COMPARISON OF DIFFERENT MACHINE LEARNING MODELS

Frenciel Anggi

	k-NN	(Regularised) Linear Least Squares	Logistic Regression	Artificial Neural Network	Support Vector Machines
Introduction	A simple intuitive method with no training.	The most widely used modelling method.	A simple and effective method for classification.		
Generalise				It can generalise, supported by advanced training strategies.	It can generalise.
Regression/ Classificiation	Can be used for both regression and classification.		It is for classification only.		
Linear/ Non-Linear	Can handle both linear and nonlinear data patterns.	Limited model expressive power, cannot deal with nonlinear data patterns.	It is a linear classifier and cannot handle nonlinear data patterns.	Capable of learning and modelling non-linear and complex relationships .	Can handle both linear and nonlinear data patterns.
Multi class classification	Easy to implement for multi-class classification.		Not very good with multi-class classification.		
Local optima				The solution is a local optimal.	Unlike ANN, it is not solved for local optima. = global?
Data	Not good at dealing with imbalanced data.	Makes efficient use of the data. Good results can be obtained with relatively small data.			Scale relatively well to high dimensional data.
Input data				Does not impose restrictions and assumptions to the input.	Usually provide competitive performance. It is a good choice when we have no idea on the data.
Model			It does not assume specific distribution, e.g., Gaussian, in the model.	Good choice for black-box modelling.	
Interpret model		Easy to explain and understand.		Difficult to interpret the learned model.	
Probabilistic Interpretation		When being used in classification, the performance is sometimes sensitive to how the target output is set, and it does not offer probabilistic interpretation.	Offer nice probabilistic interpretation.		
Cost/time	Slow with large-scale data (The computational complexity of (1) distance calculation suffers from curse of dimensionality (2) neighbour search suffers from training data size.	Low computational cost: fast training and prediction.		Long training time for large-scale data.	
Memory	High memory cost (needs to store all the training data)	Low memory.			
Set hyperparameters	There are only two things to decide, which can be relatively easy: (1) hyper-parameter value k (2) distance measure	In the least squares case, there is no hyper-parameter to set. In the regularised case, there are two things to decide: (1) the form of regularisation term (e.g., l2, l1), and regularisation parameter.	No need to set hyper-parameter.	Need to determine a “good” network architecture, which is sometimes not easy.	Need to choose a “good” kernel and select hyper-parameters, which are not easy.
Outliers	Sensitive to outliers.				
Others				Learned information is all stored in network weights.	