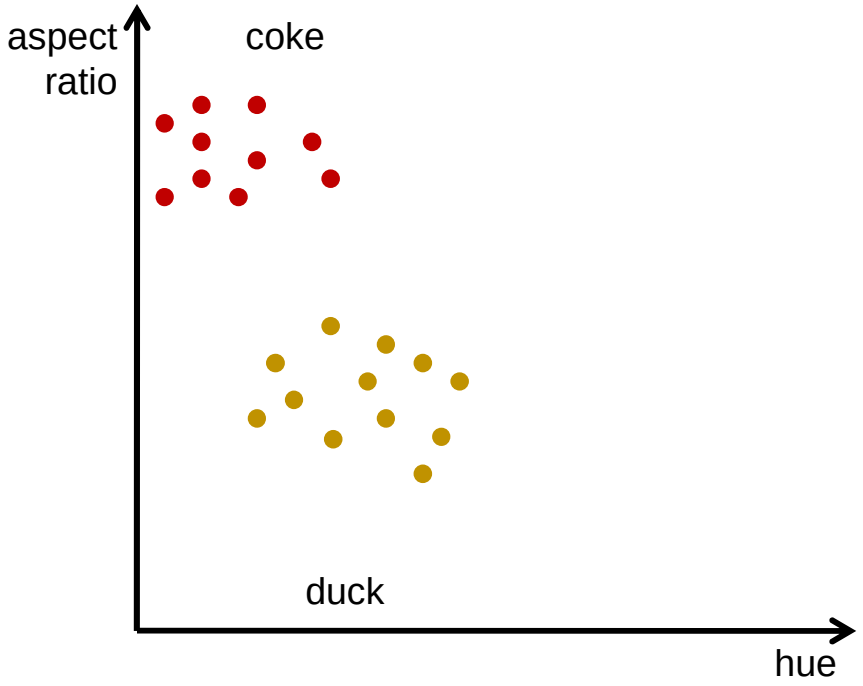
Classification cont.

- many approaches for decision rules and learning
  - linear classification
  - artificial neural networks/ deep learning
  - prototype-based methods
  - case based reasoning
  - decision trees
  - support vector machines



Classification cont.

- many approaches for decision rules and learning
  - linear classification
  - artificial neural networks/ deep learning
  - prototype-based methods
  - case based reasoning
  - decision trees
  - support vector machines
  - boosting (meta algorithm)
  - …



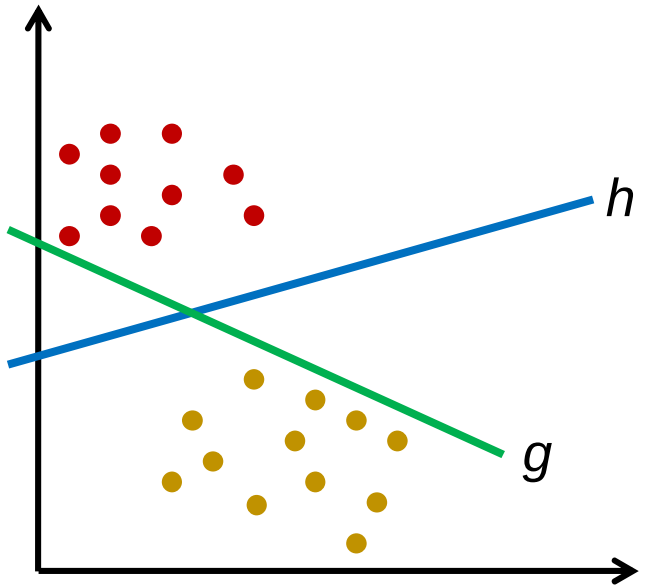
- in this lecture:

  linear classification, support vector machines, boosting, decision trees, deep learning

Slide 9

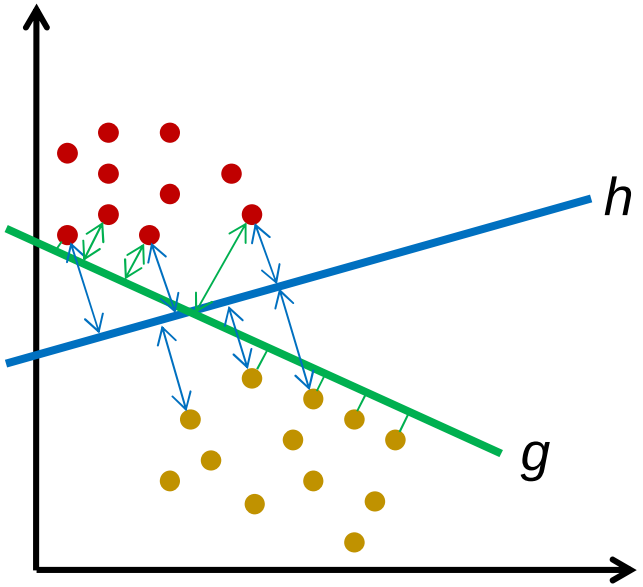
Linear Classification cont.

- many possible solutions, which one is the best?
  - *g* and *h*, both don't make classification errors



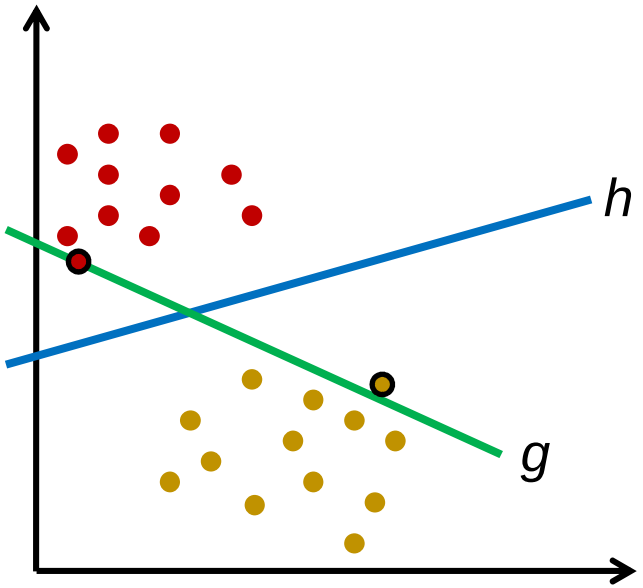
Linear Classification cont.

- many possible solutions, which one is the best?
  - *g* and *h*, both don't make classification errors
  - *g* has shorter distance to patterns than *h*



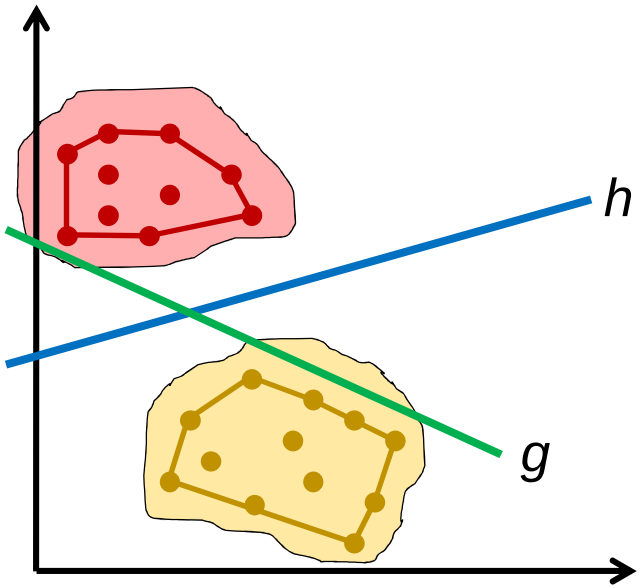
Linear Classification cont.

- many possible solutions, which one is the best?
  - *g* and *h*, both don't make classification errors
  - *g* has shorter distance to patterns than *h*
  - risk of misclassification for new pattern is larger for *g* than for *h*



Linear Classification cont.

- many possible solutions, which one is the best?
  - *g* and *h*, both don't make classification errors
  - *g* has shorter distance to patterns than *h*
  - risk of misclassification for new pattern is larger for *g* than for *h*
  - (unknown) support of the class probability distributions is similar to convex hull of training examples



Maximising the distance of the separating hyperplane to the convex hull of the training patterns means minimising the risk of misclassification (result from computational learning theory)

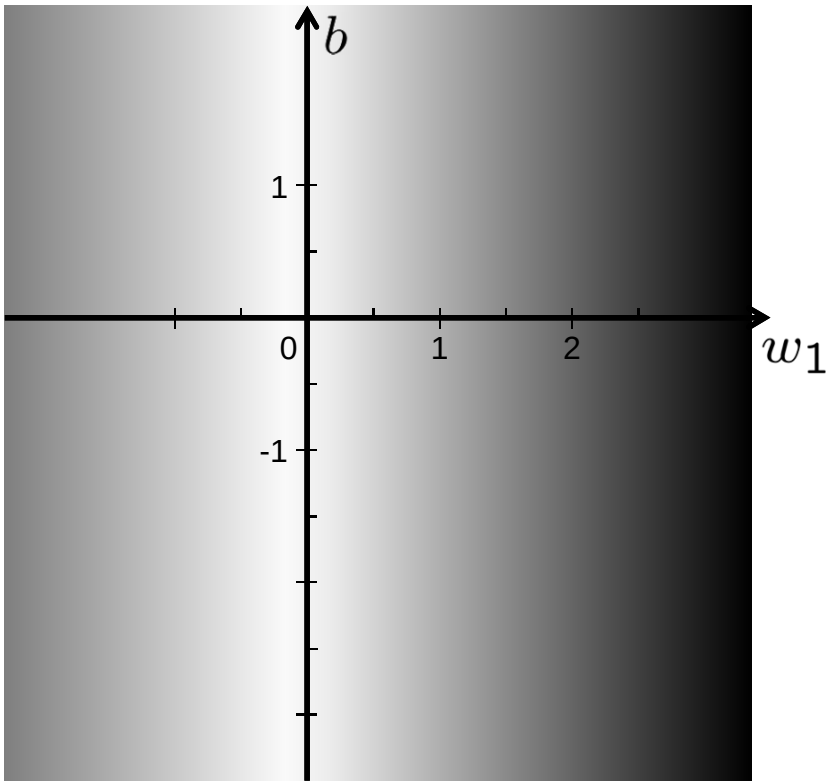
Slide 12

SVM cont.

- a simple example:
  - patterns are 1D:

    positive: 5, 10

    negative: -1, 2
  - parameters: w_1, b



  - optimization problem:

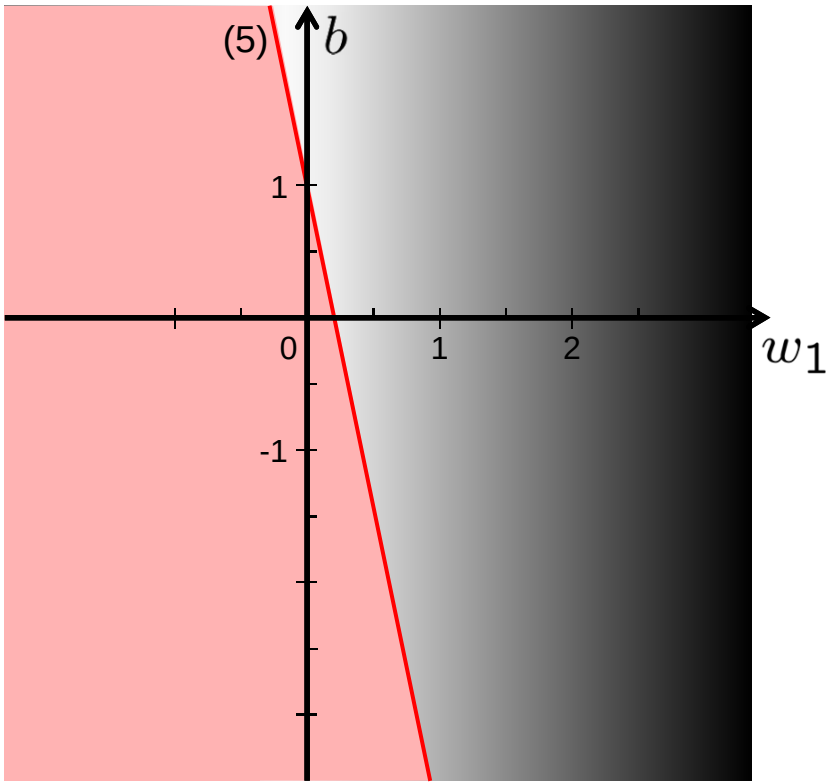
$$\begin{aligned} \underset{w_1,b}{minimise} \quad & \frac{1}{2}w_1^2 \\ subject\ to \quad & b \geq 1 - 5w_1 \\ & b \geq 1 - 10w_1 \\ & b \leq -1 + w_1 \\ & b \leq -1 - 2w_1 \end{aligned}$$

SVM cont.

- a simple example:
  - patterns are 1D:

    positive: 5, 10

    negative: -1, 2
  - parameters: w_1, b



  - optimization problem:

$$\begin{aligned} \underset{w_1,b}{minimise} \quad & \frac{1}{2}w_1^2 \\ subject\ to \quad & b \geq 1 - 5w_1 \\ & b \geq 1 - 10w_1 \\ & b \leq -1 + w_1 \\ & b \leq -1 - 2w_1 \end{aligned}$$

SVM cont.

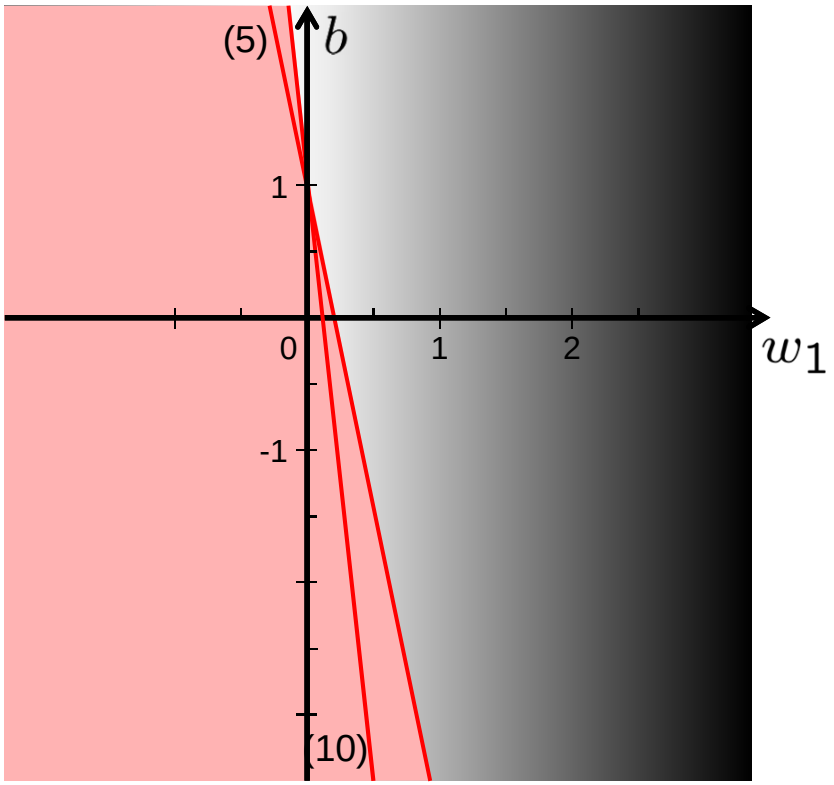
- a simple example:
  - patterns are 1D:

    positive: 5, 10

    negative: -1, 2
  - parameters: w_1, b

  - optimization problem:

$$\begin{aligned} \underset{w_1,b}{minimise} \quad & \frac{1}{2}w_1^2 \\ subject\ to \quad & b \geq 1 - 5w_1 \\ & b \geq 1 - 10w_1 \\ & b \leq -1 + w_1 \\ & b \leq -1 - 2w_1 \end{aligned}$$

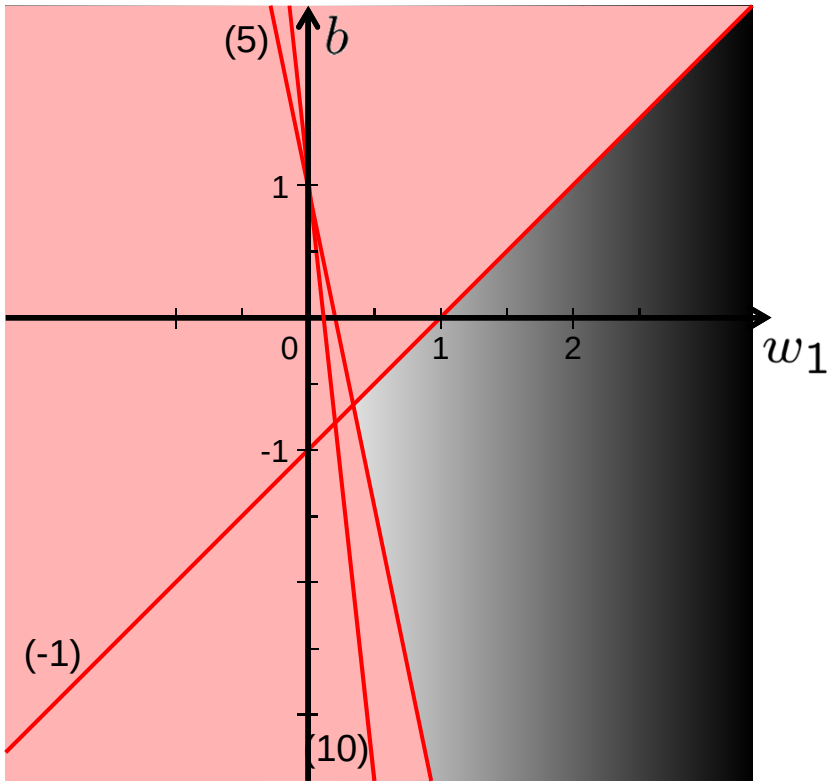


SVM cont.

- a simple example:
  - patterns are 1D:

    positive: 5, 10

    negative: -1, 2
  - parameters: w_1, b



  - optimization problem:

$$\underset{w_1,b}{minimise} \quad \frac{1}{2}w_1^2$$

$$subject\ to \quad b \geq 1 - 5w_1$$

$$b \geq 1 - 10w_1$$

$$b \leq -1 + w_1$$

$$b \leq -1 - 2w_1$$

SVM cont.

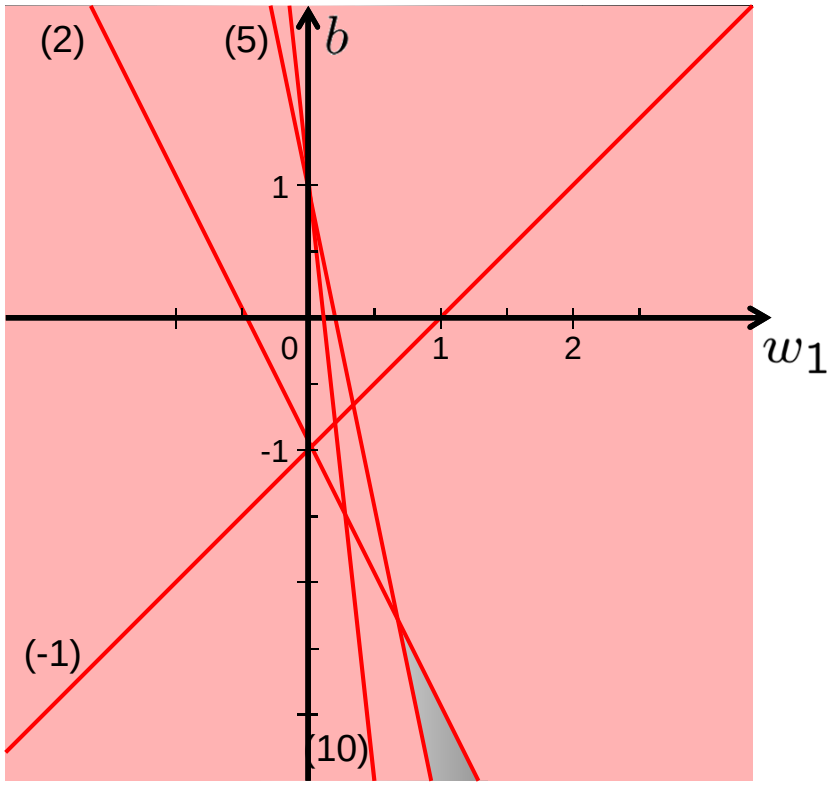
- a simple example:
  - patterns are 1D:

    positive: 5, 10

    negative: -1, 2
  - parameters: w_1, b

  - optimization problem:

$$\begin{aligned} \underset{w_1,b}{minimise} \quad & \frac{1}{2}w_1^2 \\ subject\ to \quad & b \geq 1 - 5w_1 \\ & b \geq 1 - 10w_1 \\ & b \leq -1 + w_1 \\ & b \leq -1 - 2w_1 \end{aligned}$$

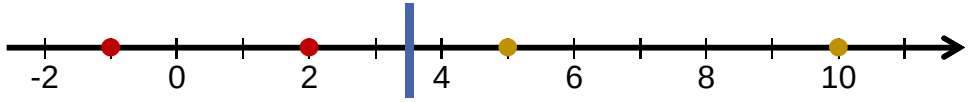


SVM cont.

- a simple example:
  - patterns are 1D:

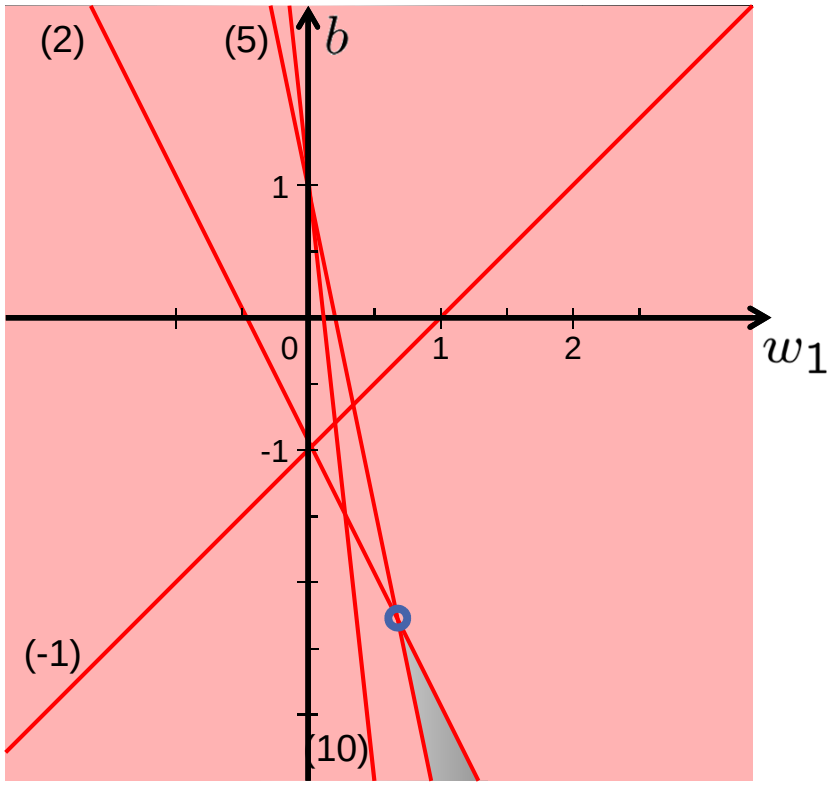
    positive: 5, 10

    negative: -1, 2
  - parameters: w_1, b



  - optimization problem:

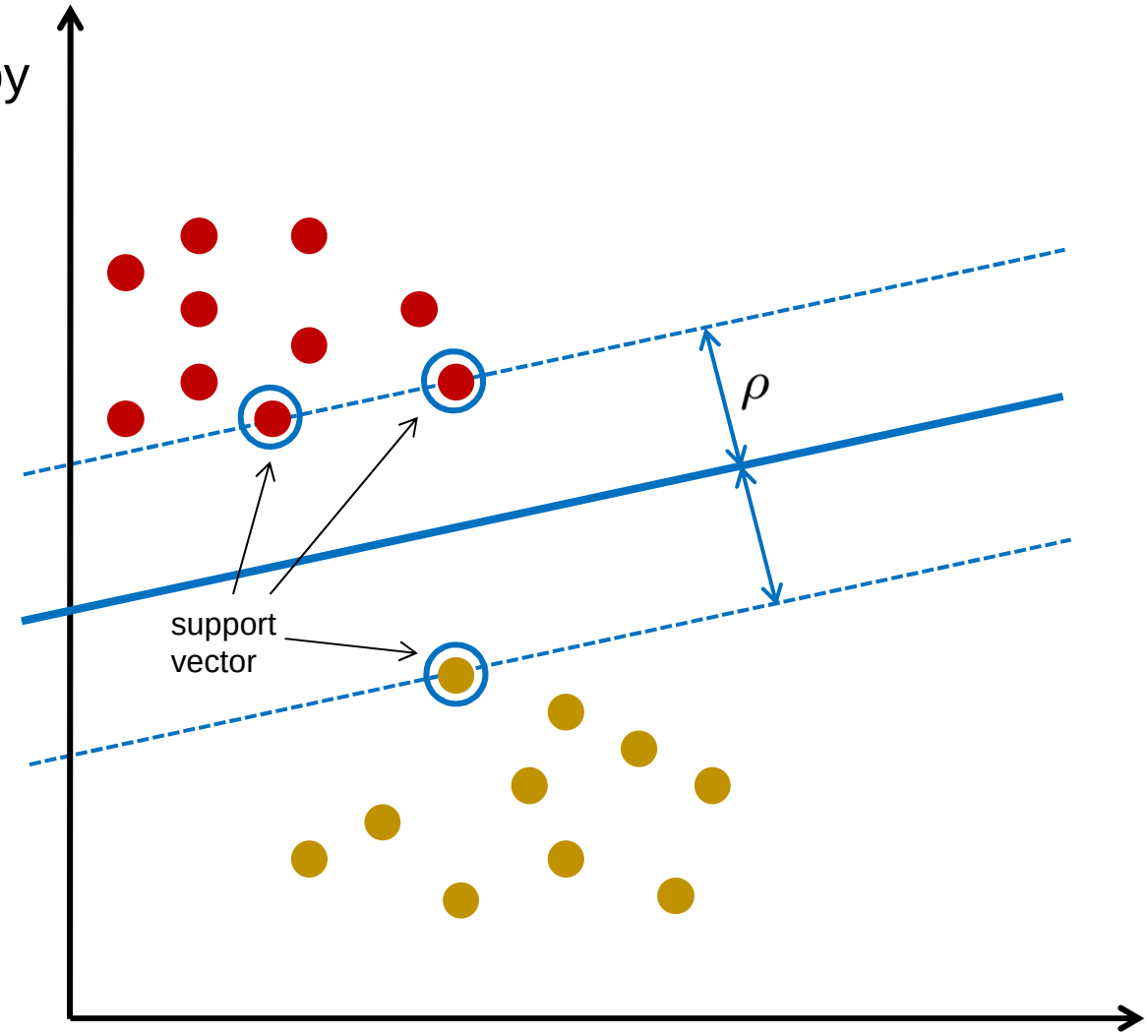
$$\begin{aligned} \underset{w_1,b}{minimise} \quad & \frac{1}{2}w_1^2 \\ subject\ to \quad & b \geq 1 - 5w_1 \\ & b \geq 1 - 10w_1 \\ & b \leq -1 + w_1 \\ & b \leq -1 - 2w_1 \end{aligned}$$



Slide 15

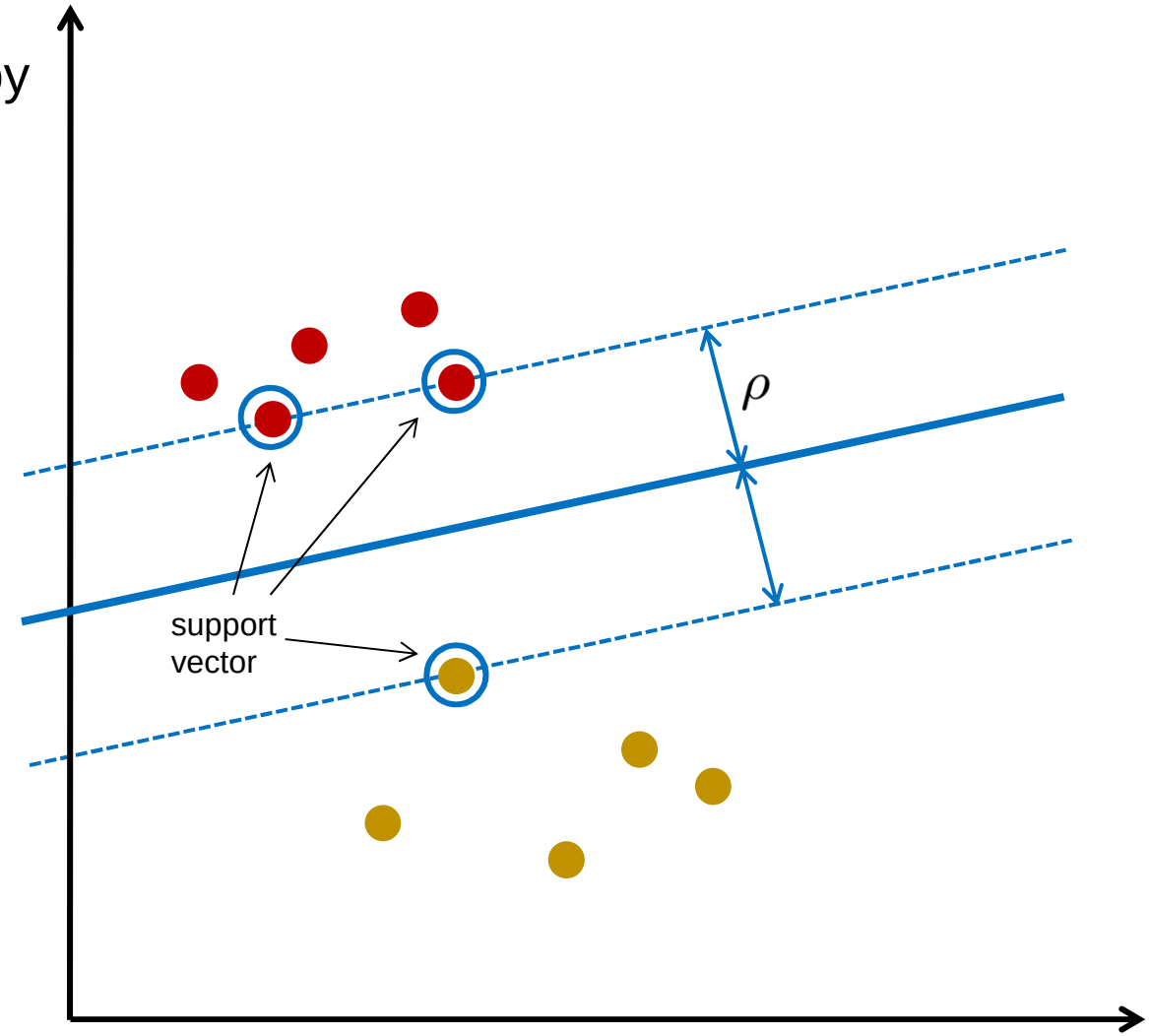
SVM cont.

- optimal separating hyperplane determined by support vectors



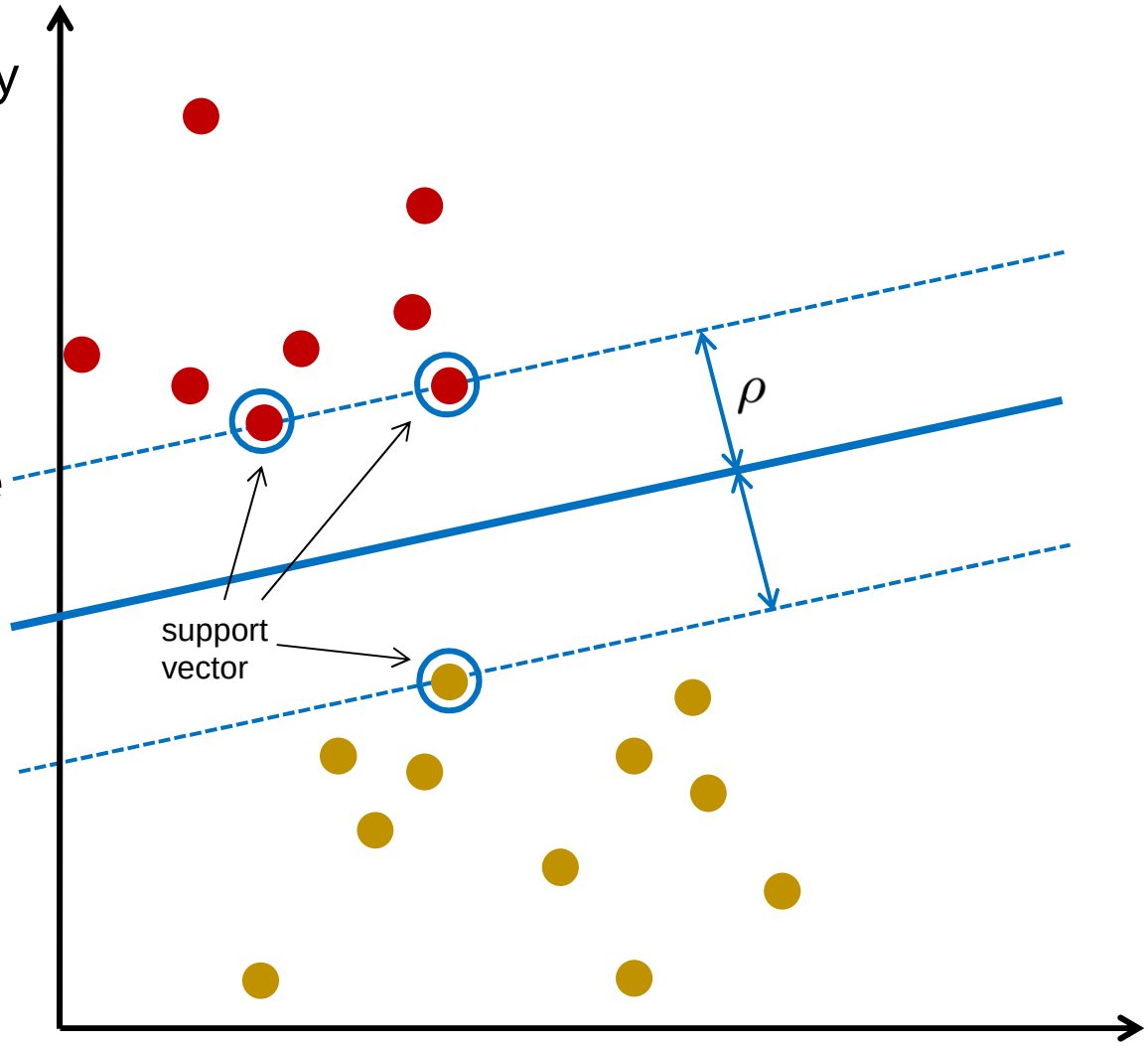
SVM cont.

- optimal separating hyperplane determined by support vectors
- removing non-support vectors does not change solution



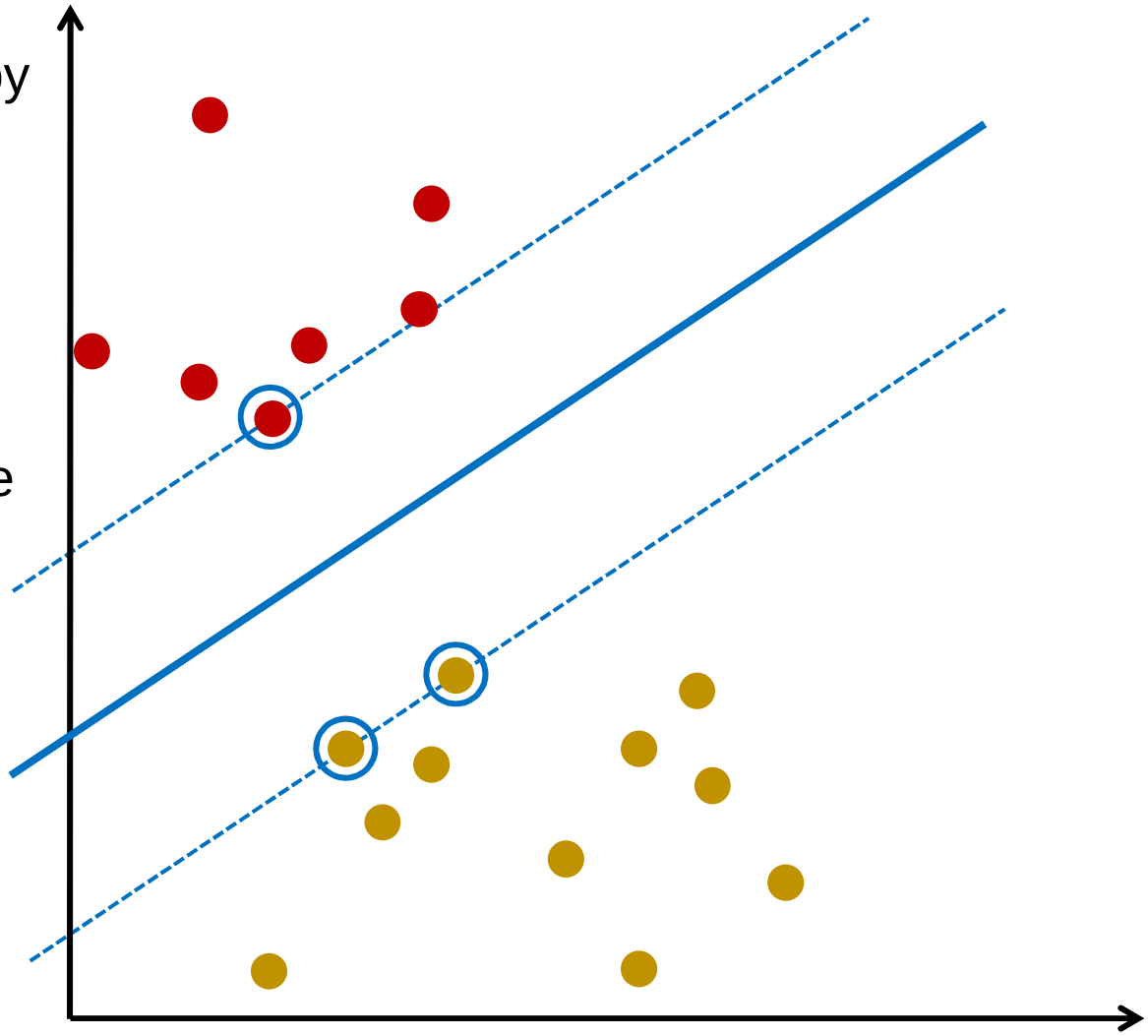
SVM cont.

- optimal separating hyperplane determined by support vectors
- removing non-support vectors does not change solution
- adding patterns with distance of more than the margin does not change solution



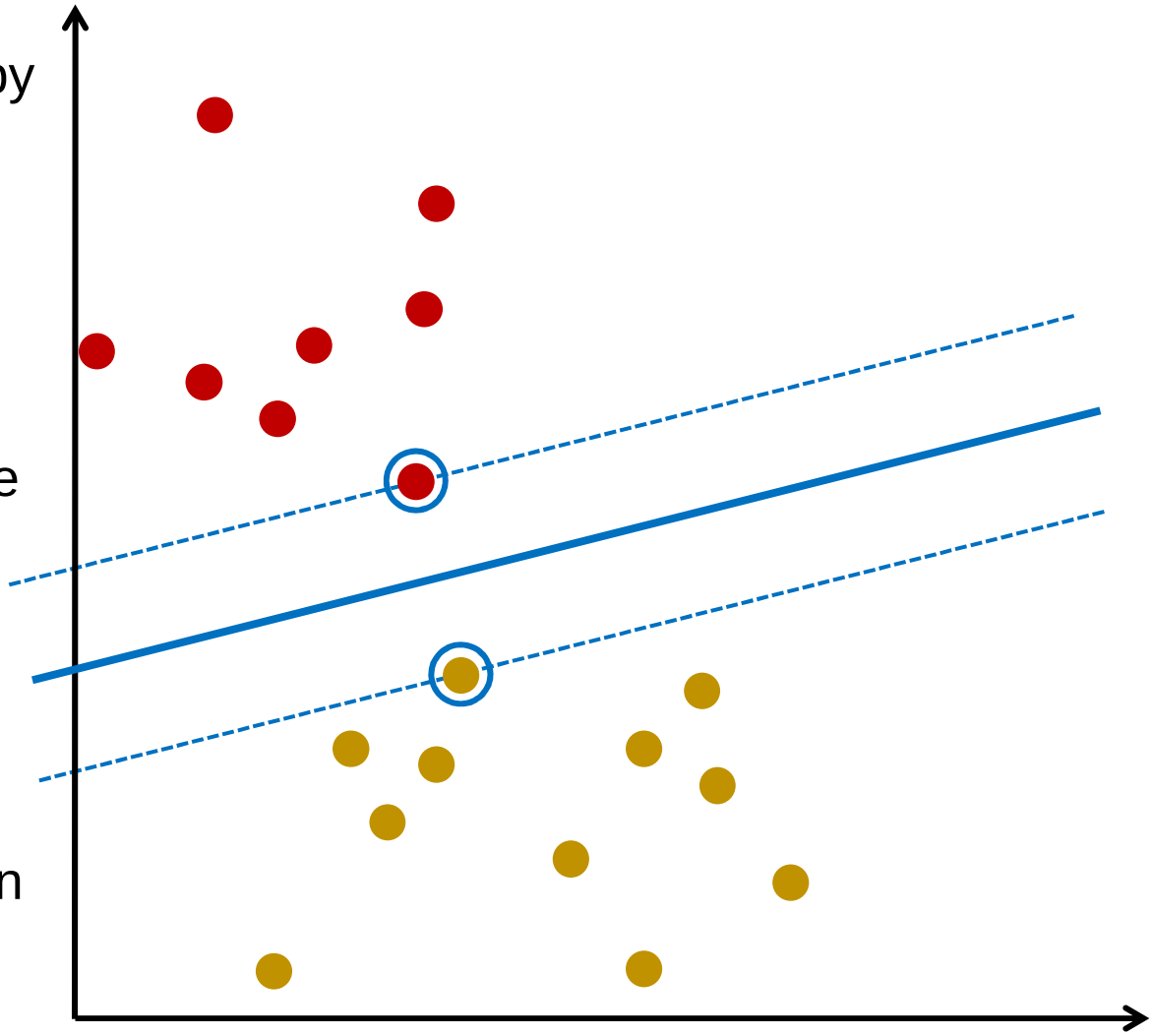
SVM cont.

- optimal separating hyperplane determined by support vectors
- removing non-support vectors does not change solution
- adding patterns with distance of more than the margin does not change solution
- removing support vector changes solution



SVM cont.

- optimal separating hyperplane determined by support vectors
- removing non-support vectors does not change solution
- adding patterns with distance of more than the margin does not change solution
- removing support vector changes solution
- adding pattern with distance less than margin changes solution



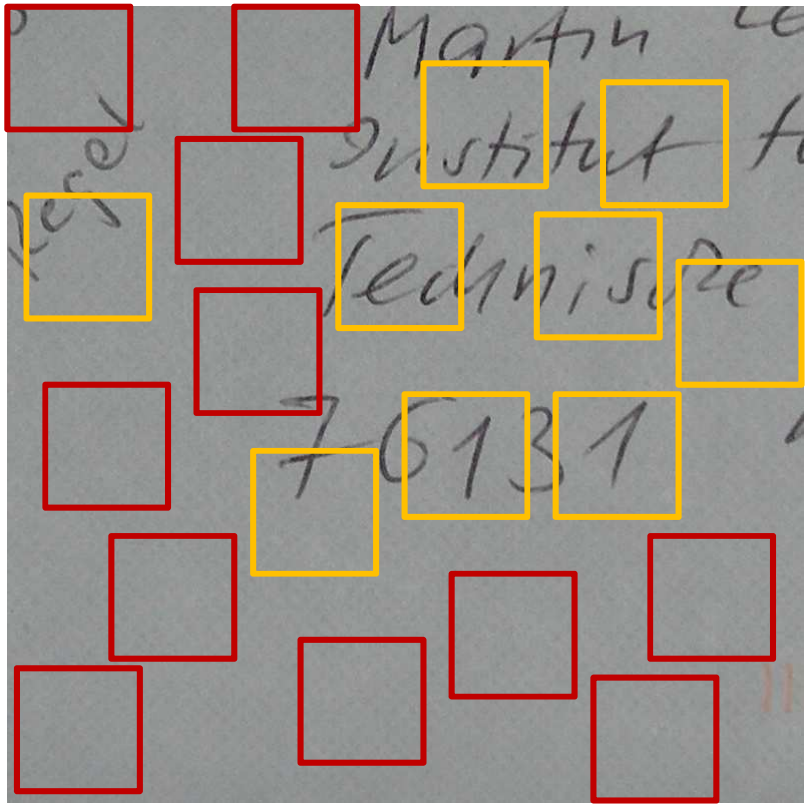
Slide 89

Searching for Objects cont.

- improved idea:

  use two classifiers

  – classifier 1
    - efficient
    - inaccurate
    - high recall
    - low precision
    - applied to all areas



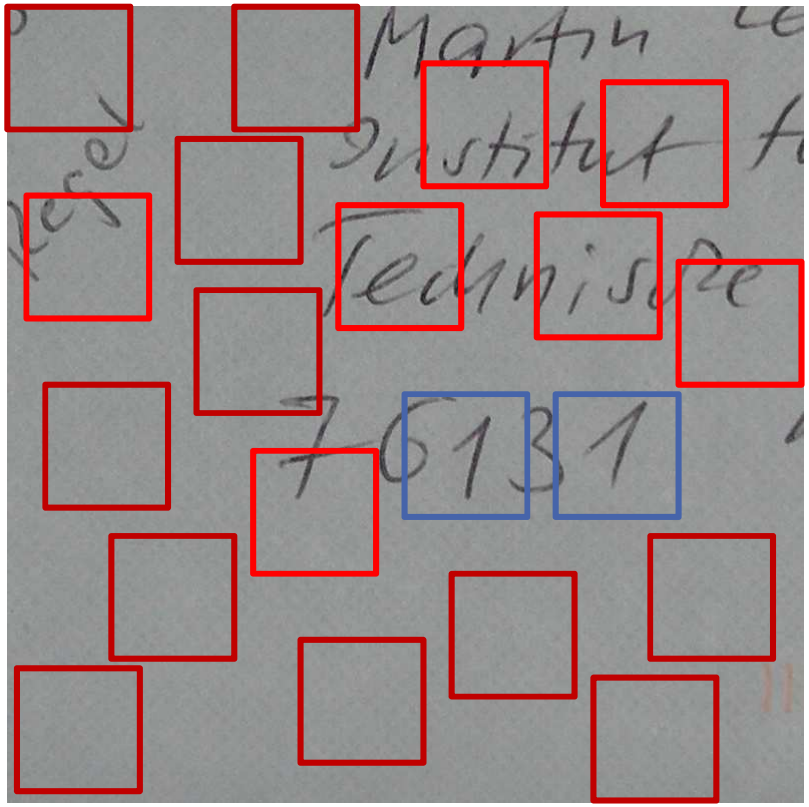
Searching for Objects cont.

- improved idea:

  use two classifiers

  – classifier 1
    - efficient
    - inaccurate
    - high recall
    - low precision
    - applied to all areas

  – classifier 2
    - inefficient
    - accurate
    - high recall
    - high precision
    - applied to areas which are found by classifier 1



- idea can be extended to a series of many classifiers

  → approach of Viola/Jones