Machine learning exam

Generalisationmeans that the machine learning algorythm that we trained can work well with new data it hasn´t seen before. It means it doesn´t overfit to the training data. To ensure good generalisaiton we can docross validation

stopping criteria

regularization ONMACHINE LEARNING EXAM SPECIFICALLY FOR YOUFOR ONLY$13. 90/PAGEOrder NowregularizationIt controls penalty for complexity. Models that are too complex might overfit frequently. So a simpler model might be better in some cases. Linear binary classifierLinear classifier seperate data in a straight line. They might be good when the data has clear boundaries and is easily distinguishable.

(a lot of dosts separated by straight line)non-linear binary classifier. if the the data is more spread out and it can not be linearly seperated a non-linear binary classifier might be better.

(a lot of dots seperated by a curvy line)the sample

distribution and the true distributionThe true distribution is the distribution that actually is happening in nature due to the fundamental properties of the issue at hand. Quite often the normal distribution is also the true distributionClassification typesBinary classification

multi-class classification

pair wise classification (m-1) x m/2the difference between Data Mining and Machine LearningData Mining is about using Statistics as well as other programming methods to find patterns hidden in the data so that you can explain some phenomenon. Data Mining builds intuition about what is really happening in some data and is still little more towards math than programming, but uses both.

Machine Learning uses Data Mining techniques and other learning algorithms to build models of what is happening behind some data so that it can predict future outcomes. Math is the basis for many of the algorithms, but this is more towards programming.

CFS pseudocodeCFS is an iterative procedure. Below are the steps your implementation should take:

1. Start with an empty set of selected features S_k, and a full set of initial features F, initialise k= 1

2. For each feature f in F, calculate the Pearson's product-moment correlation

r_cf between f and the target value t (i. e.

3. For each feature f in F, calculate the sum of correlations between f and all the features already in S_k

4. Select the feature that maximises CFS for this iteration, add it to S_k and remove it from F. Set k = k+1

5. Repeat steps 2-4 until the CFS value starts to drop (convergence)

ImplementCurse of the dimmensionality enemyBlessing of the non-uniformity. In most applications examples are not spread uniformly throughout

the instance space, but are concentrated on or near

a lower-dimensional manifold. brute-force searchor exhaustive search, also known as generate and test, is a very general problem-solving technique that consists of systematically enumerating all possible candidates for the solution and checking whether each candidate satisfies the problem's statement. Supervised Dimensionality Reduction• Neural nets: learn hidden

layer representation, designed

to optimize network prediction accuracy

• PCA: unsupervised, minimize reconstruction error

- but sometimes people use PCA to re-represent original data

before classification (to reduce dimension, to reduce overfitting)

• Fisher Linear Discriminant

- like PCA, learns a linear projection of the data

- but supervised: it uses labels to choose projectionnaive Bayes classifiersare a family of simple probabilistic classifiers based on applying Bayes' theorem with mstrong (naive) independence assumptions between the features. K-meansk-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results is a partitioning of the data space into Voronoi cells. Entropyis a measure of impurity (the opposite of information gain). ID3 algorythmThe first thing to be implemented was entropy function. It shows the purity of collection of examples:

Having the entropy we were able to calculate the gain. It shows which attribute has the best information value: