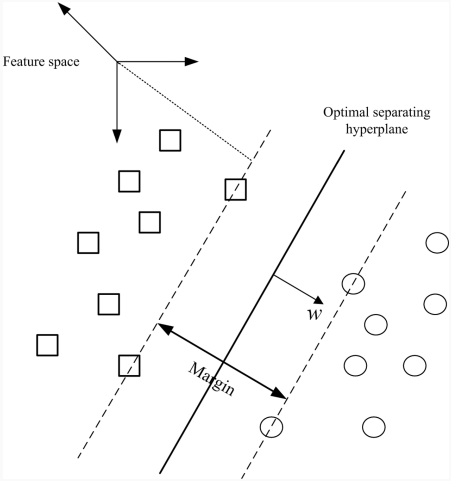
# Support vector machines

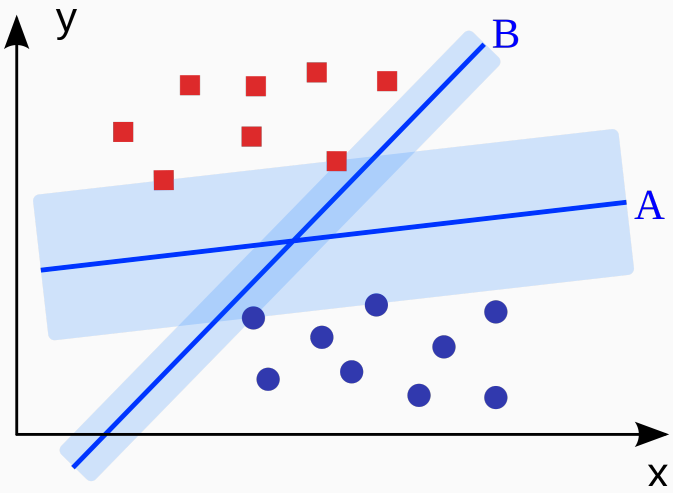
Gianluca Campanella

# Support vectors




From Li et al. (2011)

# Margin




Via *Wikimedia Commons*

# Margin

- Maximising the margin is good

→ Less overfitting

→ Model generalises better

- Only support vectors are important

- Can be done by solving a quadratic optimisation problem subject to linear constraints

# Hard and soft-margin SVM

**Hard-margin**

- Requires correct classification of all samples
- Only solvable if samples are linearly separable

# Hard and soft-margin SVM

**Hard-margin**

- Requires correct classification of all samples
- Only solvable if samples are linearly separable

**Soft-margin**

- Some misclassification is allowed
- Will ‘compromise’ on model performance to obtain a larger margin $\rightarrow$ more generalisable model

# Non-linear SVM

**Idea**

Map the original input space to some higher-dimensional space where the training set is linearly separable

**‘Kernel trick’**

- Effectively introduces new predictors
- No need to compute (and store) the expanded dataset

# Pros and cons

**Pros**

- Can handle large datasets (only support vectors matter)
- Effective in high-dimensional spaces ($p > n$)
- Mathematically ‘convenient’ (also ‘kernel trick’)

**Cons**

- Prone to overfitting ($\rightarrow$ use soft-margin)
- Do not provide probability estimates directly