1 Machine Learning techniques

1.1 Artificial Neural Networs

Artificial neurons were first proposed by McCulloch and Pitts in 1943. Motivated by biological neurons, the artificial neuron received several weighted inputs and produced an output, based on some threshold. The perceptron model (Rosenblatt 1962) built on this early work, adding a learning rule to improve the performance of the neural network. However, the perceptron model was severely limited, unable to solve non-linearly separable functions such as XOR (Perceptrons, Minsky and Papert 1969). Eventually, multilayer perceptrons were developed to address the original perceptron model's shortcomings.

A checkers-playing program used neural networks to train a player (Chellapilla Fogel 2001). The board was encoded as a vector of available board positions, with values assigned based on whether the square was empty, taken by a regular piece, or taken by a King. The neural network consisted of three hidden layers. The first hidden layer completed spatial preprocessing, representing each subsquare of the board as a node, for a total of 91 nodes. The second and third hidden layers had 40 and 10 nodes, respectively. The network outputted a value between -1 and 1, representing the goodness of the board from the current player's perspective. The weights for the network were initially specified through a uniform sample, and several networks played against one another. The winners were declared 'parents', and they generated 'offspring networks' with weights varied by a parameter vector. The process was repeated for many generations to produce an ideal neural network.

1.2 Support Vector Machines

Support vector machines are binary classifiers that can be applied to linearly separable datasets. They separate data into classes using a hyperplane. SVMs can also be used non-linearly by mapping the data to a higher-dimensional space, thus making the data separable. This mapping is done by a kernel function. (ML cs 158 notes 7) SVMs perform well with large feature spaces, as long as the data is separable with a wide margin. They also do well with sparse datasets, as in text classification. (joachims 1998). In the absence of large amounts of labeled data, pool-based active learning can be utilized with SVMs (Tong Koller 2001). The learning algorithm has access to a pool of unlabeled data, and is able to choose a subset of that pool to use as training data. The learner chooses pool data to use such that the data minimizes the size of the learner's set of hypotheses, and brings it closer to a single hyperplane. This approach allows for using less labeled data.

2 Applications to Medical Data

Medical diagnosis presents an ideal domain for machine learning algorithms. A large part of diagnosis falls under pattern recognition, based on large amounts

of data, and ML algorithms are well-suited to this task. For an algorithm to be effective in this domain, it needs to be able to handle noisy and missing data, rely on relatively few medical tests, and complement the role of physicians (Kon01). Machine learning algorithms have been applied to a variety of medical data, some examples of which are outlined below.

2.1 Self-reported input

Some diagnoses rely largely on patient-reported information, rather than biological tests. A prime example of this is diagnosis of mental disorders, which is based on how a patient's symptoms compare to criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders. Symptoms are determined through consultation with a mental health professional, and are largely reported by the patient. Automated systems have been proposed which will produce a diagnosis based on user-reported information. (Yap and Clarke 1996)

2.2 EEG and EKG data

Recordings of electrical activity in the body can be used to diagnose a variety of disorders. Electroencephalograms (EEGs) are recordings from the brain and contain a wealth of features that can be used by machine learning algorithms. A classification algorithm using EEGs was able to diagnose Alzheimer's Disease with 86.05% accuracy (Podgorolec 2012).

Electrocardiograms (EKGs) are often used to detect arrhythmia, which is any abnormality of the heartbeat. They can be indicative of heart disease and other conditions. Det 89 derived a model from the Cleveland Clinic heart disease data set and compared it to the CADENZA Bayesian algorithm. Both models were found to overpredict heart disease, though this occured more with CADENZA (Det 89).

After that, GAD98 presented a novel machine learning approach to diagnosing and classifying cardiac arrhythmia, called the VF15 algorithm. It used a genetic algorithm to learn feature weights. Then, each feature 'voted on a class prediction. The algorithm had a 62% accuracy on this task and was found to outperform Naive Bayes.

KKG+99 collected data on ischemic heart disease, including signs and symptoms, EKG, and scintigraphy. They applied Naive Bayes, neural networks, k-nearest neighbors, and two decision tree algorithms, and compared these to clinicians diagnoses. Naive Bayes had the best sensitivity/recall, whereas clinicians, followed by neural nets, had the highest specificity.

GMCL05 compared various machine learning algorithms for arrhythmia diagnosis based on EKG data, with an emphasis on minimizing false positives and dealing with noisy data. They used the UCI Machine Learning Repository Arrhythmia dataset, and highlighted the need to improve on VF15s 62% accuracy. They evaluated a Bayesian artifical neural network classifier as compared to Naive Bayes, decision trees, logistic regression, and neural networks.

2.3 Clinical Decision Support Systems

A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples. (Shin Markey 2006)

Medical decision support systems based on machine learning (Chi 2009)

3 Diagnosing Parkinson's Disease

BACKGROUND ON PD... NON ML DIAGNOSIS...

TLMR10 proposed monitoring Parkinsons disease using speech tests, due to vocal impairment being a common symptom and early indicator. The test results were passed through signal processing algorithms and a classification and regression tree to predict a rating on the unified PD rating scale.

DAlTH12 described a weakly supervised multiple instance learning approach to detecting symptoms of Parkinsons Disease. This approach addressed the issue of self-reporting resulting in inaccurate or incomplete labels.

Gil and Johnson (2009) used a multilayer network with one hidden layer and an output layer that output healthy or PD. The inputs were passed through a sigmoidal activation function, and gradient descent backpropagation was used to modify the weights. They achieved a classification accuracy of 92.31%. They also trained an SVM using the sequential minimal optimization (SMO) algorithm. SMO speeds up training of SVMs, particularly those with non-linear kernel functions (Platt 1998), using a divide and conquer approach. Gil and Johnson used a linear kernel with 91.79% accuracy, and a Pearson VII function kernel, with accuracy of 93.33%.

Mandal and Sairam also used a neural network with a sigmoidal activation function. They modified weights using backpropagation with dynamic learning rate and momentum, and achieved an accuracy of 97.6471%. They also used SVM with a linear kernel and obtained an accuracy of 97.6471%.s