Machine Learning Summary

Maximilian Pfundstein

2018-12-28

Contents

1 Terms

1.1 Bagging

1.2 Bias-Variance-Trade-Off

1.3 Curse of Dimensionality

1.4 Degrees of Freedom

1.5 Generative vs Discriminative

1.6 Kernel Trick

1.7 Mean Squared Error (MSE)

1.8 Misclassification Rate

1.9 No Free Lunch Theorem

1.10 Ordinary Least Squares Regression (OLS)

1.11 Parametric and Non-Parametric Models

1.12 Types of Learning

1.12.1 Active Learning

1.12.2 Reinforcement Learning

1.12.3 Semi-supervised

1.12.4 Supervised Learning

1.12.5 Unsupervised Learning

1.13 Under- and Overfitting

2 Useful Code Snippets

2.1 Confusion Matrix and Misclassification Rate

2.2 Custom Error Function

Gam and Tree

2.3 Feature Plot

2.4 Histogram

2.5 Importing Data

Also include how to handle encoding issues.

2.5.1 .csv

2.5.2 .xls

2.5.3 .xlsx

2.6 RMarkdown Template

3 Models

3.1 Bayesian Classification

3.2 Boosting

3.2.1 AdaBoost

3.2.2 Forward Stagewise Additive Modeling

3.2.3 Gradient Boosting

3.3 Elastic Net

3.4 Generalized Additive Model (GAM)

Family parameter mgcv

3.5 Generalized Linear Model (GLM)

- Response Poisson distributed
- Canonical Link (log) is used for regression
- probabilistic expression for the fitted model

3.6 K-Nearest Neighbour (KNN)

3.7 Lasso

3.8 Least Absolute Deviation Regression

3.9 Linear Regression

3.10 Logistic Regression

- Equation of decision boundary
- Plot classified data and decision boundary
- GLM
- Custom Classification

3.11 Naive Bayes

3.12 Nearest Shrunken Centroid Classification (NSCC)

3.13 Neural Networks (NN)

Limitations and Types

3.13.1 Backpropagation Implementation

3.13.2 Implementation

Naive Bayes that uses nonparametric density estimation method Hint: density() function does not have predict() function but it evaluates predictions on a given grid. To make prediction for a vector of new values, you may call density() several times and specify one prediction point at a time, i.e. interval [a,b]=[x(i),x(i)].

3.13.3 Library

3.13.4 Regularization

3.14 Partial Least Squares Regression (PLS)

3.15 Quadratic Discriminant Analysis

3.16 Ridge Regression

3.17 StepAIC (AIC)

3.18 Support Vector Machines (SVM)

3.19 Trees

Decision, Regression, Pruning, Deviance, Gini, Blackboost, Random Forest, CART

4 Feature Reduction

4.1 Independent Component Analysis (ICA)

4.2 Linear Discriminant Analysis (LDA)

lda() in package mass

4.2.1 Implementation

4.2.2 Library

4.3 Principal Component Analysis (PCA)

Trace Plots

4.3.1 Kernel PCA

kpca in kernlab

4.4 Regularized Discriminant Analysis

5 Miscellaneous

5.1 Benjamin-Hochberg Algorithm

5.2 Bootstrapping

5.2.1 Confidence Band and Prediction Bands

5.2.2 Parametric and Non-Parametric

5.3 Cross-Validation

5.3.1 Cross Validation Plot

5.3.2 K-Fold

5.3.3 Nested Cross Validation

5.3.4 Two-Fold

5.4 EM-Algorithm

5.5 Holdout-Principle

5.6 K-Means Algorithm

5.7 Kernel Density Estimation

Epanechnikov kernel Exam 2015 2.3

5.8 Kernel Methods

5.8.1 Histogram Classification

5.8.2 Moving Windows Classification

5.9 Loss-Matrix

5.10 Probability Model and Log-Likelihood

From the engineers

5.11 ROC-Curve

TPR, FPR

5.12 Splines