

**БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИНСТИТУТ ПО МАТЕМАТИКА И ИНФОРМАТИКА**

сигнатура:				
4.6	I	S	13	01
професионално направление	код на докт. програма	вид курс (базов/спец.)	номер	версия
<i>попълва се административно след приемане от НС на ИМИ</i>				

Утвърдил:
(проф. д-мн П. Бойваленков, Директор на ИМИ-БАН)

**Учебна програма
за специализиран докторантски курс**

Област на висше образование:	4. Природни науки, математика и информатика
професионално направление:	4.6. Информатика и компютърни науки
докторска програма:	Информатика
тема:	Машинно самообучение
лектор:	проф. д-р Десислава Панева-Маринова
данни за връзка с лектора (тел., имейл)	+359888894814, dessi@cc.bas.bg
хорариум:	20 часа лекции и 20 часа практически упражнения
кредити съгл. кредитната система на ЦО на БАН:	20

1. Анотация

Учебният курс цели запознаване с основните алгоритми за машинно самообучение – алгоритъм FINDS, класификационни дървета, генетични алгоритми, невронни мрежи, теорема на Бейс, к най-близки съседи и др.

2. Необходими предварителни знания

няма

3. Компетентности, придобити в резултат на обучението

Знания и умения за използване на различни алгоритми за машинно самообучение. След завършване на курса докторантите се очаква да могат:

- Прилагане на различни алгоритми за машинно самообучение;
- Разпознаване кога и при какви условия да се приложи съответният алгоритъм;
- Оценка на приложения алгоритъм;
- Оценка на точността на хипотеза;
- Сравнение на точността на хипотеза.

4. Тематично съдържание

Тема	брой часове лекции	брой часове практически упражнения
Алгоритъм FIND-S.	1	1
Алгоритъм за елиминирание на кандидати.	1	1
Класификационни дървета.	1	2
Класификационни дървета и обработка на непрекъснати атрибути и обработка на примери с липсващи стойности на атрибути.	1	2
Алгоритми за последователно покриване – програма – FOIL.	1	1
Генетични алгоритми.	2	1
Невронни мрежи.	2	1
Алгоритъм BACKPROPAGATION.	1	1
Теоремата на Бейс.	1	1
Оптимален Бейсов класификатор. Наивен Бейсов класификатор.	1	1
Бейсови мрежи: Представяне на Бейсови мрежи.	1	1
Мерки за сходство. Обработка на липсващите признаци. Алгоритъм на най-близкия съсед. Алгоритъм на k най-близки съседи.	2	2
Определяне на релевантни атрибути - алгоритъм Relief-F.	1	1
Машина на поддържащи вектори.	1	1
Самообучение чрез ансамбли: Bagging. Рандомизация. Boosting. Stacking.	1	1

Оценяване на точността на хипотезата. Извадкова грешка и истинска грешка. Оценяване на грешка и биномното разпределение. Доверителни интервали и нормалното разпределение.	1	1
Сравнение на точността на хипотези. Сравнение на алгоритми за обучение. t-тест по двойки. Тест на Уилкоксон за знаково ранжиране по двойки.	1	1

5. Конспект

1. Алгоритъм FIND-S.
2. Алгоритъм за елиминиране на кандидати.
3. Класификационни дървета – базов алгоритъм, обучение с класификационни дървета.
4. Класификационни дървета и обработка на непрекъснати атрибути и обработка на примери с липсващи стойности на атрибути.
5. Алгоритми за последователно покриване – програма FOIL.
6. Генетични алгоритми.
7. Невронни мрежи.
8. Алгоритъм BACKPROPAGATION.
9. Теоремата на Бейс.
10. Оптимален Бейсов класификатор. Наивен Бейсов класификатор.
11. Бейсови мрежи: Представяне на Бейсови мрежи.
12. Мерки за сходство. Обработка на липсващите признаци. Алгоритъм на най-близкия съсед. Алгоритъм на k най-близки съседи.
13. Определяне на релевантни атрибути - алгоритъм Relief-F.
14. Машина на поддържащи вектори.
15. Самообучение чрез ансамбли: Bagging. Рандомизация. Boosting. Stacking.
16. Оценяване на точността на хипотезата. Извадкова грешка и истинска грешка. Оценяване на грешка и биномното разпределение. Доверителни интервали и нормалното разпределение.
17. Сравнение на точността на хипотези. Сравнение на алгоритми за обучение. t-тест по двойки. Тест на Уилкоксон за знаково ранжиране по двойки.

6. Препоръчана литература:

1. Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.
2. Theobald, O. (2017). *Machine Learning for Absolute Beginners: A Plain English Introduction*, Scatterplot Press.
3. Burkov, A. (2019). *The Hundred-Page Machine Learning Book*, Andriy Burkov.
4. Müller, A. C., & Guido, S. (2016). *1. Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly.
5. Barber, D. (2010). *Bayesian Reasoning and Machine Learning*, Cambridge University Press.
6. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning*. Cambridge University Press.
7. Hurwitz, J., & Kirsch, D. (2018). *Machine Learning For Dummies*. John Wiley & Sons, Inc.
8. Zaki, M. J., & Meira, Jr, W. (2020). *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge University Press.

7. Ресурсно осигуряване на обучението:

Не е предвидено специализирано ресурсно осигуряване.

(хардуер, софтуер, данни, ...)

8. Критерии за оценка

Изпитът е с продължителност 4 часа и се състои от две части – писмен и устен.

На писмения изпит докторантът развива своите идеи и концепции по два въпроса от конспекта.

На устния изпит докторантът отговаря на зададени от журито въпроси, свързани с темата на курса.

Крайната оценка е от 2 до 6 (с точност до 0.5).

Тя се формира на базата на следното съответствие:

Отличен (6 или 5.50)	Отлично владее материала. Изложението е изчерпателно, последователно, компетентно, логично и хармонично. Правилно обосновава предлаганите решения, знае как да обобщава и излага материала без да прави грешки. Притежава необходимите умения за изпълнение на практически задачи.
Мн. добър (5 или 4.50)	Познава материала. Излага го правилно без да допуска съществени неточности. Може правилно да прилага теоретични принципи и притежава необходимите умения за изпълнение на практически задачи.
Добър (4 или 3.50)	Владее голяма част от материала, но допуска неточности при изложението и отговорите на въпросите. Има известни неясноти при опитите за прилагане на материала в практически ситуации.
Среден (3)	Владее само част от материала, но се затруднява в отделните детайли. Допуска неточности във формулировките и нарушава последователността при представянето на материал. Има затруднения при изпълнение на практически задачи.
Слаб (2)	Не познава значителна част от материала, допуска съществени грешки и с големи трудности изпълнява практически задачи.

Учебната програма е обсъдена и одобрена на заседание на секция „Математическа лингвистика“ на 5.01.2022 г.

Ръководител секция:

(проф. д-р Десислава Панева-Маринова)

Разгледана от Директорския съвет на ИМИ-БАН на 20.01.2022 г. (протокол № 3).

Приета от Научния съвет на ИМИ-БАН на 21.01.2022 г. (протокол № 1).