

SEMINAR: AKTIVES LERNEN

Organisation, Überblick, Themen

Überblick

- Organisatorisches
- Was ist aktives Lernen?
- Vorstellung der Themen

Organisatorisches

- „Aktives Lernen“
 - Seminar mit 2 SWS (3 LP)
- Ansprechpartner:
 - Christoph Sawade, Raum 03.04.0.20,
sawade@cs.uni-potsdam.de
 - Prof. Tobias Scheffer, Raum 03.04.0.17,
scheffer@cs.uni-potsdam.de
- Webseite:
 - <http://www.cs.uni-potsdam.de/ml/teaching/ws11/al.html>

Organisatorisches

Ablauf des Seminars

- Seminar findet als Blockveranstaltung statt

- Vorstellung der Themen und Literatur
 - Themenwahl per eMail an mich
 - Jedem Thema wird ein Betreuer zugewiesen, mit dem der Teilnehmer Termine zur Besprechung vereinbart

- Schriftliche Ausarbeitung und Seminarvortrag

Literatur / Empfehlung

- Vorlesung: „Maschinelles Lernen 1“
dienstags 10-14 Uhr
- Vorlesung: „Wissenschaftliches Arbeiten im Seminar“
als Video auf der Seminarwebseite
- Überblick: „Active Learning Literature Survey“
(B.Settles) frei im www

Was ist „überwachtes“ Lernen?

- Im maschinellen Lernen betrachten wir Modelle f_θ , die den Zusammenhang $p(y | \mathbf{x})$ zwischen Merkmalen $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ und einem Zielattribut $y \in \{-1, +1\}, \mathbb{R}, \dots$ gut approximieren.
- Beispiel: Spamfilter

$f_\theta : \text{Envelope} \mapsto$



Was ist „überwachtes“ Lernen?

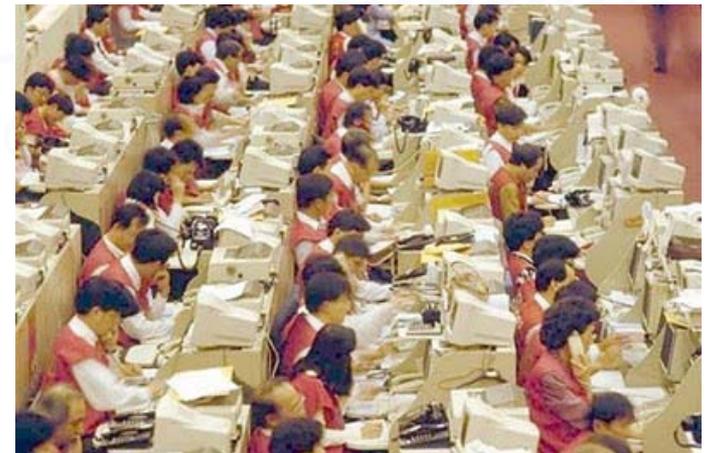
- Geg.: Ge-label-te Trainingsdaten $T = \{(\mathbf{x}_i, y_i) \mid i = 1, \dots, n\}$
- Ziel: Ein Modell f_θ mit hoher Genauigkeit auf (ungesehenen) Testdaten.



Was ist „überwachtes“ Lernen?

- Problem: Zwar sind unge-*label*-te Trainingsbeispiele häufig ausreichend verfügbar, doch die Akquirierung des zugehörigen Labels ist teuer

- Beispiel: Spamfilter
 - Genauigkeit: ~99,99%
 - Trainingsdaten: 2,5 Mio. eMails



Was ist aktives Lernen?

- Problem: Zwar sind unge-label-te Trainingsbeispiele häufig ausreichend verfügbar, doch die Akquirierung des zugehörigen Labels ist teuer
 - NLP, Textklassifikation
 - Biologische Probleme (z.B. Aussagen über Phäno- aus Genotyp)
 - Spamfilterung

- Ziel: Bestimmung des Modells mit höchster Güte für festes Labeling-Budget.

- Trivial-Ansatz: Zufälliges Ziehen bis Budget erschöpft
 - kein Bias, aber hohe Varianz

Was ist aktives Lernen?

- Geg.:
 - Pool von unge-label-ten Daten $U = \{\mathbf{x}_i \mid i = 1, \dots, m\}$
 - Festes Budget $n \ll m$

- Ziel: Bestimmung des Modells mit höchster Güte für festes Labeling-Budget.

Allgemeiner aktiver Lerner

□ Algorithmus

- Bis Labeling-Budget n erschöpft
 - Wähle eine Instanz x , die ge-label-t werden soll
 - Labeln der Instanz x
 - Update des Models f_θ
- Ausgabe Model f_θ

Themen

- Uncertainty Sampling
- Expected Error Reduction
- Duale aktive Lernverfahren
- Cluster-basierte aktive Lernverfahren
- Query by Committee
- Aktives Lernen mit Feature Feedback
- Stream-based Active Learning
- Aktives Lernen mit optimalen Ziehverteilungen
- Aktive Evaluierung

Uncertainty Sampling

- Version-Space: Menge aller Hypothesen, die mit den Trainingsdaten konsistent sind

$$VS_{\Theta, L} = \{ \theta \in \Theta \mid \forall (\mathbf{x}, y) \in T : f_{\theta}(\mathbf{x}) = y \}$$

- Ziel: Wähle die Menge an Beispielen, die den Version-Space minimiert.
- Heuristik: In jedem Schritt, das Beispiel wählen, das am nächsten an der Trennebene liegt...

S. Tong, D. Koller (2002): Support Vector Machine Active Learning with Applications to Text Classification

D. D.Lewis, W.A. Gale (1994): A Sequential Algorithm for Training Text Classifier

Expected Error Reduction

- Probabilistische Klassifikatoren liefern eine Wahrscheinlichkeitsschätzung für $p(y | \mathbf{x})$
- Damit lässt sich der erwartete Verlust eines Modells abschätzen: $\frac{1}{m} \sum_{i=1}^m \int \ell_{0/1}(f_{\theta}(\mathbf{x}), y) p(y | \mathbf{x}; \theta)$
- Idee: Erwarteten Fehler für jedes potentielle Beispiel und alle möglichen Label abschätzen; wähle das Beispiel mit kleinsten erwarteten Fehler

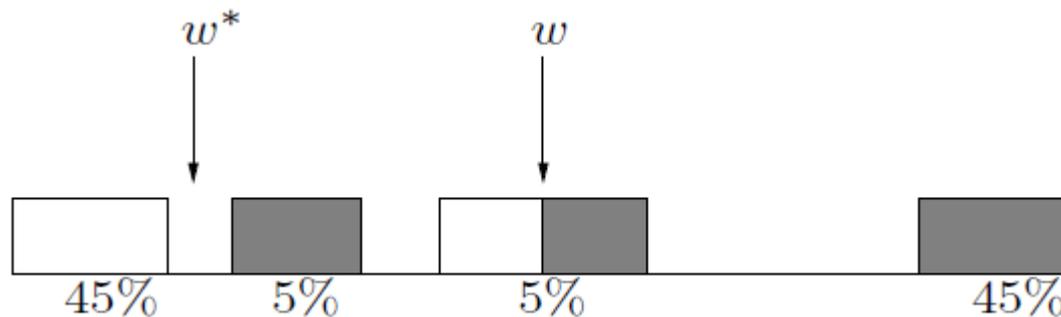
N. Roy, A. McCallum (2001): Toward Optimal Active Learning through Sampling Estimation of Error Reduction

O. Chapelle (2005): Active Learning for Parzen Window Classifier

G. Schohn, D. Cohn (2000): Less is More: Active Learning with Support Vector Machines

Missed-Cluster-Effect

- Deterministische Auswahl von Beispielen nahe der Entscheidungsgrenze kann zu lokalen Optima führen



Duale aktive Lernverfahren

- Uncertainty-Sampling leidet unter dem „Missed-Cluster-Effect“
- Exploitation vs. Exploration
- Verschiedene Heuristiken um „Uncertainty Sampling“ mit „Zufälligem Ziehen“ zu kombinieren

T. Osugi, D. Kun, S. Scott (2005): Balancing Exploration and Exploitation: A New Algorithm for Active Machine Learning

P. Donmez, J.G. Carbonell, P.N. Bennett (2007): Dual Strategy Active Learning

G. Pandey, H. Gupta, P. Mitra (2005): Stochastic Scheduling of Active Support Vector Learning Algorithms

Cluster-based Ansätze

- Im Gegensatz zur Exploration von $p(\mathbf{x})$ durch zufälliges Ziehen: Clustern des Eingaberaums
- Anschließende Sampling-Verfahren stellen sicher, dass aus jedem Cluster Beispiele gezogen werden

H.T. Nguyen, A. Smeulders (2004): Active Learning with Pre-Clustering

S. Dasgupta, D. Hsu (2008): Hierarchical Sampling for Active Learning

Query by Committee

- Menge von (unterschiedlichen) Modellen, die auf dem gleichen Trainingsdaten gelernt wurden.
- Modelle dürfen über zu *label*-ndes Beispiel abstimmen.

H. Seung, M. Opper H. Sompolinsky (1992): Query by Committee

A. McCallum, K. Nigam (1998): Employing EM in pool-based active learning for Text Classification

Aktives Lernen mit Feature Feedback

- Bisher nur Anfrage von *labels*
- Häufig ist aber auch Information über die Relevanz von *features* vorhanden.
- Sinnvoll, wenn Labels sehr komplex und zeitaufwändig zu bestimmen sind

H. Raghavan, J. Allan (2007): An InterActive Algorithm For Asking And Incorporating Feature Feedback into Support Vector Machines

G. Druck, B. Settles, A. McCallum (2009): Active Learning by Labeling Features

Stream-based Active Learning

- Algorithmus sieht eine Sequenz von (potentiell unendlich vielen) Beispielen
- Alle Daten können nicht gespeichert werden
 - Zu jedem Zeitpunkt *on-the-fly*-Entscheidung, ob aktuelles Beispiel *ge-label-ed* wird
 - *Label-n* von „alten“ Daten nicht möglich

A. Beygelzimer, S. Dasgupta, J. Langford (2009): Importance Weighted Active Learning

W. Chu, M. Zinkevich, ... (2010): Unbiased Online Active Learning in Data Streams

Aktives Lernen mit opt. Ziehverteilungen

- Beispiele werden nicht deterministisch, sondern nach einer Hilfsverteilung gezogen, die z.B.

$$\text{Erwartete Fehler} = \text{Bias}^2(f_\theta) + \text{Var}(f_\theta) + \text{Noise}$$

minimiert...

T. Kanamori (2006): Pool-based Active Learning With Optimal Sampling Distribution and Its Information Geometrical Interpretation

F. Bach (2007): Active Learning for Misspecified Generalized Linear Models

Aktive Evaluierung

- Aktive Lernverfahren bevorzugen schwer zu klassifizierende Beispiele.
- Darauf basierende Fehlerschätzung für das Modell ist negativ *ge-biased*.
- Welche Beispiele müssen für eine möglichst genaue Fehlerschätzung gezogen werden?

C. Sawade, N. Landwehr, S. Bickel, T. Scheffer (2010): Active Risk Estimation

C. Sawade, N. Landwehr, T. Scheffer (2010): Active Estimation of F-Measures

Themen

- Uncertainty Sampling
- Expected Error Reduction
- Duale aktive Lernverfahren
- Cluster-basierte aktive Lernverfahren
- Query by Committee
- Aktives Lernen mit Feature Feedback
- Stream-based Active Learning
- Aktives Lernen mit optimalen Ziehverteilungen
- Aktive Evaluierung

Termine

- Sonntag, d. 04.12.2011
 - Abgabe: 1. Version der Ausarbeitung

- Dienstag, d. 03.01.2012
 - Abgabe: Endversion der Ausarbeitung
 - Abgabe: 1. Version der Vortragsfolien

- Freitag, d. 13.01.2012
 - Seminarvortrag