

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИНСТИТУТ ПО МАТЕМАТИКА И ИНФОРМАТИКА

сигнатура:				
4.5	TPMS	S	01	v1
професионално направление	код на докт. програма	вид курс (базов/спец.)	номер	версия
<i>попълва се административно след приемане от НС на ИМИ</i>				

Утвърдил:
(акад. В. Дренски, Директор на ИМИ-БАН)

Учебна програма
за специализиран докторантски курс

Област на висше образование:	4. Природни науки, математика и информатика
професионално направление:	4.5. Математика
докторска програма:	Теория на вероятностите и математическа статистика
тема:	Непараметрично оценяване на плътността
лектор:	проф. дмн Евгения Стоименова
данни за връзка с лектора (тел., имейл)	Тел.: 0889 508 939 Email: jeni@math.bas.bg
хорариум:	30 часа лекции
кредити съгл. кредитната система на ЦО на БАН:	20

1. Анотация

Непараметричното оценяване на плътността е атрактивен подход за анализ на данни, чрез който лесно може да се реши проблема за графичната илюстрация на едномерни и многомерни данни. Други важна задача, в която непараметричните оценки на плътността намират приложение, е откриването на регресионна връзка между няколко променливи, което е част от множество изследвания в разнообразни области на природните и социалните науки. При използването на непараметрични методи за оценка на плътността се допускат по-малко предположения при конструирането на модел за данните и не се налагат допълнителни условия за разпределението, от което произлизат наблюденията.

2. Необходими предварителни знания

Изискват се основни познания по теория на вероятностите и математическа статистика.

3. Компетентности, придобити в резултат на обучението

Докторантите ще развият своята математическа интуиция и умения, свързани с непараметричния анализ на едномерни и многомерни данни.

4. Тематично съдържание

Тема	брой часове лекции
Хистограми и методи за оценка на плътността	2
Едномерни и многомерни оценки, базирани на ядра	8
Непараметрична регресия, базирана на ядра	2
Многомерна непараметрична регресия и адитивни модели	2
Метод на най-близкото съседство	2
Оценки, базирани на максимално правдоподобие с пенализация	2
Приложения на изглаждане чрез ядра	4
Непараметричен дискриминантен анализ	2
Непараметрични и полупараметрични обобщени линейни модели	4
Многомерни адаптивни хистограми	2

5. Конспект

1. Хистограми и методи за оценка на плътността (Наивни оценки; Оценки, базирани на ядра; Метод на най-близкото съседство; Оценки, базирани на ортогонални редове; Оценки, базирани на максимално правдоподобие с пенализация; Оценки, базирани на обобщени теглови функции).
2. Едномерни оценки, базирани на ядра. Критерии, основани на средна квадратична грешка (MSE) и средна интегрирана квадратична грешка (MISE). (Асимптотични приближения на MSE и MISE; Точни изчисления, свързани с MISE).
3. Канонични ядра и оптимална теория за ядра. Ядра от по-висок порядък. Модификации на оценките на плътността, базирани на ядра (Локални оценки, базирани на ядра; Трансформирани оценки, базирани на ядра).
4. Избор на параметъра на изглаждане (Методи, основани на нормална плътност; Метод на преизглаждане; Методи, базирани на cross-validation).
5. Сравнение на методите за избор на параметъра на изглаждане (Теоретично представяне; Изчислителни съображения).
6. Асимптотични свойства на едномерните оценки, базирани на ядра (Състоятелност; Скорост на сходимост).
7. Многомерни оценки, базирани на ядра (Асимптотични приближения на MISE; Точни изчисления, свързани с MISE).

8. Избор на многомерното ядро и параметъра на изглаждане (Изчислителни съображения; Трудности в пространства от по-висока размерност).
9. Непараметрична регресия, базирана на ядра (Оценка на Nadaraya–Watson; Локално полиномни оценки, базирани на ядра; Асимптотични приближения на MSE; Избор на параметъра на изглаждане).
10. Многомерна непараметрична регресия и адитивни модели. Сравнение между непараметричните регресионни методи.
11. Метод на най-близкото съседство (Избор на параметъра на изглаждане; Изчислителни съображения).
12. Адаптивни оценки, базирани на ядра (Метод на Breiman, Meisel и Purcell; Избор на параметъра на изглаждане).
13. Оценки, базирани на максимално правдоподобие с пенализация (Подход на Good и Gaskins; Метод на дискретно правдоподобие с пенализация; Пенализация на логаритъма на плътността).
14. Оценки на плътността, базирани на ядра, при зависими наблюдения, изместени наблюдения, дясно-цензурирани данни и данни с грешки при измерванията
15. Приложения на изглаждане чрез ядра (Оценка на хазартната функция; Оценка на спектралната плътност; Регресионни модели, базирани на правдоподобие; Оценка на функцията на интензивност).
16. Изглаждане на наредени категорни данни (Честотни таблици; Оценка на плътността; Регресия).
17. Непараметричен дискриминантен анализ (Подход чрез непараметрична оценка на плътността; Обобщения за дискретни и смесени данни).
18. Приложения на оценка на плътността в клъстерния анализ (Йерархичен метод; Метод, базиран на редици).
19. Непараметрични и полупараметрични обобщени линейни модели (Оценки, базирани на максимално правдоподобие с пенализация; Fisher scoring; Cross-validation)
20. Полупараметрични и структурни модели (Едноиндексен модел; Адитивен модел; Частично линеен модел).
21. Многомерни адаптивни хистограми (Greedy хистограми; Хистограми, базирани на ма класификация и регресионни дървета; Bootstrap aggregation).

6. Препоръчана литература:

1. Silverman, B. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. London - New York: Chapman and Hall.
2. Wand, M. and Jones, M. (1995). *Kernel smoothing*. Chapman & Hall., London.
3. Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press.
4. Simonoff, J. S. (1996). *Smoothing methods in statistics*. Springer, New York.
5. Klemelä, J. (2009). *Smoothing of multivariate data. Density estimation and visualization*. Hoboken, NJ: John Wiley & Sons.
6. Klemelä, J. (2014). *Multivariate Nonparametric Regression and Visualization: With R and Applications to Finance*. Wiley.

7. Ресурсно осигуряване на обучението:

Не е необходимо.

8. Критерии за оценка

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

На писмения изпит докторантът решава една задача по темата на курса и развива своите идеи и концепции по един въпрос 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)	Мн.добър (5)	Добър (4)	Среден (3)	Слаб (2)
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.	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.	Владее голяма част материала, но допуска неточности при изложението и отговорите на въпросите. Има известни неясноти при опитите за прилагане на материала в практически ситуации.	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.	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 „Изследване на операциите, вероятности и статистика“ на 10.03.2020 г.

Ръководител секция:

(проф. д-р Николай Живков)

Учебната програма е разгледана от Директорския съвет на ИМИ-БАН на 12.03.2020 г. (протокол № 10).

Учебната програма е приета от Научния съвет на ИМИ-БАН на 13.03.2020 г. (протокол № 4).