1 Support Vector Machines

1.1 Recap: Machine Learning Systems

There are two main types of machine learning systems, some are instance based some are model based.

1.1.1 Instance-Based Learning

Uses the entire dataset as a model (eg; k-NN).

- Compare new data points to known datapoints
- Non-parametric approaches
- Memory-based approaches
- Prediction can be expensive

1.1.2 Model-Based Learning

Use the training data to create a model that has parameters learned from the training datasets (e.g SVM).

- Detect a pattern in the training data
- Build a predictive model
- Prediction is extremely fast

2 Outline

Data characteristics...

- Linearly separable data
- Non-Linearly separable data

SVM come in three different flavours

- linearly separable: hard-margin SVM's
- non-linearly separable: soft-margin SVM's
- non-linearly separable: kernelized SVM's

3 Data Characteristics

SVM's, we only deal with binary classifications (ending this section we generalise). A classification method for both linear and nonlinear data...

So can we separate data using a linear function... Yay or nay? Margin of error (soft/hard margin SVM's).

4 Types of SVM's

SVM selects the maximum margin linear classifier... or SVM selects the maximum margin linear classifier with partial misclassification's allowed.

5 Problem Definition: Margin Maximization

Given a set of linearly separable training data $S = \{(x_1, y_1), ..., (x_n, y_n)\}, y_i \in \{+1, -1\}$. We want to find a linear decision boundary, hyperplane, to separate the 2 classes... There's an infinite number of lines separating the two classes. SVM tries to solve for this by finding the optimal hyperplane (maximises the margin).

We do this because of the assumption; The hyperplane with the largest margin will generalise best on unseen data.

5.1 Some intuitive definitions

Ofcourse your linear hyperplane will be defined precisely as $\vec{w} \cdot \vec{x} + b = 0$, same thing being equal to -1 is the lower boundary (less classed as -1) or equal to 1 is the upper boundary (more classified as 1). As you can see, this neatly boils into a classic optimisation problem, slides show how.

$$min_{w,b} \frac{||\vec{w}||}{2}, \text{ st. } y_i(\vec{w} \cdot \vec{x_i} + b) \geq 1$$

Any points that fall on either H_1 (+1) or H_2 (-1) are called support vectors.