Maschinelles Lernen und Data Mining (Arthur Flexer, WS05) – inoffizieller Übungsfragenkatalog

Zusammengestellt von Murmel (Murmel.vienna@gmx.at). Dies sind keine offiziellen Prüfungsfragen sondern nur selbsterdachte zur Vorbereitung. Laut Arthur kommen zur Prüfung 8 Fragen, jeweils eine zu jedem Kapitel (exklusive Intro).

The very basics of statistical Pattern Recognition

1) Was braucht man zum Klassifizieren? Wieso sind Histogramme dafür nicht gut genug? Wie lässt sich das Klassifikationsproblem formulieren?
2) Was ist maschinelles Lernen? Was der Konzeptraum? Welche Arten von Bias gibt es?
3) Was ist der Unterschied zwischen Klassifikation und Regression? Wovon sind sie eigentlich eine Generalisierung und worauf basiert das Ganze?
4) Was versteht man unter Bias-Variance tradeoff?

Bayes and all that

5) Wie funktioniert der naivste Klassifikationsansatz, der sich nur auf a-priori Wahrscheinlichkeiten stützt?
6) Formulieren Sie das Bayessche Theorem für das Klassifikationsproblem und leiten Sie es her. Wie nennt man die 4 Terme, die darin vorkommen? Zeigen Sie anhand einer Grafik, wieso durch dessen Verwendung beim Klassifikationsproblem der Fehler minimiert wird.
7) Warum ist das Bayssche Theorem so wichtig und was erlaubt es? Was ist wenn die A-priori Wahrscheinlichkeiten gleichhoch sind? Wie generalisiert man das Bayssche Theorem auf Wahrscheinlichkeitsdichten?

Probability Density Estimation

8) Welche Anwendungen gibt es für Wahrscheinlichkeitsdichtefunktionen? Welche Arten der Wahrscheinlichkeitsdichteschätzung existieren, was sind ihre Vor- und Nachteile?
9) Worauf basieren parametrische Methoden der PDE? Wodurch charakterisiert sich die Gaussche Verteilung in d Dimensionen und wie schätzt man ihre Parameter? Geben Sie die Formel für Likelihood und negative-Log likelihood an. Wie und wieso verwendet man letztere? Was ergibt sich daraus für die Parameter der Gausschen Verteilung?
10) Welche zwei Arten von nichtparametrischen Methoden gibt es für die PDE? Was versteht man unter dem „Parzen Window“? Was sind die Nachteile dieser Methoden?
11) Worauf basiert die semiparametrische Methode für die PDE? Wie schätzt man die Parameter eines GMM und was ergibt sich daraus? Wie geht man zur Expectation Maximization vor, welche 2 Schritte sind darin wichtig und wie heisst daher der dazugehörige Algorithmus?
12) Auf welchen Phasen basiert die auf spektraler Ähnlichkeit basierende Genre Klassifikation? Wie lassen sie sich auf Spracherkennung umlegen? In welche Phasen gliedert sich die Vorverarbeitung in MFCCs?
13) Wie entsteht die Similarity Matrix? Wie funktioniert K-nearest-neighbor Classification und wie gut ist One-nearest-neighbor? Welche Probleme kann es bei der Genre Klassifikation geben und wie kann man sie vermeiden?

Advanced Classification

14) Was versteht man unter Novelty (outlier) detection? Wie funktioniert ratio-reject? (nach welchen Formeln) Was ergibt dabei die Variation von s?
15) Was ist das doubt level? Auf welche 2 Arten kann man KNN dafür benutzen? Wann eignet sich outlier rejection besser und wann doubt levels?
16) Wie läuft die Evaluierung mit 3 verschiedenen Genres ab? Woraus setzt sich eine ROC in diesem Kontext zusammen? Geben Sie die Formel für Sensitivität und Spezifität an.

Statistical Evaluation of Machine Learning Experiments

17) Wieso ist statistische Evaluierung so wichtig und was ist ein Experiment?
18) Was versteht man unter re-sampling und k-fold cross validation? Wie sollte man insbesondere beim Vergleich von zwei Methoden vorgehen? Welche Varianten von cross-validation gibt es? Was ist der Zweck des dritten Testsets und wie groß sollte es sein?
19) Was sind H0, H1, alpha, der Typ 1 und der Typ 2 Fehler? Wieso verwendet man parametrische Tests? Welche Ergebnisse sollte man immer angeben? Was ist wenn N << 100?
20) Was sind die Vorteile von Konfidenzintervallen und was ist der McNemar Test? Welche Möglichkeiten gibt es, um den Effekt des multiplen Testens zu mildern?

Unsupervised Learning I: Visualization

21) Was sind die Ziele von unbeaufsichtigtem Lernen? Was die beiden Hauptthemen? Was ist insbesondere Visualisierung (Dimensionalitätsreduktion) und welche Lösungen dafür gibt es?
22) Was macht die PCA und wie funktioniert sie? Was ist insbesondere die SVD? Was sind die Grenzen der PCA und was wendet man in diesem Fall an? Was ist ICA und was sind ihre Vorteile?
23) Was ist statistische Unabhängigkeit, was der Unterschied zwischen Korreliertheit und Unabhängigkeit? Was hat letzteres mit Gaussheit zu tun und welche Arten von Algorithmen können damit in der ICA angewendet werden? Was ergibt sich daraus?
24) Welchen Zusammenhang gibt es zwischen FA, PCA und ICA? Wo kann die ICA praktisch angewendet werden?

Unsupervised Learning II: Clustering

25) Was versteht man unter Clustering? Was nimmt man hier meist als Distortion an? Wie funktioniert K-means clustering? Wo sind die Parallelen zum EM-Algorithmus und was sind die Nachteile von K-means?
26) Was für andere Arten von Algorithmen zur Clustererstellung existieren? Wie entscheidet man, wie viele Cluster es werden sollen? Wie kann man weitere Modelleigenschaften feststellen?
27) Durch welche Modifikation kann man beim Einsatz von GMMs zeitliche Information miteinbeziehen? Was sind MMs und was HMMs, wo liegt der Unterschied? Was ist die Markovketteneigenschaft?
28) Welche Wahrscheinlichkeiten braucht man für den Einsatz von HMMs, welche drei Probleme gilt es für deren Einsatz zu lösen und welche Algorithmen verwendet man dafür? Lohnt sich der Aufwand, HMMs für die Genreklassifikation zu verwenden?

Supervised Learning

29) Welche Skalenarten gibt es und wodurch charakterisieren sie sich? Für welche Art Merkmale kann Naive Bayes eingesetzt werden und was sind dessen Eigenschaften?
30) Was ist decision tree learning und wie entscheidet man, bei welchen Variablen gesplittet wird? Was sind Entropie und Informationsgewinn? Wie wird der Baum aufgebaut, wann kommt es zu overfitting und was kann dagegen getan werden? Was ist wenn keine Attribute mehr übrig sind?
31) Auf welchen Gedanken basieren SVMs? Was sind deren 2 „Tricks“? Wie reformuliert man deren Problemstellung und was macht man, wenn die Entscheidungsgrenze nichtlinear ist? Was sind damit die großen Vorteile von SVMs?
32) Woraus besteht das MLP? Was gibt seine softmax-Funktion aus?