

Hinweise aus letzter Vorlesung

THEMEN, DIE BESONDERS WICHTIG SIND

- **Lerntheorie**
- **SVM**
 - Basisprinzip: Expertenwissen, um SVM anzuwenden (Merkmalsraum kann kompliziert werden)
 - Erweiterung auf nichtlineare Probleme: siehe Beispiel aus Vorlesung (Klassifikation)
 - Merkmalsraum \leftrightarrow Hypothesenraum
- **Neuronale Netze**
 - Basisprinzip
 - Fehlerfunktion und Fehler ableiten
 - Gradientenabstieg \rightarrow Parameter anpassen
 - Aufbau Neuron
 - Aktivierungsfunktionen
- **Faltungsnetze**
 - Kerne, mit denen gefaltet wird (scheinbar nicht wirklich wichtig)
 - Schichten und Kerne: Aufbau
 - Dimensionen von Featuremaps: Veränderungen, Methoden
 - Max-Pooling
 - Verbindung von CNNs
- **Entscheidungsbäume**
 - Aufbau
 - Bestimmung von Attributen (*information gain*)
- **Sum-Product-Netze**
 - Wann ist SPN gültig?
 - SPN Verteilung, Scopes?
 - Mit Regeln testen, ob es passt
 - Kleine Beispiele kennen
 - Lernansätze (zB Parameterlernen, Strukturlernen mit vollständig beobachteten Daten) \rightarrow vor allem grundsätzliche Methoden
- **Reinforcement Learning**
 - Was heißt das?
 - Markovsche Entscheidungsfunktion und andere Funktionen
 - An kleinem Beispiel durchrechnen
- **Bayessches Lernen**
 - Bedingte Abhängigkeiten
 - Bayes-Ansatz durchrechnen
 - Annahmen (zB naive Bayes-Annahme)
 - Wohin führt Berechnung?

WEITERE TIPPS

- Definition VC-Dimension kennen (hängt von Topologie ab), zusammen mit CNNs \rightarrow Anzahl Neuronen in Topologie lässt Schlüsse auf VC zu
- Zöllner hat paar Aufgaben von WS17/18 gezeigt, will aber nicht alles zeigen, **weil einiges ähnlich ist!**
- Altklausur 3e: VC-Dimension zu groß, es gibt Overfitting \rightarrow Verbesserung: Anpassen der VC-Dim oder der Kapazität (verringern), mehr Beispiele
- Altklausur 3f: cascade correlation
- Aufgabe zu Q-Algorithmus wurde gezeigt, aber nicht im Detail auf Folien. Allerdings **V-Algorithmus** durchrechnen "ist auch recht wichtig" \rightarrow Berechnung zu V-Algorithmus kommt höchstwahrscheinlich dran
- Ergebnisse reichen, aber Bellmann-Gleichung schadet nicht
- Bei Wahrscheinlichkeiten: Aufgabenstellung beachten! Arithmetischer Ausdruck reicht oftmals
- Gradienten: Verwendung
- Boosting und Bagging mit anderen Verfahren

Altklausur WS17/18

AUFGABE 1 – ALLGEMEINE FRAGEN

1. Geben Sie 3 Einordnungskriterien von Lernverfahren, sowie deren Abstufungen bzw. Ausprägungen an.
2. Beschreiben Sie das Problem des **Overfitting**. Wie kann es verhindert werden?
3. Was verbirgt sich hinter der **Vapnik-Chervonenkis-Dimension**? Beschreiben Sie den theoretischen Nutzen in der Anwendung.

AUFGABE 2 – LERNEN VON PROBABILISTISCHEN MODELLEN

1. Der folgende Datensatz beschreibt Beobachtungen des Status eines bestimmten Zuges an einem bestimmten Bahnhof gegeben der ebenfalls beobachteten Attribute: Tag (T) \in {Wochentag, Wochenende}, Wind (Wi) \in {Kein, Leicht, Stark} und Wetter (We) \in {Sonne, Regen, Wind}.

Tag (T)	Wind (Wi)	Wetter (We)	Status (S)
Wochentag	Kein	Sonne	Verspätet
Wochentag	Kein	Nebel	Verspätet
Wochentag	Stark	Regen	Pünktlich
Wochenende	Leicht	Regen	Verspätet
Wochentag	Stark	Sonne	Pünktlich
Wochenende	Stark	Nebel	Pünktlich
Wochenende	Kein	Sonne	Verspätet
Wochentag	Leicht	Regen	Pünktlich
Wochenende	Leicht	Nebel	Verspätet

Berechnen Sie folgende (a-priori und bedingte) Wahrscheinlichkeiten:

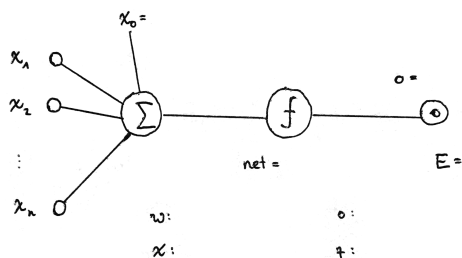
- $P(S = \text{Verspätet}) =$
 - $P(S = \text{Pünktlich}) =$
 - $P(T = \text{Wochenende} \mid S = \text{Pünktlich}) =$
2. Heute ist ein Wochentag mit Nebel und leichtem Wind. Ein naiver Bayes-Klassifikator soll genutzt werden, um den Status des Zuges vorherzusagen. Welches ist heute der wahrscheinlichste Status des Zuges? Begründen Sie Ihre Entscheidung.
 3. Das Szenario aus der vorherigen Teilaufgabe wird nun zusätzlich um das Attribut Jahreszeit (J) erweitert. Folgende Abhängigkeiten zwischen den Attributen sind gegeben:
 - Wetter (We) und Wind (Wi) sind abhängig von der Jahreszeit (J)
 - Der Status (S) ist abhängig von Wetter (We), Wind (Wi) und Tag (T)Zeichnen Sie ein Bayessches Netz, welches das Szenario beschreibt.
 4. Was kann bei einem Bayesschen Netz gelernt werden? Mit welcher Methode erfolgt dies, wenn die Struktur bekannt ist und Variablen nur teilweise beobachtbar sind?
 5. Ein HMM (*hidden markov model*) ist definiert als

$\lambda = \{S - \text{Zustände},$
 $V - \text{Ausgabezeichen},$
 $A - \text{Übergangswahrscheinlichkeiten},$
 $B - \text{Emmissionswahrscheinlichkeiten},$
 $\Pi - \text{Verteilung Anfangswahrscheinlichkeiten}\}.$

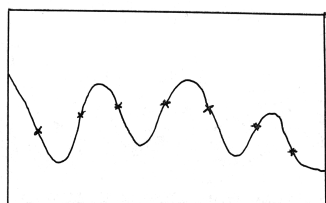
Beschreiben Sie das Lernproblem (was ist gegeben?, was ist gesucht?). Welcher Lernansatz kann dafür verwendet werden und welche Parameter werden dabei gelernt?

AUFGABE 3 – NEURONALE NETZE

1. Ergänzen Sie fehlende Begriffe und Formeln in der untenstehenden Abbildung eines Neurons.



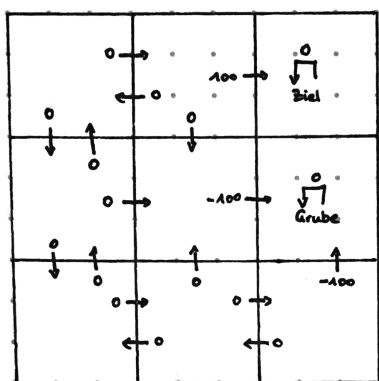
2. Nennen Sie zwei typische nichtlineare Aktivierungsfunktionen, die jeweils dazugehörige Formel und die jeweiligen Ableitungen.
3. Welche Bedingung muss die Aktivierungsfunktion eines Neurons für das Lernen unter Verwendung des Gradientenabstiegs erfüllen?
4. Nennen Sie ein Lernverfahren, um vorwärts gerichtete, mehrschichtige neuronale Netze zu trainieren und eine Herausforderung, die beim involvierten Gradientenabstieg auftreten kann. Mit welcher Methode kann dabei das Lernen verbessert werden?
5. Was lässt sich über die VC-Dimension des neuronalen Netzes sagen, das aus den untenstehenden Lerndaten die eingezeichnete Kurve approximiert? Nennen Sie das resultierende Phänomen und beschreiben Sie kurz, wie sich die Approximation verbessern lässt.



6. Nennen Sie ein konstruktives Verfahren zur schrittweisen Optimierung der Netzwerktopologie.

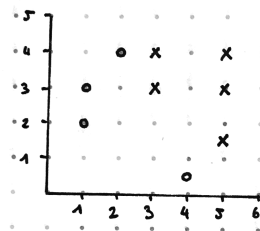
AUFGABE 4 – REINFORCEMENT LEARNING

1. Durch welches Modell lässt sich die Problemstellung beim Reinforcement Learning formal darstellen? Welche vier Bestandteile werden für die Modellierung benötigt?
2. Beschreiben Sie kurz die Markov-Bedingung.
3. Wie lautet die rekursive Definition der Q-Funktion (Bellmann-Gleichung)?
4. Nennen Sie die zwei wesentlichen Unterschiede zwischen den Suchstrategien *Exploration* und *Exploitation*.
5. Betrachten Sie das untenstehende Labyrinth. Ein Agent kann sich mit den angezeigten Zustandsübergängen von Raum zu Raum bewegen. Der Reward für einen Übergang ist jeweils in der Zeichnung abgebildet. Zu Beginn des Trainings gilt $\forall s, a : Q(s, a) = 0$. Führen Sie den Q-Learning-Trainingsalgorithmus durch und tragen Sie die Schätzung $\hat{Q}(s, a)$ für alle (s, a) nach konvergieren des Lernalgorithmus (Diskontierungsfaktor $\gamma = 0.8$) ein. Runden Sie die Ergebnisse auf natürliche Zahlen.



AUFGABE 5 – BOOSTING

Mit Hilfe von AdaBoost soll eine Klassifizierung durchgeführt werden. Hierzu soll ein Entscheidungsstumpf (1-Merkmal-Entscheidungsbaum mit lediglich einer Wurzel und zwei Blättern) verwendet werden. In jeder Iteration wählen Sie den Stumpf, der den gewichteten Trainingsfehler minimiert. Nutzen Sie hierzu die untenstehende Zeichnung.



1. Zeichnen Sie die Entscheidungsgerade in die obige Abbildung ein. Markieren Sie die positive und negative Seite der Klassifikation.
2. Berechnen Sie die Gewichtung jedes Datenpunktes nach der ersten Iteration und markieren Sie den Datenpunkt, der die höchste Gewichtung nach der ersten Iteration besitzt.
3. Erklären Sie anhand von AdaBoost die Anwendung der strukturellen Risikominimierung. Gehen Sie dabei auf die wesentlichen Aspekte der Hypothesenraumstrukturierung und der Fehlerminimierung ein.

AUFGABE 6 – SUPPORT VECTOR MACHINES

1. Beschreiben Sie kurz die Grundidee, die der Methode der Support-Vektor-Klassifikation zugrunde liegt und wie das Verfahren einzuordnen ist.
2. Geben Sie die Formeln für das Optimierungskriterium der **optimalen** Hyperebene und für die Randbedingung einer korrekten Klassifikation an (gegeben Trainingsbeispiele der Form (\vec{x}, y)).
 - Optimierungskriterium:
 - Randbedingung:
3. Erklären Sie die Dualität zwischen Hypothesenraum und Merkmalsraum im Kontext des SVM-Verfahrens (*version space duality*).
4. Welche Vorteile ergeben sich durch den sogenannten "Kerneltrick"? Welche Beobachtung liegt dem Kerneltrick zugrunde?