Wrap-up of Supervised Learning Algorithms

Nathan Sprague
CS444

Supervised Learning: The Story So Far

Algorithm	Notes	Classification	Regression
Linear Regression	Results are relatively easy to interpret: weights tell us something about the importance of input dimensions. Only works well if there is a linear relationship between input and output variables. Optimal weights can be found analytically and efficiently. (This can be viewed as a single-layer NN with no non-linearity)		X
Logistic Regression	Simple, can have good performance. Scales to large data sets. No direct analytical solution, but the optimization surface is convex so gradient descent (or fancier optimization algorithms) can be used without worrying about local minima. Only works well if the data is linearly separable. (This can be viewed as a single-layer NN with a sigmoidal non-linearity. Outputs are generally interpreted as the probability of the input belonging to the target class.)	X	

Supervised Learning: The Story So Far

Algorithm	Notes	Classification	Regression
Multi-Layer Neural Netwoks (Also known as Multi-Layer Perceptrons or MLP’s)	Flexible, general, powerful. Scale to very large data sets. Recent results with deep architectures have achieved state-of-the-art results on many tasks. Can be finicky and slow to train. Generally require lots of training data.	X	X
Convolutional Neural Networks	Specialized architecture for computer vision tasks. State of the art approach for many problems in computer vision. Can be finicky and slow to train. Generally require lots of training data.	X	X
K-Nearest Neighbors (and variations)	Appropriate for on-line learning. Very few parameters to tune. Doesn’t always work well in high-dimensional spaces. Classification speed is a function of the size of the data set.	X	X

Supervised Learning: The Story So Far

Algorithm	Notes	Classification	Regression
Decision Trees	Works well for both categorical and continuous data. Relatively easy to interpret. Prone to over-fitting. Doesn’t scale easily to huge data sets.	X	X
Random Forests	Works well for both categorical and continuous data. Resistant to over-fitting. Not many parameters to tune. Doesn’t scale easily to huge data sets. GOOD STARTING POINT AS LONG AS THE DATA SET IS NOT TOO BIG.*	X	X

* Fernández-Delgado, Manuel, et al. "Do we need hundreds of classifiers to solve real world classification problems." J. Mach. Learn. Res 15.1 (2014): 3133-3181.

Supervised Learning: The Story So Far

Algorithm	Notes	Classification	Regression
Support Vector Machines	No worries about local minima. Good theoretical performance guarantees. Resistant to over-fitting. Not as many fiddly hyper-parameters as neural networks, but more than random forests. Doesn’t scale easily to huge data sets. Best-suited for two-class classification. Can be tweaked for regression and mult-class classification, but not as natural ALSO A REASONABLE CHOICE FOR MODERATELY SIZED DATA SETS*	X	X

* Fernández-Delgado, Manuel, et al. "Do we need hundreds of classifiers to solve real world classification problems." J. Mach. Learn. Res 15.1 (2014): 3133-3181.