

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Кафедра ІМІ

## КОМПЛЕКС НАВЧАЛЬНО-МЕТОДИЧНОГО ЗАБЕЗПЕЧЕННЯ

навчальної дисципліни

«Математичні методи аналізу даних та обробки мультимедійної інформації»  
підготовки  
докторів філософії  
за спеціальністю 172 «Телекомунікації та радіотехніка»

Розробник                      А.В. Омельченко, доц. каф. ІМІ, к.т.н., доцент

Схвалено на засіданні кафедри ІМІ

Протокол № 9 від «10» 01 2020 р.

Схвалено методичною комісією факультету Інфокомунікацій

Протокол № 6 від «30» 01 2020 р.

Харків  
2020 р.

## ЗМІСТ

1. Робоча програма .....3
2. Методичні вказівки до самостійної роботи .....15

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій  
(повна назва)

Кафедра Інформаційно-мережної інженерії  
(повна назва)

ЗАТВЕРДЖУЮ  
Декан факультету Інфокомунікацій

А.В. Снігуров  
(підпис, ініціали, прізвище)

« \_\_\_\_\_ » \_\_\_\_\_ 2020 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ  
Математичні методи аналізу даних та обробки мультимедійної інформації  
(назва навчальної дисципліни)

рівень вищої освіти третій (освітньо-науковий)  
(бакалаврський, магістерський, освітньо-науковий)

спеціальність 172 Телекомунікації та радіотехніка  
(код і повна назва спеціальності)

освітньо-наукова програма  
(професійна або наукова)

172 Телекомунікації та радіотехніка  
(повна назва програми)

Харків – 2020 р.

Розробник(и): А.В. Омельченко, доцент кафедри ІМІ, к.т.н., доцент  
(ініціали, прізвище, посада, науковий ступінь, вчене звання)

Робочу програму схвалено на засіданні кафедри ІМІ

Протокол від « 10 » 01 2020 р. № 9

Завідувач кафедри \_\_\_\_\_ В.М. Безрук  
(підпис) (ініціали, прізвище)

Керівник групи забезпечення спеціальності \_\_\_\_\_ М.В. Москалець  
(підпис) (ініціали, прізвище)

Схвалено методичною комісією факультету інфокомунікацій

Протокол від “ 30 ” 01 2020 р. № 6

Голова методичної комісії \_\_\_\_\_ Д.В Чеботарьова  
(підпис) (ініціали, прізвище)

## 1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни			
		денна форма навчання	заочна форма навчання		
Кількість кредитів ЄКТС <u>8</u>	Галузь знань 17 «Електроніка та телекомунікації»	Вибіркова компонента ОП			
	Спеціальність підготовки 172 «Телекомунікації та радіотехніка»				
Модулів	Освітній ступінь: доктор філософії	Рік підготовки:			
Змістових модулів 2		1-й			
		Семестр			
Загальна кількість годин <u>240</u>		1-й	2-й		
		Кількість годин			
		120	120		
		Аудиторні: 1) лекції, год			
Тижневих годин для денної форми навчання: аудиторних – 2 самостійної роботи аспіранта - 5		2) практичні, год			
		-	-		
	3) лабораторні, год				
	-	-			
	4) консультації, год				
	20	20			
	Самостійна робота, год				
	100	100			
	в тому числі: 1) інд. завд., год.				
	-	-	-	-	
2) курсова робота, год					
	-	-	-		
Вид контролю: залік					

## 2 МЕТА І ЗАВДАННЯ ДИСЦИПЛІНИ

### 2.1 Мета навчальної дисципліни

Методи аналізу даних (Data mining) та обробки мультимедійної інформації – важливий напрямок в області інформаційних технологій, що активно розвивається. При дослідженні складних систем, якими є інформаційні мережі зв'язку, виникає необхідність в обробці та узагальненні великих обсягів різномірної інформації. Для цього використовуються методи дискретної математики, регресійного аналізу, перевірки гіпотез, класифікації та кластеризації.

Метою дисципліни є формування у аспірантів знань, умінь і навичок з методів і алгоритмів отримання знань із експериментальних даних та обробки мультимедійних даних.

Дисципліна «Математичні методи аналізу даних та обробки мультимедійної інформації» відноситься до вибіркових дисциплін професійної та практичної підготовки за освітньо-професійною програмами Телекомунікації та радіотехніка за спеціальністю 172 Телекомунікації та радіотехніка.

Навчальна дисципліна використовується для формування наступних компетентностей: здатність до пошуку, систематичного вивчення та аналізу науково-технічної інформації, світового досвіду, пов'язаного із застосуванням методів телекомунікацій та радіотехніки для дослідження різноманітних процесів, явищ та систем.

### 2.2 Завдання дисципліни

Відповідно до мети викладення дисципліни головними завданнями викладення дисципліни є:

- засвоєння аспірантами математичних методів аналізу даних, достатніх для проведення наукових досліджень;
- ознайомлення аспірантів з технологіями аналізу даних на рівні, достатньому для формування уявлень, які б розширювали професійний кругозір;
- засвоєння аспірантами методів та алгоритмів обробки мультимедійних даних.

Програмні результати навчання: набуття знань та розуміння основних методів аналізу даних та вміння застосовувати інструменти та моделі аналізу даних (апаратно-програмні ресурси, пакети прикладних програм, онлайн ресурси й відповідні технології) в дослідженні реальних систем та презентації результатів наукових досліджень у різних формах; здійсненню науково-педагогічної діяльності з використанням цих ресурсів та технологій.

Після вивчення дисципліни аспіранти мають знати:

1. Критерії перевірки гіпотез Колмогорова, Смирнова,  $\chi^2$ -квадрат Пірсона.
2. Методи кореляційного аналізу.

3. Методи регресійного аналізу.
4. Методи локального ядерного аналізу.
5. Моделі процеси авторегресії і процеси ковзного середнього.
6. Методи прогнозування часових рядів.
7. Байєсівський підхід до побудови оптимальних вирішальних правил розпізнавання.
8. Алгоритми класифікації на кластеризації даних.
9. Методи кодування мовних сигналів.
10. Методи подання зображень в цифровій формі та методи покращення зображень.
11. Методи кодування цифрових зображень та відео послідовностей.
12. Неперервні вейвлет-перетворення SWT та дискретні вейвлет-перетворення DWT.

### 3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

#### Осінній семестр

##### **Змістовий модуль 1. Методи статистичного аналізу даних**

Тема 1. Критерії перевірки статистичних гіпотез

Тема 2. Методи класифікації даних

Тема 3. Кластерний аналіз

Тема 4. Кореляційний аналіз

Тема 5. Регресійний аналіз

#### Весняний семестр

##### **Змістовий модуль 2. Методи обробки мультимедійної інформації**

Тема 6. Аналіз часових рядів та їх прогнозування

Тема 7. Методи кодування мовних сигналів. Вокодери

Тема 8. Подання зображень в цифровій формі та методи покращення зображень

Тема 9. Кодування цифрових зображень та відеопослідовностей

Тема 10. Вейвлет-аналіз і його використання при стисненні зображень.

## 4 СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин											
	денна форма						Заочна форма					
	Усього го	у тому числі					Усього	у тому числі				
		л	п	лб	конс	с.р.		л	п	лб	конс	с.р.
<b>Змістовий модуль 1. Методи статистичного аналізу даних</b>												
<b>Тема 1.</b> Критерії перевірки статистичних гіпотез	16				2	14	16				2	14
<b>Тема 2.</b> Методи класифікації даних	30				6	24	30				6	24
<b>Тема 3.</b> Кластерний аналіз	28				4	24	28				4	24
<b>Тема 4.</b> Кореляційний аналіз	16				2	14	16				2	14
<b>Тема 5.</b> Регресійний аналіз	30				6	24	30				6	24
Разом за зміст. мод. 1	120				20	100	120				20	100
<b>Змістовий модуль 2. Методи обробки мультимедійної інформації</b>												
<b>Тема 1.</b> Аналіз часових рядів та їх прогнозування	24				4	20	24				4	20
<b>Тема 2.</b> Методи кодування мовних сигналів. Вокодери	24				4	20	24				4	20
<b>Тема 3.</b> Подання зображень в цифровій формі та методи покращення зображень	24				4	20	24				4	20
<b>Тема 4.</b> Кодування цифрових зображень та відеопослідовностей	24				4	20	24				4	20
<b>Тема 5.</b> Вейвлет-аналіз і його використання при стисненні зображень.	24				4	20	24				4	20
Разом за зміст. мод. 2	120				20	100					20	100
<b>Усього годин</b>	<b>240</b>	-	-	-	<b>40</b>	<b>200</b>	<b>240</b>	-	-	-	<b>40</b>	<b>200</b>



## 5 ЛАБОРАТОРНІ РОБОТИ

Лабораторні роботи за навчальним планом для даної дисципліни відсутні.

## 6 ПРАКТИЧНІ ЗАНЯТТЯ

Практичні заняття за навчальним планом для даної дисципліни відсутні.

## 7 САМОСТІЙНА РОБОТА

№ зміст модулю	Теми самостійної роботи	обсяг, год.	вид контролю	літ. джер.
Змістовий модуль 1. Методи статистичного аналізу даних				
1	Перевірка гіпотез	14	усне опитування	1-3, 7,8
1	Методи класифікації даних	24	усне опитування	4-6
1	Кластерний аналіз	24	усне опитування	4-6
1	Кореляційний аналіз	14	усне опитування	4
1	Регресійний аналіз	24	усне опитування	1, 3, 7, 8
Змістовий модуль 2. Методи обробки мультимедійної інформації				
2	Аналіз часових рядів та їх прогнозування	24	усне опитування	9, 10
2	Методи кодування мовних сигналів. Вокодерери	24	усне опитування	14-17
2	Подання зображень в цифровій формі та методи покращення зображень	24	усне опитування	18, 19
2	Кодування цифрових зображень та відеопослідовностей	24	усне опитування	18, 19
2	Вейвлет-аналіз і його використання при стисненні зображень	24	усне опитування	20
<b>Загальна кількість</b>		<b>200</b>		

## 8 ІНДИВІДУАЛЬНІ ЗАВДАННЯ

За результатами навчання аспіранти готують реферат за узгодженою з викладачем темою, що пов'язана з темою їх наукових досліджень.

## 9 МЕТОДИ НАВЧАННЯ

Основні методи навчання:

- робота з навчально-методичною літературою (конспектування, самостійне опрацювання заданих розділів, написання реферату тощо).
- перевірка знань та умінь (за відповідями на контрольні запитання, результатами виконання контрольних робіт, індивідуальних завдань, доповіді за темою реферату).

## 10 РЕЙТИНГОВА ОЦІНКА ЗА ДИСЦИПЛІНОЮ

### 10.1 Кількісні критерії оцінювання

Для оцінювання роботи аспіранта протягом семестру підсумкова рейтингова оцінка  $Q_{\text{сем}}$  розраховується як сума оцінок по кожній темі та оцінка за реферат. Кожна тема оцінюється в 5 балів. Оцінка за реферат – 50 балів. Максимальна рейтингова оцінка протягом семестру – 100 балів.

### 10.2 Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

1. Критерій узгодженості Колмогорова.
2. Критерій однорідності Смирнова.
3. Критерій  $\chi^2$ -квадрат Пірсона.
4. Модель регресійного аналізу.
4. Оцінювання коефіцієнтів регресії за методом найменших квадратів.
5. Локальний ядерний аналіз.
6. Фільтри Савицького-Голея.
7. Різницеві фільтри і їх застосування для інтерполяції даних.
8. Байєсівський підхід до побудови оптимальних вирішальних правил розпізнавання.
9. Вирішальні правила розпізнавання з використанням нормальних розподілів векторів ознак.
10. Алгоритми класифікації:  $k$  найближчих сусідів; наївний байєсівський класифікатор; класифікація з використанням дерев рішень.
11. Метрики класифікації і кластеризації.
12. Алгоритми кластерного аналізу:  $k$ -means; fuzzy  $k$ -means; EM; FOREL; з використанням штучних нейронних мереж.
13. Кореляційний аналіз з використанням коефіцієнтів кореляції Пірсона; Спірмена та Кендала.
14. Процеси авторегресії і процеси ковзного середнього.
15. Прогнозування часових рядів.
16. Умовне сподівання як оптимальний прогноз.

17. Лінійне прогнозування стаціонарного часового ряду.
18. Алгоритми ІКМ та АДІКМ
19. Класифікація вокодерних алгоритмів стиснення мовних сигналів.
20. Алгоритми стиснення мовних сигналів на основі моделі лінійного передбачення.
21. Стандарти стиснення даних JPEG та JPEG 2000.
22. Стандарти стиснення відеопослідовностей.
23. Неперервні вейвлет-перетворення SWT.
24. Дискретні вейвлет перетворення DWT.

Необхідний обсяг умінь для одержання позитивної оцінки.

1. Вміти застосовувати критерії перевірки гіпотез про функцію розподілу, однорідність і незалежність вибірок.
2. Вміти оцінювати коефіцієнти регресії за методом найменших квадратів.
3. Вміти знаходити коефіцієнти фільтрів Савицького-Голя та різницевих фільтрів.
5. Вміти обчислювати коефіцієнти моделі авторегресії.
6. Вміти прогнозувати стаціонарні часові ряди.
7. Вміти будувати вирішальні правила класифікації з використанням нормальних розподілів векторів ознак.
8. Вміти застосовувати алгоритми кластеризації даних.
9. Вміти вибирати параметри дискретизації і квантування мовних сигналів і зображень.
10. Вміти застосовувати алгоритми стиснення мовних сигналів.
11. Вміти застосовувати алгоритми стиснення зображень.
12. Вміти здійснювати вейвлет-перетворення Хаара і застосовувати його для стиснення зображень.

10.3 Критерії оцінювання роботи аспіранта протягом семестру.

**Задовільно, D, E (60-74).** Мати мінімум знань та вмінь. Опрацювати усі теми, виконати реферат.

**Добре, C (75-89).** Твердо знати матеріал. Опрацювати усі теми, виконати реферат. Знати методи аналізу даних і вміти застосовувати їх на практиці.

**Відмінно, A, B (90-100).** Твердо знати всі теми. Орієнтуватися в підручниках та учбових посібниках. Опрацювати усі теми, виконати реферат. Досконально знати методи аналізу даних і вміти застосовувати їх на практиці.

## Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
96 – 100	<b>A</b>	відмінно добре задовільно	зараховано
90-95	<b>B</b>		
75-89	<b>C</b>		
66-74	<b>D</b>		
60-65	<b>E</b>		
35-59	<b>FX</b>	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	<b>F</b>	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

## НАВЧАЛЬНО–МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ДИСЦИПЛІНИ ТА РЕКОМЕНДОВАНА ЛІТЕРАТУРА

### 11.1 Основна література:

1. Костин В.Н., Тишина Н.А. Статистические методы и модели: Учебное пособие. – Оренбург: ГОУ ОГУ, 2004. – 138 с.
2. Ивченко Г.И., Медведев Ю.И. Математическая статистика. М. 1984.
3. Лавренчик В.Н. Постановка физического эксперимента и статистическая обработка его результатов: Учеб.пособие для вузов. –М.: Энергоатомиздат, 1986.- 272 с.
4. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
5. Барсегян А. А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.
6. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М.: Мир, 1976.
7. Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. М.Финансы и статистика. -1987.
8. Себер Дж. Линейный регрессионный анализ. М.: Мир. 1980.
9. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. – М.:Мир, 1974. – Вып.1,2.
10. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. – М.: Финансы и статистика, 2003. – 414 с.
11. Циплаков А. «Эконометрический ликбез: прогнозирование временных рядов», Квантиль, № 1, 2006 г., с 3-62.
12. Омельченко А.В., Роздымаха Е.А., Федоров А.В. Прогнозирование временных рядов со степенными регрессорами при коррелированном шуме наблюдения // Наукоемкие технологии в инфокоммуникациях: обработка и защита информации: коллективная монография / под ред. В.М. Безрука и В.В. Баранника. – Харьков: Компания СМІТ. -2013. – С. 262-272.
13. Беллами Дж. Цифровая телефония – М.: Эко-Тренз, 2004. – 640 с.  
Иващенко П.В. Основи теорії інформації: навч. посіб. / П.В. Иващенко – Одеса: ОНАЗ ім. О.С. Попова, 2015. – 53 с.
14. Сапожков М.А., Михайлов В.Г. Вокодерная связь. М.: Радио и связь, 1983. – 248 с.
16. Шелухин О. И., Лукьянцев Н. Ф.. Цифровая обработка и передача речи - М. : Радио и связь, 2000. – 454 с.
17. Тимченко О., Колодій Р. Порівняння методів оцінки якості кодеків мовного сигналу для VoIP // Радіоелектроніка та телекомунікації. Вісник НУ “Львівська політехніка”, №557. – Львів: 2006. – С.68-77
18. Фисенко В.Т.,Фисенко Т.Ю. Компьютерная обработка и распознавание изображений: учебное пособие. – СПб: СПбГУ ИТМО, 2008. – 192 с.

19. Р. Гонсалес, Р. Вудс. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
20. Смоленцев Н.К. Основы теории вейвлетов. Вейвлеты в MATLAB. – М., ДМК Пресс, 2005. – 303с.
21. Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. Методы сжатия данных. - М.: Диалог-МИФИ, 2003. - 384 с.
22. Столлингс В. Современные компьютерные сети. 2-е изд. – СПб: СПбГУ, 2003. – 783 с

## 11.2 Додаткова література

1. Кендалл М.Дж., Стьюарт А. Статистические выводы и связи. М.: Наука. 1973
2. Бикел П., Доксам К. Математическая статистика. Вып. 1, 2. М.: Финансы и статистика. 1983.
3. Тихонов А.Н., Уфимцев М.В. Статистическая обработка результатов экспериментов. М.: изд. МГУ. 1988.
4. Уфимцев М.В. Методы многомерного статистического анализа. М.: изд-во МГУ. 1997
5. Олійник А. О. Інтелектуальний аналіз даних : Навчальний посібник / А. О. Олійник, О. О. Олійник, С. О. Субботін. – Запоріжжя : ЗНТУ, 2011. – 278 с.
6. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посібник. – К.: КНЕУ, 2007. – 376 с.
7. Імовірності моделі випадкових сигналів та полів у прикладах і задачах. Прикладах і задачах: Навч. посібник / За ред. В.О. Омельченка. – К.: ІСДО, 1996. – 272 с.
8. Джеффри Расин. Непараметрическая эконометрика: вводный курс. Квантиль, №4, 2008. – С. 7-55.
9. У. Прэтт. Цифровая обработка изображений. - М.: Мир, 1982- 790 с.
10. Методы компьютерной обработки изображений / Под ред. В. А. Соифера – 2-е изд., испр. – М.: Физматлит, 2003. - 784 с.
11. Л. Шапиро, Дж. Стокман. Компьютерное зрение. Бином. Лаборатория знаний, М., 2006.
12. Б. Хорн. Зрение роботов. - М.: Мир, 1989.

## 12 ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

Система комп'ютерної математики Wolfram Mathematica  
Середовище програмування R

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

МЕТОДИЧНІ ВКАЗІВКИ

до самостійної роботи для підготовки  
докторів філософії  
за спеціальністю 172 «Телекомунікації та радіотехніка»  
з дисципліни

«МАТЕМАТИЧНІ МЕТОДИ АНАЛІЗУ ДАНИХ ТА ОБРОБКИ  
МУЛЬТИМЕДІЙНОЇ ІНФОРМАЦІЇ»

Харків 2020

## ЗМІСТ

<b>1 РОБОЧА ПРОГРАМА ДИСЦИПЛІНИ .....</b>	<b>4</b>
1.1 НОРМАТИВНІ ДАНІ З ДИСЦИПЛІНИ.....	4
1.2 МЕТА І ЗАВДАННЯ ДИСЦИПЛІНИ .....	5
1.2.1 Мета навчальної дисципліни.....	5
1.2.2 Завдання дисципліни.....	5
1.3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ .....	6
1.4 СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ.....	7
1.5 ЛАБОРАТОРНІ РОБОТИ.....	8
1.6 ПРАКТИЧНІ ЗАНЯТТЯ.....	8
1.7 САМОСТІЙНА РОБОТА .....	8
1.8 ІНДИВІДУАЛЬНІ ЗАВДАННЯ.....	8
1.9 МЕТОДИ НАВЧАННЯ .....	9
1.10 РЕЙТИНГОВА ОЦІНКА ЗА ДИСЦИПЛІНОЮ.....	9
1.10.1 Кількісні критерії оцінювання .....	9
1.10.2 Якісні критерії оцінювання .....	9
1.11 НАВЧАЛЬНО–МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ДИСЦИПЛІНИ ТА РЕКОМЕНДОВАНА ЛІТЕРАТУРА.....	12
1.11.1 Основна література:.....	12
1.11.2 Додаткова література .....	13
<b>2 ХАРАКТЕРИСТИКА ПІДРУЧНИКІВ І НАВЧАЛЬНИХ ПОСІБНИКІВ.....</b>	<b>14</b>
<b>3 МЕТОДИЧНІ ВКАЗІВКИ З ВИВЧЕННЯ ДИСЦИПЛІНИ .....</b>	<b>14</b>
3.1 Критерії перевірки статистичних гіпотез [1-3, 8].....	14
3.2 Методи класифікації даних [4 – 6] .....	15
3.3 Кластерний аналіз [4 – 6] .....	16
3.4 Кореляційний аналіз [4].....	17
3.5 Регресійний аналіз [1, 3, 7, 8] .....	18
3.6 Аналіз часових рядів та їх прогнозування [9 -12].....	19
3.7 Методи кодування мовних сигналів. Вокодерн [14 – 17].....	19
3.8 Подання зображень в цифровій формі та методи покращення зображень [18, 19] .....	20
3.9 Кодування цифрових зображень та відео послідовностей [18,19,21,22] .....	21
3.10 Вейвлет-аналіз і його використання при стисненні зображень [20] .....	22



## ВСТУП

При підготовці спеціалістів рівня доктора філософії самостійна робота є основною формою навчання. Така форма навчання дозволяє найкраще підготувати недавніх студентів до майбутньої наукової діяльності та дозволяє отримати стійкі знання.

Дисципліна «Математичні методи аналізу даних та обробки мультимедійної інформації» призначена для фахівців освітнього рівня доктор філософії за спеціальністю 172 «Телекомунікації та радіотехніка». Її метою є формування знань, умінь і навичок з методів і алгоритмів отримання знань із експериментальних даних та обробки мультимедійних даних.

# 1 РОБОЧА ПРОГРАМА ДИСЦИПЛІНИ

## 1.1 НОРМАТИВНІ ДАНІ З ДИСЦИПЛІНИ

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни			
		денна форма навчання	заочна форма навчання		
Кількість кредитів ЄКТС <u>8</u>	Галузь знань 17 «Електроніка та телекомунікації»	Вибіркова компонента ОП			
	Спеціальність підготовки 172 «Телекомунікації та радіотехніка»				
Модулів		Рік підготовки:			
Змістових модулів 2		1-й			
		Семестр			
Загальна кількість годин <u>240</u>		1-й	2-й		
	Кількість годин				
	120	120			
	Аудиторні: 1) лекції, год				
	2) практичні, год				
	-	-			
	3) лабораторні, год				
	-	-			
	4) консультації, год				
	20	20			
	Самостійна робота, год				
	100	100			
	в тому числі: 1) інд. завд., год.				
	-	-	-	-	
	2) курсова робота, год				
		-	-	-	
	Вид контролю: залік				

## 1.2 МЕТА І ЗАВДАННЯ ДИСЦИПЛІНИ

### 1.2.1 Мета навчальної дисципліни

Методи аналізу даних (Data mining) та обробки мультимедійної інформації – важливий напрямок в області інформаційних технологій, що активно розвивається. При дослідженні складних систем, якими є інформаційні мережі зв'язку, виникає необхідність в обробці та узагальненні великих обсягів різноманітної інформації. Для цього використовуються методи дискретної математики, регресійного аналізу, перевірки гіпотез, класифікації та кластеризації.

Метою дисципліни є формування у аспірантів знань, умінь і навичок з методів і алгоритмів отримання знань із експериментальних даних та обробки мультимедійних даних.

Дисципліна «Математичні методи аналізу даних та обробки мультимедійної інформації» відноситься до вибіркової дисципліни професійної та практичної підготовки за освітньо-професійною програмою Телекомунікації та радіотехніка за спеціальністю 172 Телекомунікації та радіотехніка.

Навчальна дисципліна використовується для формування наступних компетентностей: здатність до пошуку, систематичного вивчення та аналізу науково-технічної інформації, світового досвіду, пов'язаного із застосуванням методів телекомунікацій та радіотехніки для дослідження різноманітних процесів, явищ та систем.

### 1.2.2 Завдання дисципліни

Відповідно до мети викладення дисципліни головними завданнями викладення дисципліни є:

- засвоєння аспірантами математичних методів аналізу даних, достатніх для проведення наукових досліджень;
- ознайомлення аспірантів з технологіями аналізу даних на рівні, достатньому для формування уявлень, які б розширювали професійний кругозір;
- засвоєння аспірантами методів та алгоритмів обробки мультимедійних даних.

Програмні результати навчання: набуття знань та розуміння основних методів аналізу даних та вміння застосовувати інструменти та моделі аналізу даних (апаратно-програмні ресурси, пакети прикладних програм, онлайн ресурси й відповідні технології) в дослідженні реальних систем та презентації результатів наукових досліджень у різних формах; здійсненню науково-педагогічної діяльності з використанням цих ресурсів та технологій.

Після вивчення дисципліни аспіранти мають знати:

1. Критерії перевірки гіпотез Колмогорова, Смирнова,  $\chi^2$ -квадрат Пірсона.
2. Методи кореляційного аналізу.
3. Методи регресійного аналізу.

4. Методи локального ядерного аналізу.
5. Моделі процеси авторегресії і процеси ковзного середнього.
6. Методи прогнозування часових рядів.
7. Байєсівський підхід до побудови оптимальних вирішальних правил розпізнавання.
8. Алгоритми класифікації на кластеризації даних.
9. Методи кодування мовних сигналів.
10. Методи подання зображень в цифровій формі та методи покращення зображень.
11. Методи кодування цифрових зображень та відео послідовностей.
12. Неперервні вейвлет-перетворення SWT та дискретні вейвлет-перетворення DWT.

### 1.3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

#### Осінній семестр

##### **Змістовий модуль 1. Методи статистичного аналізу даних**

Тема 1. Критерії перевірки статистичних гіпотез

Тема 2. Методи класифікації даних

Тема 3. Кластерний аналіз

Тема 4. Кореляційний аналіз

Тема 5. Регресійний аналіз

#### Весняний семестр

##### **Змістовий модуль 2. Методи обробки мультимедійної інформації**

Тема 6. Аналіз часових рядів та їх прогнозування

Тема 7. Методи кодування мовних сигналів. Вокодери

Тема 8. Подання зображень в цифровій формі та методи покращення зображень

Тема 9. Кодування цифрових зображень та відеопослідовностей

Тема 10. Вейвлет-аналіз і його використання при стисненні зображень.

## 1.4 СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин											
	денна форма						Заочна форма					
	Усього го	у тому числі					Усього	у тому числі				
		л	п	лб	конс	с.р.		л	п	лб	конс	с.р.
<b>Змістовий модуль 1. Методи статистичного аналізу даних</b>												
<b>Тема 1.</b> Критерії перевірки статистичних гіпотез	16				2	14	16				2	14
<b>Тема 2.</b> Методи класифікації даних	30				6	24	30				6	24
<b>Тема 3.</b> Кластерний аналіз	28				4	24	28				4	24
<b>Тема 4.</b> Кореляційний аналіз	16				2	14	16				2	14
<b>Тема 5.</b> Регресійний аналіз	30				6	24	30				6	24
Разом за зміст. мод. 1	120				20	100	120				20	100
<b>Змістовий модуль 2. Методи обробки мультимедійної інформації</b>												
<b>Тема 1.</b> Аналіз часових рядів та їх прогнозування	24				4	20	24				4	20
<b>Тема 2.</b> Методи кодування мовних сигналів. Вокодери	24				4	20	24				4	20
<b>Тема 3.</b> Подання зображень в цифровій формі та методи покращення зображень	24				4	20	24				4	20
<b>Тема 4.</b> Кодування цифрових зображень та відеопослідовностей	24				4	20	24				4	20
<b>Тема 5.</b> Вейвлет-аналіз і його використання при стисненні зображень.	24				4	20	24				4	20
Разом за зміст. мод. 2	120				20	100					20	100
<b>Усього годин</b>	<b>240</b>	-	-	-	<b>40</b>	<b>200</b>	<b>240</b>	-	-	-	<b>40</b>	<b>200</b>

### 1.5 ЛАБОРАТОРНІ РОБОТИ

Лабораторні роботи за навчальним планом для даної дисципліни відсутні.

### 1.6 ПРАКТИЧНІ ЗАНЯТТЯ

Практичні заняття за навчальним планом для даної дисципліни відсутні.

### 1.7 САМОСТІЙНА РОБОТА

№ зміст модулю	Теми самостійної роботи	обсяг, год.	вид контролю	літ. джер.
Змістовий модуль 1. Методи статистичного аналізу даних				
1	Перевірка гіпотез	14	усне опитування	1-3, 7,8
1	Методи класифікації даних	24	усне опитування	4-6
1	Кластерний аналіз	24	усне опитування	4-6
1	Кореляційний аналіз	14	усне опитування	4
1	Регресійний аналіз	24	усне опитування	1, 3, 7, 8
Змістовий модуль 2. Методи обробки мультимедійної інформації				
2	Аналіз часових рядів та їх прогнозування	24	усне опитування	9, 10
2	Методи кодування мовних сигналів. Вокодер	24	усне опитування	14-17
2	Подання зображень в цифровій формі та методи покращення зображень	24	усне опитування	18, 19
2	Кодування цифрових зображень та відеопослідовностей	24	усне опитування	18, 19
2	Вейвлет-аналіз і його використання при стисненні зображень	24	усне опитування	20
<b>Загальна кількість</b>		<b>200</b>		

### 1.8 ІНДИВІДУАЛЬНІ ЗАВДАННЯ

За результатами навчання аспіранти готують реферат за узгодженою з викладачем темою, що пов'язана з темою їх наукових досліджень.

## 1.9 МЕТОДИ НАВЧАННЯ

Основні методи навчання:

- робота з навчально-методичною літературою (конспектування, самостійне опрацювання заданих розділів, написання реферату тощо).
- перевірка знань та умінь (за відповідями на контрольні запитання, результатами виконання контрольних робіт, індивідуальних завдань, доповіді за темою реферату).

## 1.10 РЕЙТИНГОВА ОЦІНКА ЗА ДИСЦИПЛІНОЮ

### 1.10.1 Кількісні критерії оцінювання

Для оцінювання роботи аспіранта протягом семестру підсумкова рейтингова оцінка  $Q_{\text{сем}}$  розраховується як сума оцінок усного опитування по кожній темі та оцінка за реферат. Кожна тема оцінюється в 5 балів. Оцінка за реферат – 50 балів. Максимальна рейтингова оцінка протягом семестру – 100 балів.

### 1.10.2 Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

1. Критерій узгодженості Колмогорова.
2. Критерій однорідності Смирнова.
3. Критерій  $\chi^2$ -квадрат Пірсона.
4. Модель регресійного аналізу.
4. Оцінювання коефіцієнтів регресії за методом найменших квадратів.
5. Локальний ядерний аналіз.
6. Фільтри Савицького-Голея.
7. Різницеві фільтри і їх застосування для інтерполяції даних.
8. Байєсівський підхід до побудови оптимальних вирішальних правил розпізнавання.
9. Вирішальні правила розпізнавання з використанням нормальних розподілів векторів ознак.
10. Алгоритми класифікації:  $k$  найближчих сусідів; наївний байєсівський класифікатор; класифікація з використанням дерев рішень.
11. Метрики класифікації і кластеризації.
12. Алгоритми кластерного аналізу:  $k$ -means; fuzzy  $k$ -means; EM; FOREL; з використанням штучних нейронних мереж.
13. Кореляційний аналіз з використанням коефіцієнтів кореляції Пірсона; Спірмена та Кендала.
14. Процеси авторегресії і процеси ковзного середнього.
15. Прогнозування часових рядів.
16. Умовне сподівання як оптимальний прогноз.

17. Лінійне прогнозування стаціонарного часового ряду.
18. Алгоритми ІКМ та АДІКМ
19. Класифікація вокодерних алгоритмів стиснення мовних сигналів.
20. Алгоритми стиснення мовних сигналів на основі моделі лінійного передбачення.
21. Стандарти стиснення даних JPEG та JPEG 2000.
22. Стандарти стиснення відеопослідовностей.
23. Неперервні вейвлет-перетворення SWT.
24. Дискретні вейвлет перетворення DWT.

Необхідний обсяг умінь для одержання позитивної оцінки.

1. Вміти застосовувати критерії перевірки гіпотез про функцію розподілу, однорідність і незалежність вибірок.
2. Вміти оцінювати коефіцієнти регресії за методом найменших квадратів.
3. Вміти знаходити коефіцієнти фільтрів Савицького-Голея та різницевих фільтрів.
5. Вміти обчислювати коефіцієнти моделі авторегресії.
6. Вміти прогнозувати стаціонарні часові ряди.
7. Вміти будувати вирішальні правила класифікації з використанням нормальних розподілів векторів ознак.
8. Вміти застосовувати алгоритми кластеризації даних.
9. Вміти вибирати параметри дискретизації і квантування мовних сигналів і зображень.
10. Вміти застосовувати алгоритми стиснення мовних сигналів.
11. Вміти застосовувати алгоритми стиснення зображень.
12. Вміти здійснювати вейвлет-перетворення Хаара і застосовувати його для стиснення зображень.

1.10.3 Критерії оцінювання роботи аспіранта протягом семестру.

**Задовільно, D, E (60-74).** Мати мінімум знань та вмінь. Опрацювати усі теми, виконати реферат.

**Добре, C (75-89).** Твердо знати матеріал. Опрацювати усі теми, виконати реферат. Знати методи аналізу даних і вміти застосовувати їх на практиці.

**Відмінно, A, B (90-100).** Твердо знати всі теми. Орієнтуватися в підручниках та учбових посібниках. Опрацювати усі теми, виконати реферат. Досконально знати методи аналізу даних і вміти застосовувати їх на практиці.



### Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
96 – 100	<b>A</b>	відмінно	зараховано
90-95	<b>B</b>		
75-89	<b>C</b>		
66-74	<b>D</b>		
60-65	<b>E</b>		
35-59	<b>FX</b>	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	<b>F</b>	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

## 1.11 НАВЧАЛЬНО–МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ДИСЦИПЛІНИ ТА РЕКОМЕНДОВАНА ЛІТЕРАТУРА

### 1.11.1 Основна література

1. Костин В.Н., Тишина Н.А. Статистические методы и модели: Учебное пособие. – Оренбург: ГОУ ОГУ, 2004. – 138 с.
2. Ивченко Г.И., Медведев Ю.И. Математическая статистика. М. 1984.
3. Лавренчик В.Н. Постановка физического эксперимента и статистическая обработка его результатов: Учеб.пособие для вузов. –М.: Энергоатомиздат, 1986.- 272 с.
4. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
5. Барсегян А. А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.
6. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М.: Мир, 1976.
7. Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. М.Финансы и статистика. -1987.
8. Себер Дж. Линейный регрессионный анализ. М.: Мир. 1980.
9. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. – М.:Мир, 1974. – Вып.1,2.
10. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. – М.: Финансы и статистика, 2003. – 414 с.
11. Циплаков А. «Эконометрический ликбез: прогнозирование временных рядов», Квантиль, № 1, 2006 г., с 3-62.
12. Омельченко А.В., Роздымаха Е.А., Федоров А.В. Прогнозирование временных рядов со степенными регрессорами при коррелированном шуме наблюдения // Научные технологии в инфокоммуникациях: обработка и защита информации: коллективная монография / под ред. В.М. Безрука и В.В. Баранника. – Харьков: Компания СМІТ. -2013. – С. 262-272.
13. Беллами Дж. Цифровая телефония – М.: Эко-Тренз, 2004. – 640 с.
14. Иващенко П.В. Основы теорії інформації: навч. посіб. / П.В. Иващенко – Одеса: ОНАЗ ім. О.С. Попова, 2015. – 53 с.
15. Сапожков М.А., Михайлов В.Г. Вокодерная связь. М.: Радио и связь, 1983. – 248 с.
16. Шелухин О. И., Лукьянцев Н. Ф.. Цифровая обработка и передача речи - М. : Радио и связь, 2000. – 454 с.
17. Тимченко О., Колодій Р. Порівняння методів оцінки якості кодеків мовного сигналу для VoIP // Радіоелектроніка та телекомунікації. Вісник НУ “Львівська політехніка”, №557. – Львів: 2006. – С.68-77
18. Фисенко В.Т., Фисенко Т.Ю. Компьютерная обработка и распознавание изображений: учебное пособие. – СПб: СПбГУ ИТМО, 2008. – 192 с.

19. Р. Гонсалес, Р. Вудс. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
20. Смоленцев Н.К. Основы теории вейвлетов. Вейвлеты в MATLAB. – М., ДМК Пресс, 2005. – 303с.
21. Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. Методы сжатия данных. - М.: Диалог-МИФИ, 2003. - 384 с.
22. Столлингс В. Современные компьютерные сети. 2-е изд. – СПб: СПбГУ, 2003. – 783 с

### 1.11.2 Додаткова література

1. Кендалл М.Дж., Стьюарт А. Статистические выводы и связи. М.: Наука. 1973
2. Бикел П., Доксам К. Математическая статистика. Вып. 1, 2. М.: Финансы и статистика. 1983.
3. Тихонов А.Н., Уфимцев М.В. Статистическая обработка результатов экспериментов. М.: изд. МГУ. 1988.
4. Уфимцев М.В. Методы многомерного статистического анализа. М.: изд-во МГУ. 1997
5. Олійник А. О. Інтелектуальний аналіз даних : Навчальний посібник / А. О. Олійник, О. О. Олійник, С. О. Субботін. – Запоріжжя : ЗНТУ, 2011. – 278 с.
6. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посібник. – К.: КНЕУ, 2007. – 376 с.
7. Імовірності моделі випадкових сигналів та полів у прикладах і задачах. Прикладах і задачах: Навч. посібник / За ред. В.О. Омельченка. – К.: ІСДО, 1996. – 272 с.
8. Джеффри Расин. Непараметрическая эконометрика: вводный курс. Квантиль, №4, 2008. – С. 7-55.
9. У. Прэтт. Цифровая обработка изображений. - М.: Мир, 1982- 790 с.
10. Методы компьютерной обработки изображений / Под ред. В. А. Сойфера – 2-е изд., испр. – М.: Физматлит, 2003. - 784 с.
11. Л. Шапиро, Дж. Стокман. Компьютерное зрение. Бином. Лаборатория знаний, М., 2006.
12. Б. Хорн. Зрение роботов. - М.: Мир, 1989.

## 1.12 ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

Система комп'ютерної математики Wolfram Mathematica  
Середовище програмування R

## 2 ХАРАКТЕРИСТИКА ПІДРУЧНИКІВ І НАВЧАЛЬНИХ ПОСІБНИКІВ

Питання перевірки статистичних гіпотез та регресійного аналізу розглядаються у літературі з математичної статистики. В обсязі, необхідному для вивчення дисципліни, вони викладаються в книгах [1-3]. Більш детально питання регресійного аналізу розглядаються в [7, 8].

Тісно пов'язані з регресійним аналізом питання кореляційного аналізу розглядаються в літературі по аналізу даних [4].

Методи класифікації і кластеризації викладаються у літературі з аналізу даних [4, 5]. Крім того, в дещо іншій термінології ці ж питання розглянуті в літературі з розпізнавання образів [6].

Питання по темі «Аналіз часових рядів та їх прогнозування» розглядаються в літературі з аналізу часових рядів [9]. Методам експоненційного згладжування присвячена книга [10]. Питання прогнозування часових рядів у стислій формі розглянуті у оглядовій статті [11] та [12].

Питання з кодування мовних сигналів розглядаються в літературі по цифровій телефонії [13], а також у книгах [14, 16]. Питання мовоутворення і вокодерного зв'язку викладаються в [15], а також в [16], де детально розглядаються сучасні вокодери лінійного передбачення. Методи оцінювання якості мови гарно викладені в оглядовій статті [17].

Методам обробки зображень присвячено багато книг. У необхідному для освоєння даної дисципліни питання з обробки зображень розглянуті в посібнику [18], а більш детально у книгах [19,21,22].

Для опрацювання теми «Вейвлет-аналіз і його використання при стисненні зображень» слід використати книги [19, 20].

У наступному розділі література, необхідна для вивчення заданої теми, вказана у квадратних дужках після назви теми.

## 3 МЕТОДИЧНІ ВКАЗІВКИ З ВИВЧЕННЯ ДИСЦИПЛІНИ

### 3.1 Критерії перевірки статистичних гіпотез [1-3, 8]

При вивченні даної теми необхідно опрацювати наступні питання.

Поняття статистичної гіпотези. Помилки 1-го та 2-го роду. Рівень статистичної значущості. Допустима та критична області. Статистичний критерій. Потужність критерію. Загальна задача перевірки гіпотез. Алгоритм вибору статистичного критерію. Перевірка гіпотез про розподіли та середні значення. Гіпотеза про дисперсію нормального розподілу. Критерії згоди омега-квадрат та хі-квадрат. Перевірка гіпотез про однорідність дисперсій: критерій Кохрена,

критерій Фішера. Непараметричні критерії: Крамера-Уелча, Уїлкоксона, критерій серій знаків.

Особливу увагу слід звернути на принципи побудови допустимої та критичної області прийняття гіпотез, а також на критерії перевірки гіпотез про розподіли, зокрема критерій  $\chi^2$ -квадрат.

### **Питання для самоконтролю**

1. Що називається критерієм згоди?
2. Що таке нульова гіпотеза?
3. Що таке альтернативна гіпотеза?
4. Що таке статистика критерію (тесту)?
5. З яких етапів складається алгоритм перевірки гіпотез?
6. Охарактеризуйте види помилок, що виникають при перевірці статистичних гіпотез.
7. Що таке рівень значущості?
8. Що таке критичні значення критерію?
9. Що називають потужністю критерію? Для чого використовують цей параметр?
10. Як можна підвищити потужність критерію?
11. У чому полягає різниця між критеріями згоди  $\omega^2$ -квадрат та  $\chi^2$ -квадрат?
12. Як перевірити гіпотезу про вид закону розподілу випадкової величини?
13. Яка мета перевірки гіпотез про рівність середніх значень та дисперсій?
14. Які критерії перевірки гіпотез про рівність генеральних середніх?
15. Які умови застосування критеріїв про рівність генеральних середніх?
16. Які критерії перевірки гіпотез про рівність генеральних дисперсій ви знаєте?
17. Які умови застосування критеріїв про рівність генеральних середніх?
18. Що називається рангом?
19. Які існують непараметричні критерії перевірки гіпотез?

## **3.2 Методи класифікації даних [4 – 6]**

При вивченні даної теми необхідно опрацювати наступні питання.

Гіпотеза компактності. Постановка задачі класифікації. Метод  $k$  найближчих сусідів. Міри близькості. 1-правило. Теорема Байєса і класифікація. Принцип максимуму апостеріорної імовірності. Байєсівський класифікатор. Наївний байєсівський класифікатор - Naive bayes. Оцінка ймовірностей в наївному байєсівському класифікаторі. Використання багатовимірного нормального розподілу в задачі розпізнавання. Лінійний дискримінант Фішера. Деревя рішень. Побудова дерев рішень. Критерії розщеплення. Алгоритми ID3 и C4.5. Алгоритм CART.

Особливу увагу слід звернути на використання теореми Байеса для класифікації даних, наївний байєсівський класифікатор і використання багатовимірного нормального розподілу в задачі розпізнавання об'єктів.

### *Контрольні питання*

1. Що називають класифікацією даних?
2. Який критерій найчастіше застосовують для розробки процедур класифікації даних?
3. Що називають функцією втрат? Для чого її використовують?
4. У чому полягає основна різниця між методами класифікації з навчанням та без навчання?
5. Якою є основна передумова застосування параметричних методів класифікації без навчання?
6. Як формують задачу побудови оптимальної класифікації у методах класифікації з навчанням?
7. Що називають вирішувальним правилом або дискримінантною функцією?
8. У якому випадку вирішувальне правило називають оптимальним?
9. У чому полягає сутність дискримінантного аналізу?
10. Якою є базова процедура параметричного дискримінантного аналізу у випадку нормальних класів?
11. Якою є загальна схема байєсівських методів класифікації з навчанням?
12. Якою є базова процедура лінійного дискримінантного аналізу Фішера?

### **3.3 Кластерний аналіз [4 – 6]**

При вивченні цієї теми слід опрацювати такі питання.

Цілі кластеризації. Постановка задачі кластеризації. Відмінність задач класифікації і кластеризації. Метрики кластеризації. Класифікація кластерних методів. Агломеративні і девізимні методи та їх дендрограми. Кластеризація методами теорії графів. Метод k-Means. Метод fuzzy c-means. Алгоритм EM:Expectation–Maximization. Алгоритм FOREL. Алгоритм мережі Кохонена. Критерії якості кластеризації. Особливості програмної реалізації в R. Метод головних компонент (PCA).

Особливу увагу слід звернути на метрики кластеризації, методи k-Means, fuzzy c-means, алгоритм EM.

### *Контрольні питання*

1. В чому полягає задача кластерного аналізу?
2. Для яких задач обробки експериментальних даних використовуються методи ієрархічного кластерного аналізу?
3. Наведіть основні міри порівняння об'єктів між собою.

4. Що таке дендрограма?
5. Що являють собою і ієрархічні агломеративні методи кластерного аналізу?
6. Що являють собою ієрархічні дивізімні методи кластерного аналізу?
7. Наведіть основні способи зв'язування об'єктів у кластери.
8. В чому полягають основні етапи ієрархічного кластерного аналізу?
9. Яким чином визначити значущу кількість кластерів?
10. Геометрична інтерпретація кластерного аналізу?
11. Що називають класом в кластерному аналізі?
12. Які міри відстані найчастіше використовують у кластерному аналізі?
13. Які вимоги має задовольняти функція, яку використовують як міру відстані?
14. Що називають мірою Махаланобіса? Як вона пов'язана з іншими мірами відстані?
15. Як визначається евклідова відстань?
16. Які міри близькості кластерів один до одного використовують у кластерному аналізі?
17. У чому полягає сутність ієрархічних методів кластерного аналізу?
18. Якими є основні переваги й недоліки ієрархічних методів кластерного аналізу?
19. Яким є базовий алгоритм методу ближнього зв'язку в кластерному аналізі?
20. Що називають дендрограмою? Як можна побудувати дендрограму?
21. Яким є базовий алгоритм методу  $k$  – середніх Мак-Куїна у кластерному аналізі?

### **3.4 Кореляційний аналіз [4]**

При вивченні даної теми необхідно опрацювати такі питання.

Багатовимірна вибірка. Функціональний та стохастичний зв'язок між величинами. Перевірка гіпотези про кореляційний зв'язок. Поле кореляції. Емпіричний коефіцієнт лінійної кореляції Пірсона. Довірчий інтервал для коефіцієнта кореляції. Кореляційне відношення. Множинна кореляція. Коефіцієнти рангової кореляції Спірмена та Кендела.

Порівняння однорідності декількох груп. Застосування критерію Кохрена. Порівняння однорідності двох груп. Застосування критерію Фішера. Застосування критерію Стьюдента. Порівняння часток. Таблиці спряженості. Застосування критерію  $\chi^2$ -квадрат.

Особливу увагу слід звернути на оцінку коефіцієнта лінійної кореляції Пірсона та довірчі інтервали для неї.

#### **Питання для самоконтролю**

1. Дайте означення поняттю «кореляція»
2. Наведіть приклади додатної та від'ємної кореляції

3. Що називається кореляційним аналізом? Яка мета кореляційного аналізу?
4. Що називається кореляційним зв'язком? Статистичним зв'язком?
5. Що називається полем кореляції? Кореляційною таблицею?
6. Як побудувати кореляційну таблицю?
7. Як за вибірковими даними визначити вид зв'язку?
8. Чим відрізняються коефіцієнти кореляції Пирсона та Спірмена?
9. Які спільні риси мають коефіцієнти кореляції Пирсона і Спірмена?
10. Який зв'язок між факторами вважається сильним? Середнім? Слабким?
11. Чи показує коефіцієнт кореляції спрямованість зв'язку?
12. Який висновок робиться, якщо коефіцієнт кореляції є додатнім?

Від'ємним?

13. Що означає поняття «статистична значущість»?
14. Як перевірити зв'язок між декількома факторами?
15. Що таке множинний кореляційний зв'язок?
16. Які коефіцієнти кореляції обчислюються при множинному кореляційному аналізі?

### 3.5 Регресійний аналіз [1, 3, 7, 8]

При вивченні даної теми необхідно опрацювати такі питання.

Задача регресійного аналізу. Криві регресії. Лінійна регресія. Наближена лінія регресії. Властивості коефіцієнтів рівняння регресії та їх геометрична інтерпретація. Метод найменших квадратів. Метод максимальної правдоподібності. Довірчий інтервал для коефіцієнта регресії. Перевірка статистичної значущості коефіцієнтів регресії. Перевірка адекватності відтвореної статистичної моделі. Коефіцієнт детермінації  $R^2$ . Множинна (багатовимірна) лінійна регресія. Нелінійна парна регресія. Поліноміальна регресія. Побудова нелінійних емпіричних залежностей з використанням ортогональних поліномів Чебишева. Ортогоналізація рівняння регресії.

Особливу увагу слід звернути на застосування методу найменших квадратів для оцінювання коефіцієнтів регресії, знаходження коефіцієнтів поліноміальної регресії, та також на зміст коефіцієнта детермінації  $R^2$ .

#### Питання для самоконтролю

1. Дайте означення поняттю «регресія»
2. Як побудувати емпіричну лінію регресії? Теоретичну лінію регресії?
3. За яким методом знаходяться коефіцієнти рівняння регресії?
4. У чому полягає особливість лінійної регресії?
5. Як перевірити правильність побудованої регресійної моделі?
6. Що називається коефіцієнтом детермінації і як він використовується?
7. Як перевірити статистичну значущість побудованої регресійної моделі?
8. Що називається багатofакторною лінійною регресією?
9. Як здійснюється прогноз за багатofакторною лінійною регресійною моделлю?



10. Як знайти критичне значення критерія?
11. Як з декількох регресійних моделей обрати найбільш адекватну?

### **3.6 Аналіз часових рядів та їх прогнозування [9 -12]**

При вивченні даної теми необхідно опрацювати такі питання.

Метод лінійного передбачення. Процеси авторегресії-ковзного середнього. Процеси авторегресії. Передатні функції ARMA-систем. Оцінювання коефіцієнтів передбачення. Коваріційний і автокореляційний метод. Рівняння Юла-Уокера. Процеси ARIMA. Параметричні методи спектрального оцінювання.

Оптимальне у середнькваратичному сенсі прогнозування. Умовне середнє як оптимальний прогноз. Експоненційне згладжування і прогноз методом експоненційного згладжування. Лінійне прогнозування стаціонарних часових рядів. Прогнозування за моделлю Бокса–Дженкіса. Прогнозування стаціонарного процесу ARMA. Прогнозування процесу ARIMA.

Особливу увагу слід звернути на моделі процесів авторегресії, авторегресії-ковзного середнього і їх узагальнення у вигляді процесу авторегресії про інтегрованого ковзного середнього ARIMA. Слід дакож звернути увагу на умовне середнє як найкращий у середньквдратичному сенсі лінійний прогноз.

#### **Питання для самоконтролю**

1. У чому полягає суть моделі лінійного передбачення?
2. Дайте означення процесу авторегресії-ковзного середнього
3. Дайте означення процесу авторегресії
4. Дайте означення процесу ARIMA.
5. Які існують методи оцінювання параметрів процесів авто регресії?
6. У чому різниця коваріційного і автокореляційног методу оцінування параметрів процесу авторегресії?
7. Які переваги мають параметричні методи спектрального оцінювання у порівнянні з непараметричними?
8. Який алгоритм побудови експоненційного згладжування?
9. Як здійснюється прогноз за методом експоненційного згладжування?
10. Яким чином константа згладжування впливає на прогноз?
11. Як розраховується помилка прогнозу?
12. Як обґрунтовується оптимальність моделі для прогнозування?
13. Яким чином прогнозуються значення процесу авторегресії?
14. Яким чином прогнозуються значення процесу ARIMA?

### **3.7 Методи кодування мовних сигналів. Вокодерн [14 – 17]**

При вивченні даної теми необхідно опрацювати такі питання.

Модель процесу мовоутворення. Поняття формант і фонем. Формування мовного сигналу в кодеку G.711. Адаптивна диференційна ІКМ.

Принцип побудови вокодерів. Типи вокодерів. Канальні вокодери. Форматні вокодери. Різновиди фокодерів лінійного передбачення. Вокодери зізбудженням від сигналу основного тону. RPE-LTP вокодер стандарту GSM. Багатоімпульсне кодування вокодери MPLPC. Векторне квантування і вокодери CELP. Оцінювання якості мови. Середня експертна оцінка – MOS. Модель E і розрахунок R-фактору.

Особливу увагу слід звернути на методи кодування мовних сигналів, що побудовані на основі моделі лінійного передбачення.

### **Питання для самоконтролю**

1. Назвіть основні параметри мовних сингалів і поясніть їх фізичну суть.
2. З яких основних етапів складається процес формування мови?
3. У чому полягає різниця між формуванням вокалізованих і невокалізованих фонем?
4. Як визначаються інтервал дискретизації та частота дискретизації?
5. Що таке крок квантування та як він вибирається?
6. Яким чином можна збільшити відношення сигнал-шум квантування?
7. У чому полягають переваги кодування з компандуванням?
8. Як визначається швидкість цифрового сигналу при кодуванні за методом ІКМ?
9. Пояснити принцип дії цифрових систем із лінійним передбаченням.
10. Пояснити принцип кодування за методом ДІКМ.
11. Назвіть основні операції формування сигналу ДІКМ.
12. Назвіть основні принципи функціонування смугового вокодеру.
13. Назвіть основні принципи функціонування форматного вокодеру.
14. З яких етапів складається обробка сигналів в RPE-LTP вокодері стандарту GSM?
15. З яких етапів складається обробка сигналів в MPLPC вокодері?
16. Що таке векторне квантування?
17. У чому полягає особливість роботи вокодеру CELP?
18. Яким чином знаходиться оцінка MOS?
19. Які фактори враховує E-модель?
20. Яким чином розраховується оцінковий фактор R?

### **3.8 Подання зображень в цифровій формі та методи покращення зображень [18, 19]**

При вивченні даної теми необхідно опрацювати такі питання.

Дискретизація зображень. Вибір частоти дискретизації у відповідність з теоремою відліків. Оптимальне квантування. Квантування при наявності завад. Похибка дискретизації. Методи інтерполяції.

Гістограми зображень. Яскравісні перетворення зображень. Лінійне контрастування. Нормалізація і еквалізація гістограм. Перетворення зображень в негатив. Лінійна просторова фільтрація. Локальна фільтрація. Частотна

фільтрація зображень. Порогова і медіанна фільтрація. Виділення контурів шляхом фільтрації. Різницеві фільтри.

Особливу увагу слід звернути на поняття просторової частоти, вибором частоти дискретизації, яскравіших перетвореннях зображень та виділення контурів шляхом фільтрації.

### **Питання для самоконтролю**

1. Поясніть загальну схему обробки зображень.
2. Що таке піксел зображення?
3. Як кодується зображення?
4. Які моделі подання кольорів ви знаєте?
5. Що показує гістограма яскравостей пікселів зображення?
6. Для яких цілей застосовується лінійне контрастування? Наведіть алгоритм лінійного контрастування.
7. Що розуміють під контрастом і яскравістю під час цифрової обробки зображень? Яка міра контрасту?
8. Що таке локальна просторова фільтрація? У чому принципова відмінність між локальною та глобальною фільтрацією?
9. У чому полягає суть двомірного перетворення Фур'є?
10. Охарактеризуйте рекурсивні і нерекурсивні фільтри.
11. Що таке лінійна рекурсивна фільтрація зображень?
12. Охарактеризуйте лінійні фільтри для виділення і підкреслення контурів.
13. Поясніть роботу алгоритму медіанної фільтрації. Для чого застосовується медіанна фільтрація
12. Що буде з зображенням, якщо з його спектру вилучити складові з високими просторовими частотами ?

### **3.9 Кодування цифрових зображень та відео послідовностей [18,19,21,22]**

При вивченні даної теми необхідно опрацювати такі питання.

Особливості зорової системи людини. Стиснення без втрат. Стиснення із втратами. Кодування довжин серій. Кодування по методу LZW. Арифметичне стискання. Метод кодування Хафмана. Метод диференціальної ІКМ. Кодування з використанням ортогональних перетворень. Дискретне косинусне перетворення. Стиснення даних по методу JPEG.

Стандарти стиснення відео послідовностей: MPEG-1, MPEG-2, MPEG-4.

Особливу увагу слід звернути на стиснення даних по методу JPEG і стандарти стиснення відео послідовностей.

### **Питання для самоконтролю**

1. В чому переваги і недоліки методів стиснення зображень з втратами?
2. Яке граничне стиснення забезпечує алгоритм Хафмана ? Пояснити.

3. Чи може LZW-алгоритм стиснення використовуватись не для зображень?
4. Назвіть основні етапи, з яких складається JPEG-алгоритм стиснення цифрових відеозображень.
5. Який результат застосування ДКП до цифрового відеозображення?
6. Як обчислюється пряме і обернене ДКП?
7. Назвіть переваги і недоліки різних методів стиснення цифрових відеозображень
8. Як впливає стиснення відеозображень на їх якість?
9. На яких етапах кодування MPEG відбуваються невідновні втрати інформації?
10. В чому полягає різниця кодування I- та P-,B-кадрів?
11. За рахунок чого у передбачених кадрах зменшуються похибки ДКП та квантування опорних кадрів?
12. Пояснити необхідність буферного ЗП у кодері.
13. В чому полягає операція компенсації руху?
14. Як забезпечується часова синхронізація стандарту MPEG?

### **3.10 Вейвлет-аналіз і його використання при стисненні зображень [20]**

При вивченні даної теми необхідно опрацювати такі питання. Вейвлет-перетворення. Вейвлетні функції. Поняття вейвлету. Материнські вейвлети і скейлін-функції. Неперервні вейвлет-перетворення SWT. Дискретні вейвлет-перетворення DWT. Перетворення Хаара. Узагальнене перетворення Хаара. Використання двохмірного перетворення Хаара для стиснення зображень. Приклад двохмірного перетворення Хаара. Багатократні перетворення Хаара. Вейвлет-стиснення зображень. Стандарт JPEG 2000.

Особливу увагу слід звернути на перетворення Хаара і його застосування для стиснення зображень, а також на застосування вейвлет-стиснення зображень у стандарти JPEG 2000.

#### **Питання для самоконтролю**

1. Що таке материнські і дочерні вейвлети?
2. Які властивості мають материнські вейвлети?
3. Наведіть приклади материнської вейвлетів.
4. Що таке скейлін-функції?
5. Яким чином задається неперервне вейвлет-перетворення SWT?
6. Яким чином задається дискретне вейвлет-перетворення DWT?
7. У чому полягає основна ідея вейвлет аналізу?
8. Яким чином будується перетворення Хаара?
9. У чому полягає суть узагальненого перетворення Хаара?
10. Наведіть приклад перетворення Хаара.
11. Наведіть приклад двохвимірного перетворення Хаара.
12. Яким чином будується багатократне вейвлет-перетворення ?

*13. У чому суть вейвлет-стиснення зображень?*

*14. З яких етапів складається обробка зображень при кодуванні за стандартом JPEG 2000.*