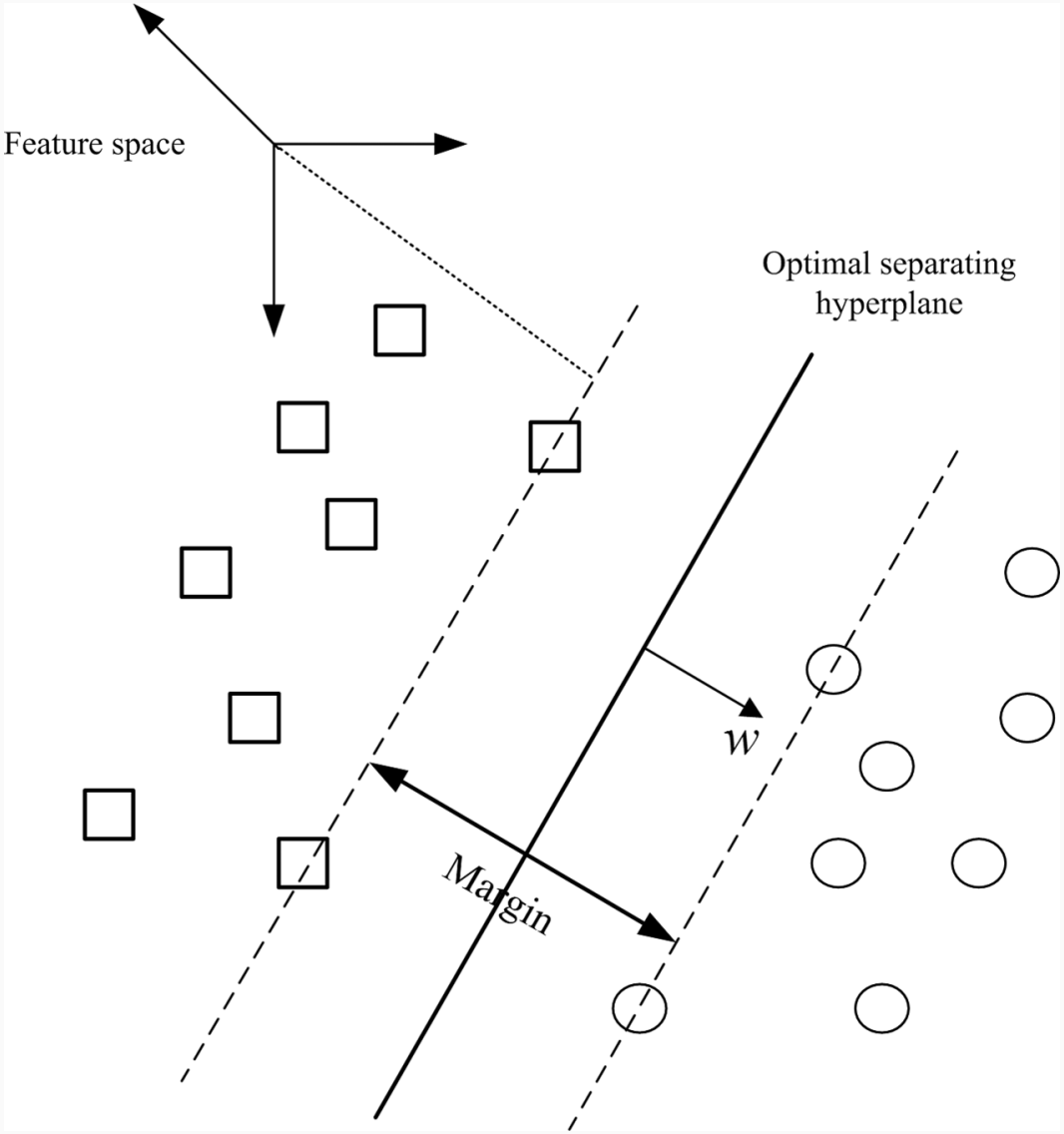
estimand

Support vector machines

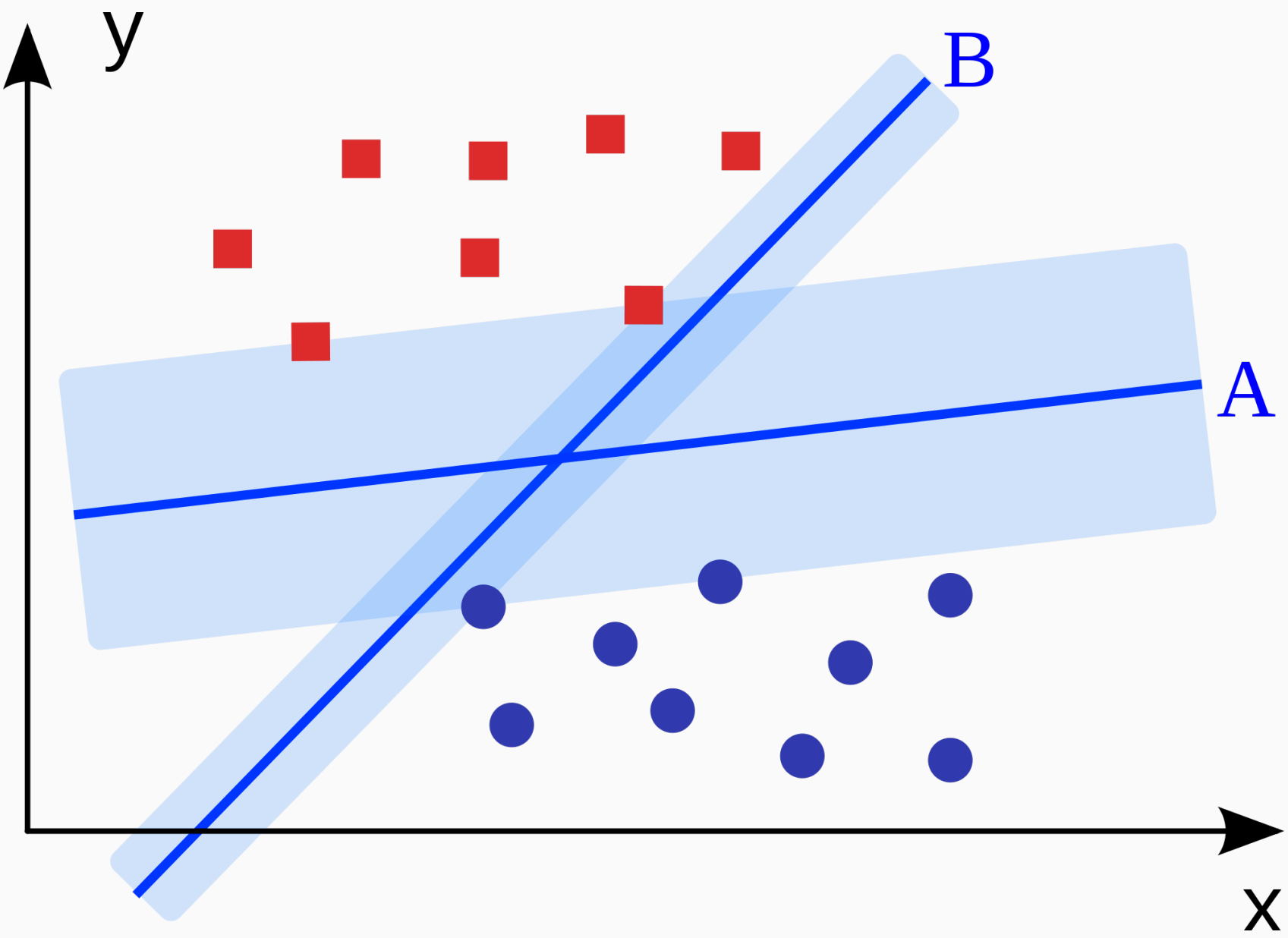
Dr Gianluca Campanella

9th June 2016

Support vectors



Margin



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, for example of 'difficult' samples
- Will 'compromise' on model performance to obtain a larger margin (more generalisable model)

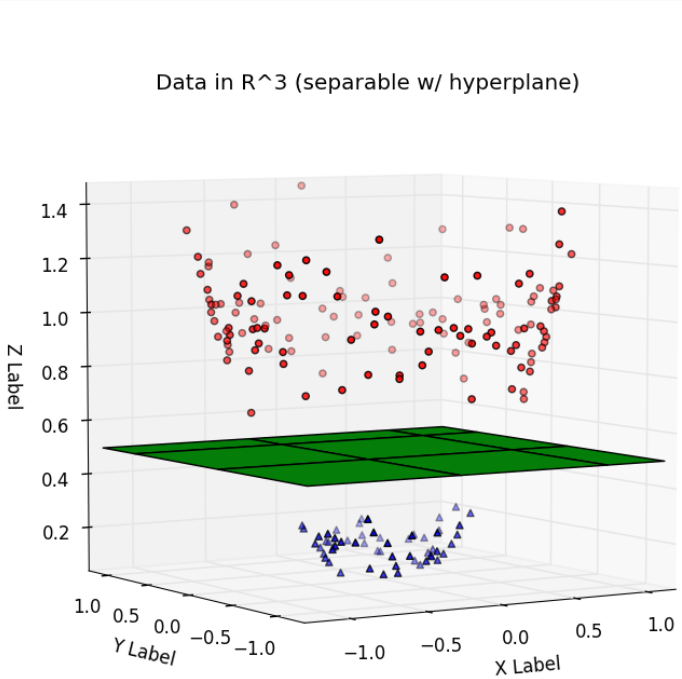
Non-linear SVM

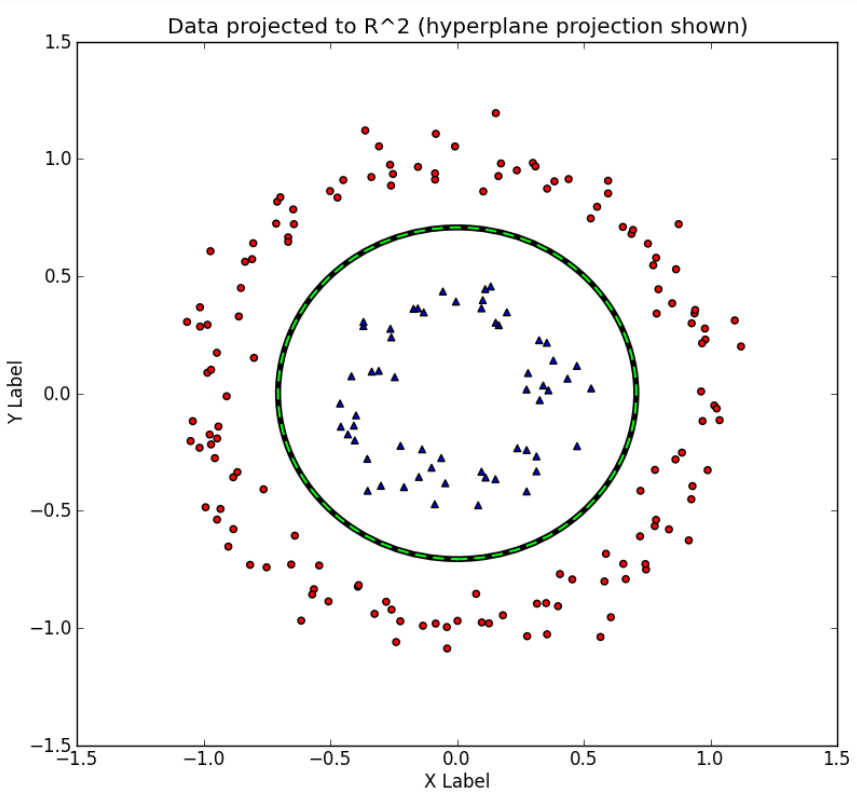
Idea

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

- Effectively ‘expands’ the dataset without introducing new predictors
- Using the ‘kernel trick’, this can be done **without computing the expanded dataset** $\rightarrow$ efficient

Non-linear SVM





Pros and cons

Pros

- 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