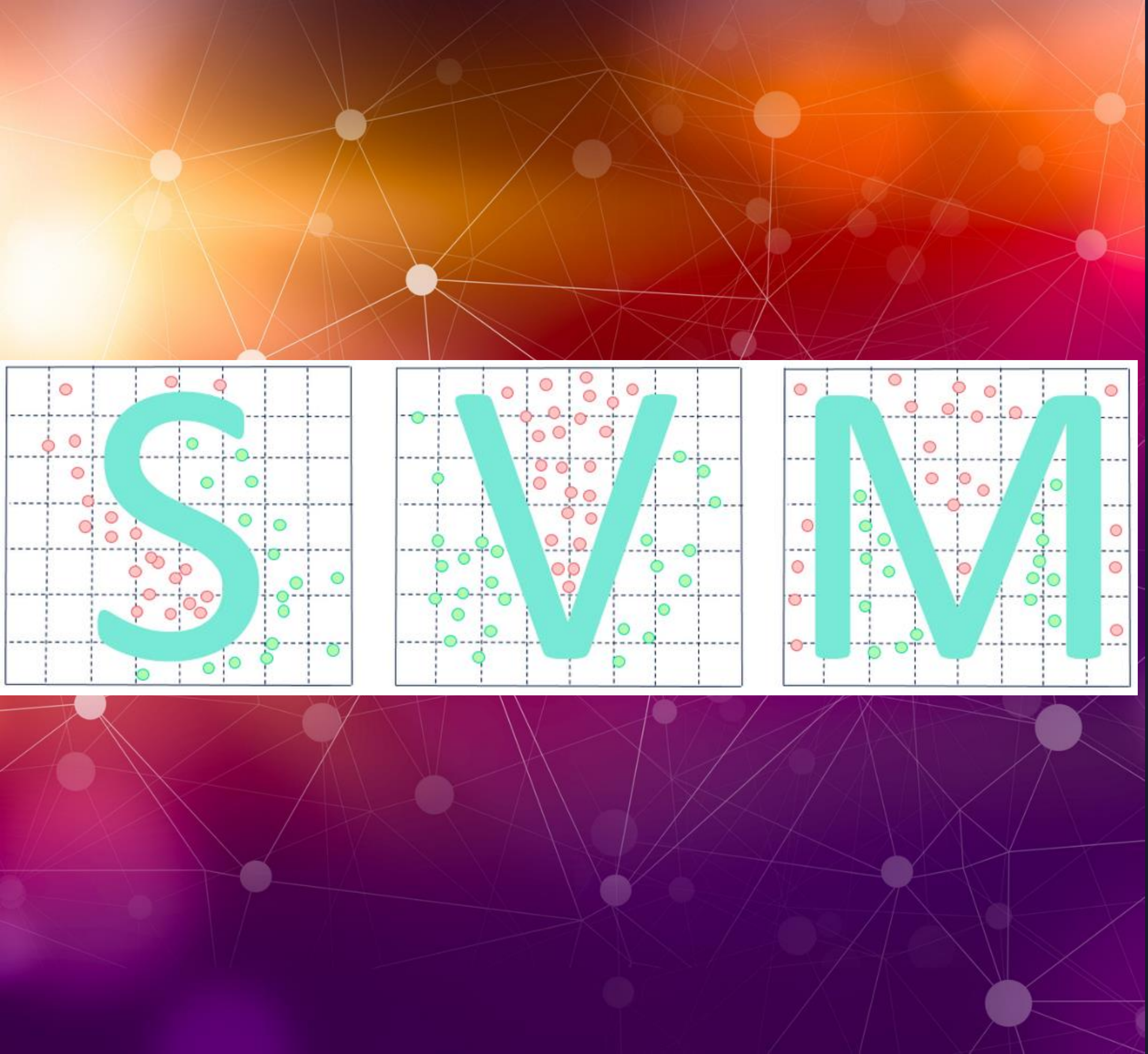
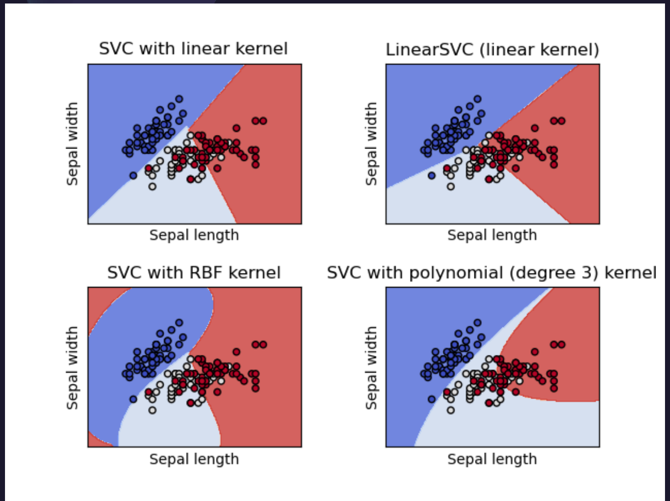


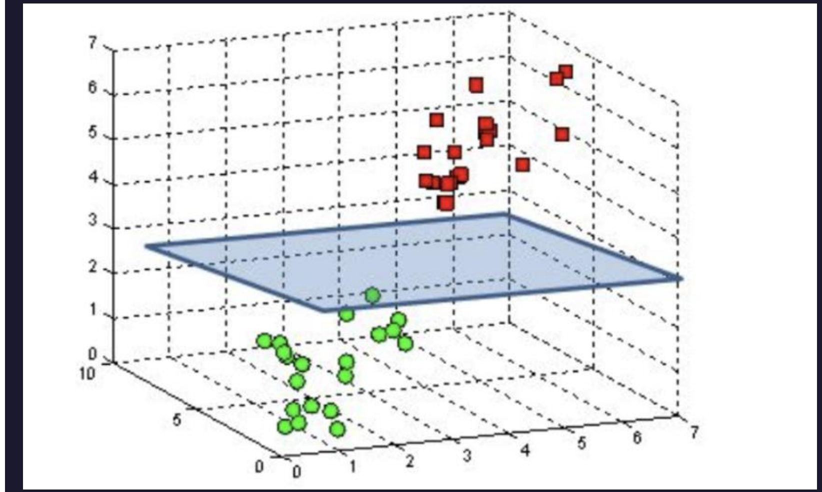
SVMs (Support Vector Machines)

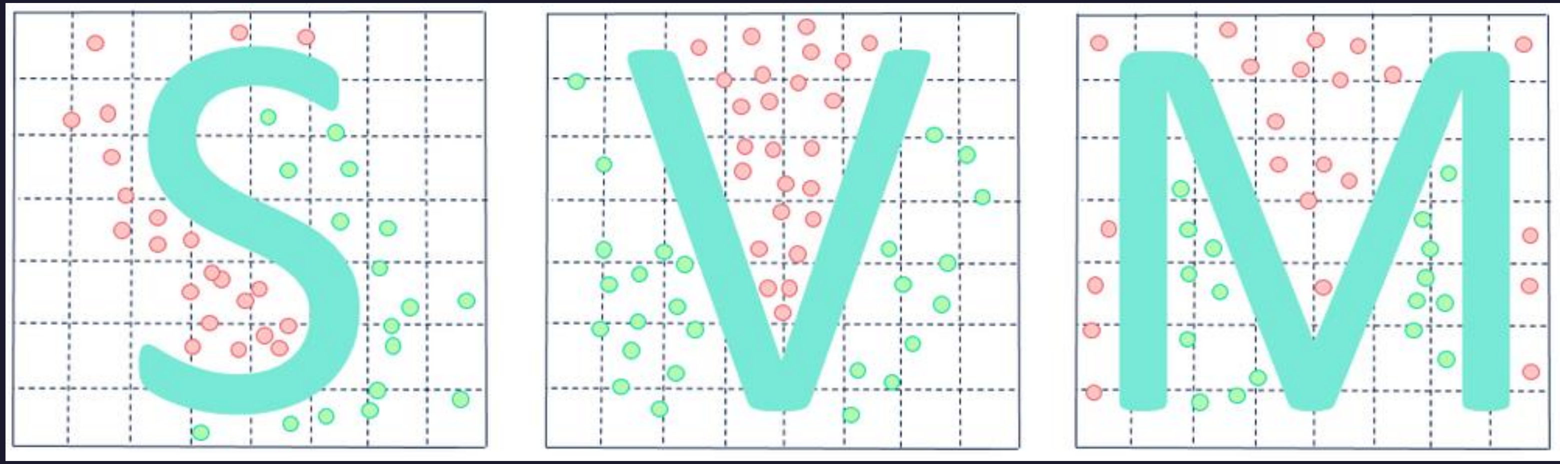
Group 2

Ahad, Jake, Ryan, Yansong

What is an SVM



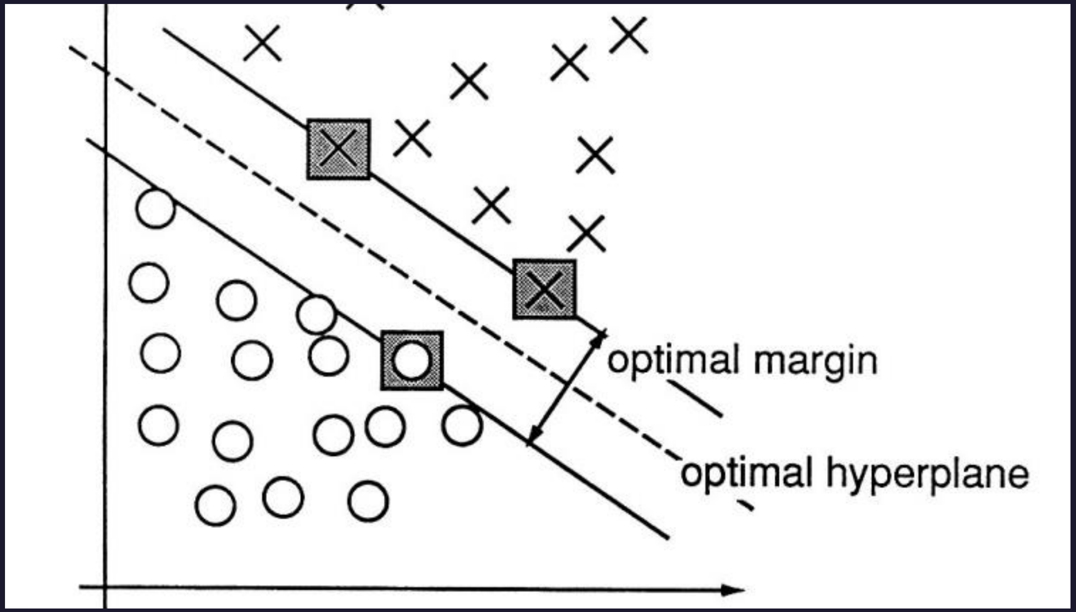


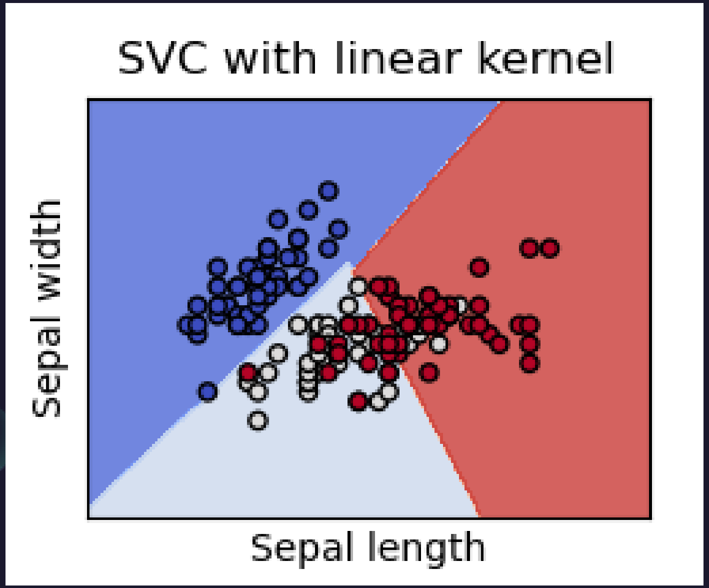


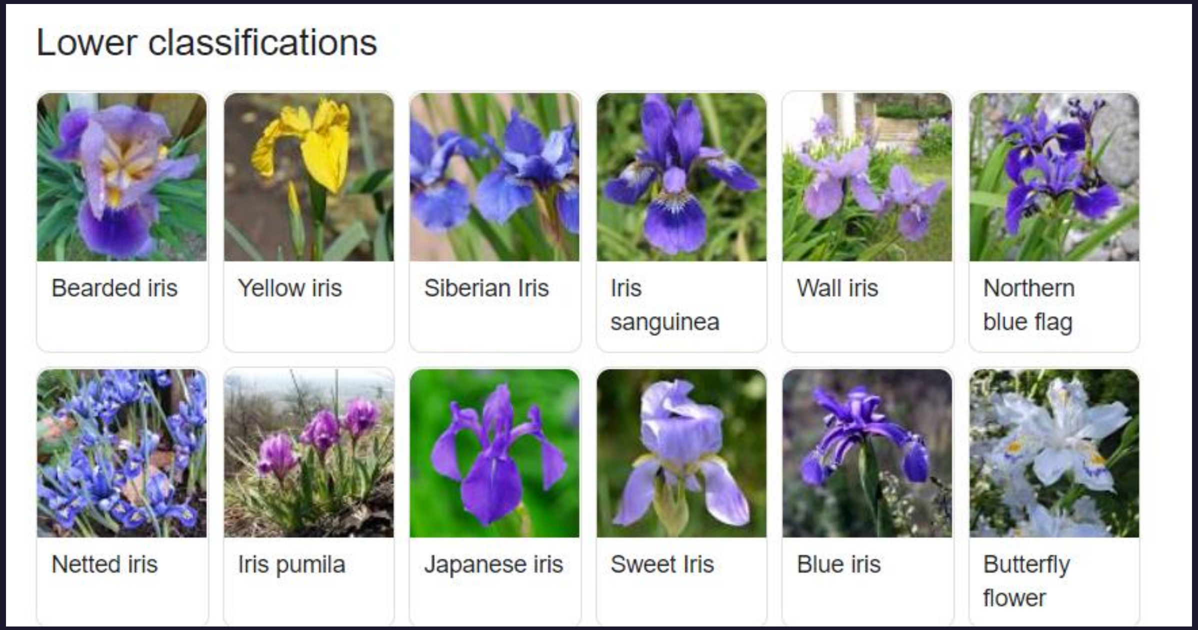
- Support Vector Machine
- Primarily used for classification
- Logistic Regression, not just binary
- Supervised classification
- Very popular

What it does

- Maximize "Road" size
- Support vectors on sides of roads

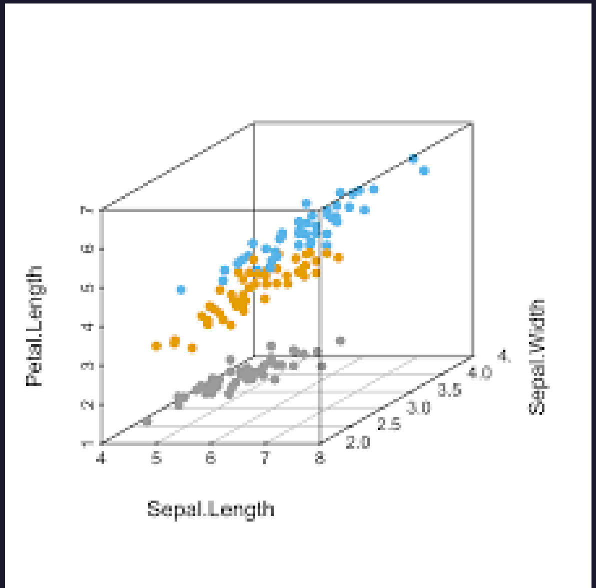


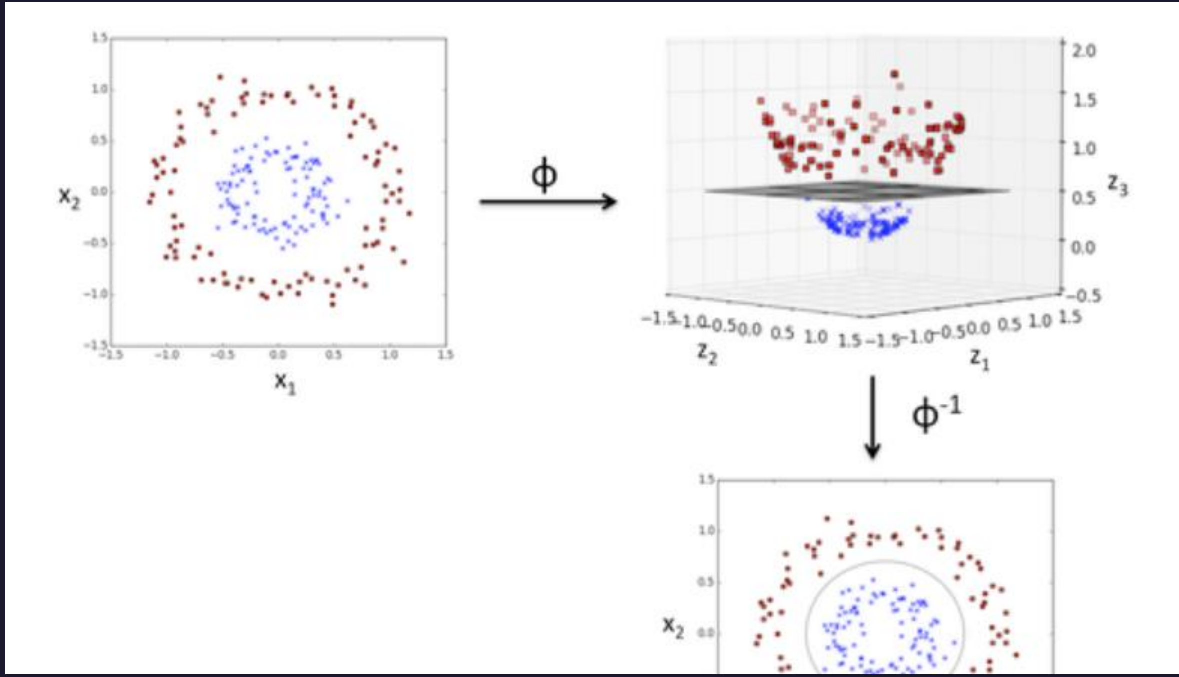




How it works

- Works in higher dimensions
- Hyperplanes instead of lines





Data Processing requirements

DATA FORMATTING

- Must be similarly dense to be fitted properly, if determining vars are densely populated cannot accept sparse input from dependent vars
- Should be scaled (mean 0 variance 1)
- Data must be labeled (we want to know what group it's supposed to be in)

COMPUTER SIDE

- Needs kernel cache size to be large for larger datasets to run well
- Needs to be C-Contiguous (stored sequentially in memory)

Disadvantages

Overfitting

The number of features cannot be too much larger than the number of samples.

Not designed for numerical linear regression

The best kind of data to use SVM's for are predicting a discrete outcome, rather than a numerical one.

Similarly, this is not suited to large datasets, or datasets with overlapping classes. K-means clustering is better for overlapping classes.

No direct probability estimates

Success is measured by how many violations are produced, and how large of a gap can be produced in the line between clusters.

Advantages

High accuracy

Highly efficient with memory

Only uses a subset of training points for the support vectors

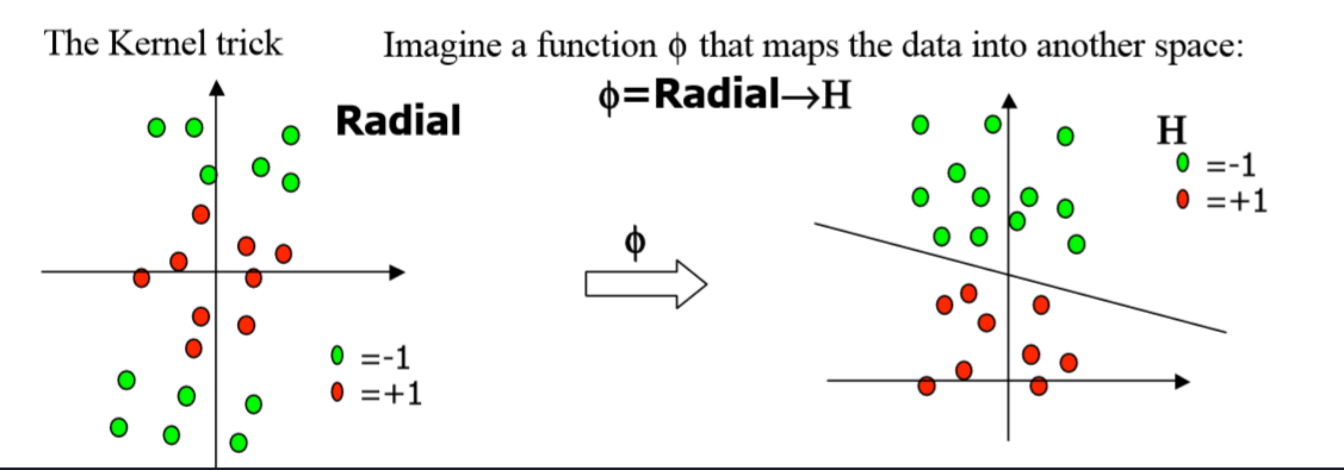
High versatility

Different kernel functions can be specified, allowing for a high degree of customization

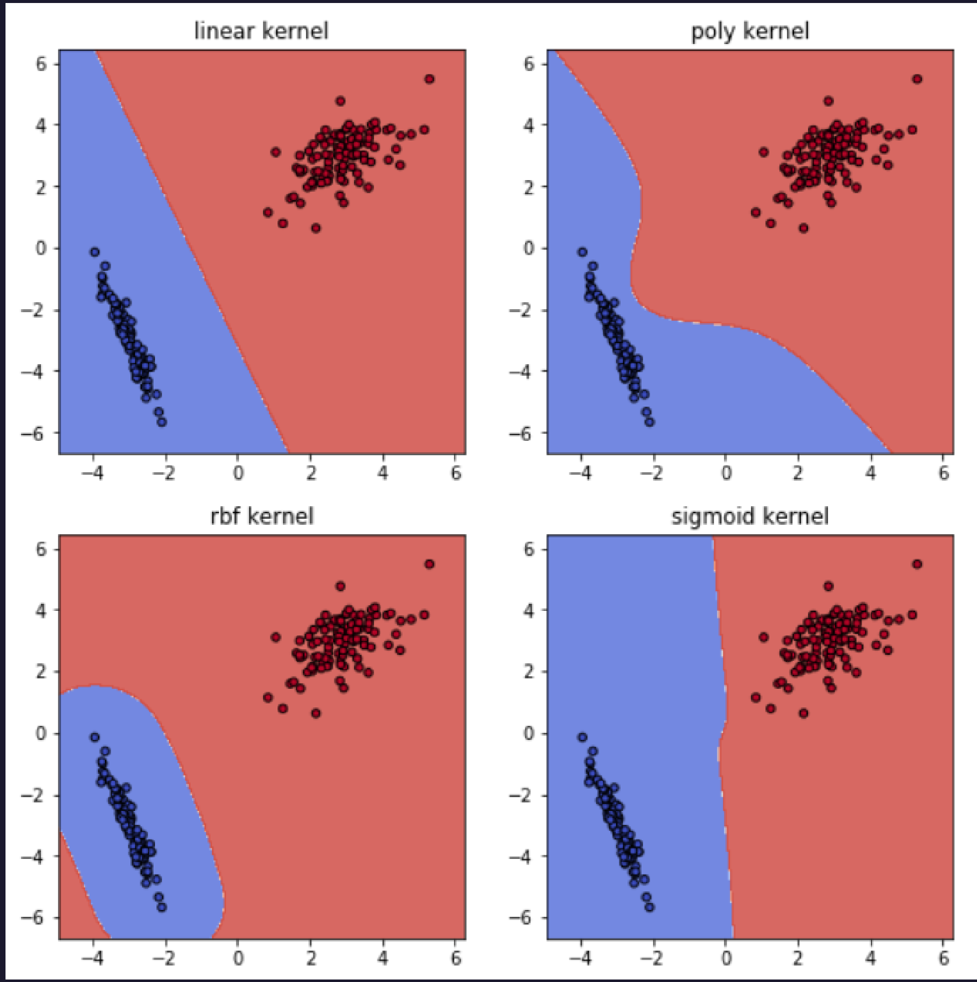
Gives the user more control over the outcome compared to K-means clustering

Hyperparameters: Kernel

- "Kernel tricks" project data into a transformed space where it may be easier to draw a line separating the data
- Example values: 'linear', 'rbf', 'poly', 'sigmoid'



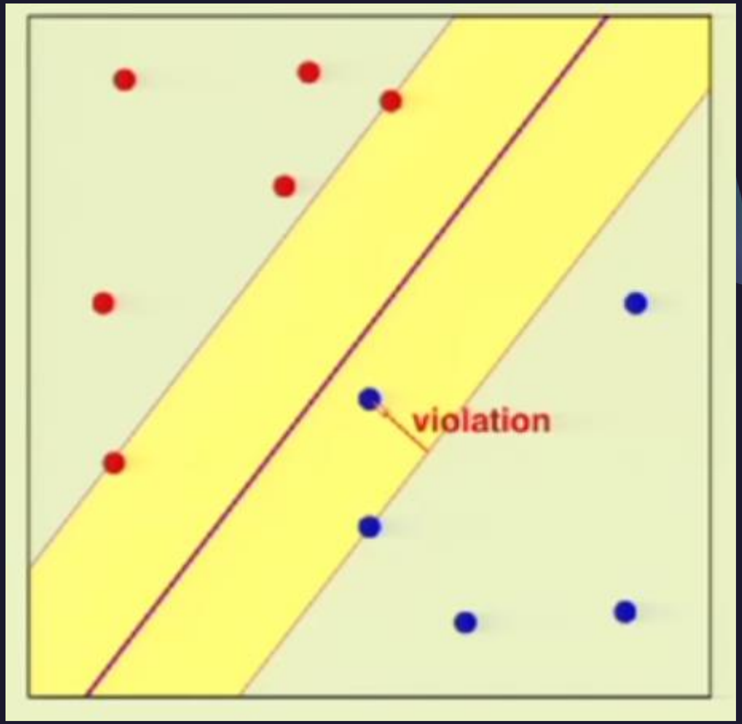
An example transformation where kernel = 'rbf' [2]



One dataset using different kernels [1]

Hyperparameters: C

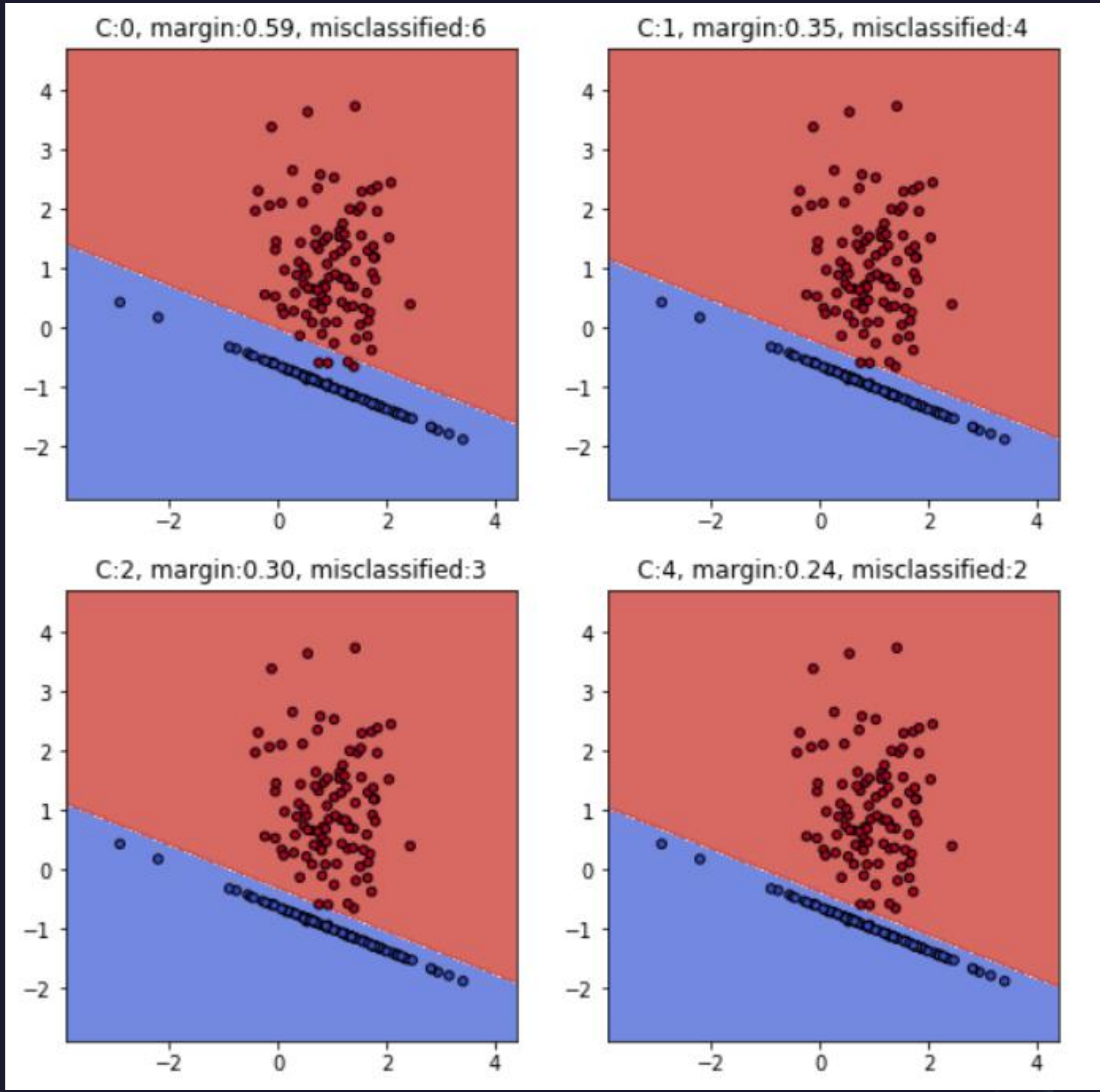
- C is a "penalty term" that determines how much the model cares about violations
- A value of 0 means the model doesn't avoid violations at all
- A high value means the model sacrifices margin to prevent any violations, can cause overfitting



An example of a violation [1]

Hyperparameters: C

- As C increases, violations decrease
- As C increases, the margin shrinks
- Note high C is more computation-heavy



Effects of Various values of C [1]

Hyperparameters: Various

- Different kernels require different parameters, but all kernels use C
- Polynomial uses 'degree' and 'coef0'
- Radial Basis Function (RBF) uses 'gamma'
- Sigmoid uses 'coef0'
- For kernel details, see scikit 1.4.6 'Kernel Functions' [4]
- For an RBF tuning example, see [5]

Appendix

1. Support Vector Machines Explained
2. An Idiot's Guide to SVMs (MIT)
3. Caltech Lecture on Kernel Methods
4. Scikit Learn: SVM
5. Grid Search for Hyperparameter tuning in SVM
6. Support-Vector Networks (Original paper)
7. SVM Tutorial with Example
8. SVM for Text Classification
9. SVMs for Dummies
10. Semantic Parsing using SVMs

Thank You

Tuesday, February 2, 20XX