# Початок

Доброго дня пані та панове. Мене звуть Микита Красницький, а тема мого дипломного проекта звучить як – «Методи класифікації нот в акустичному сигналі»

# Актуальність

На сьогодні теми машинного навчання та штучного інтелекту тримаються на вершині тренду. Вони э дуже поширеними в великій кількості застосувань. В загалом в машинному навчанні з вчителем існують регресивні моделі, що видають числа/функції/передбачення трендів та класифікатори, що видають на виході один клас з кінечної вибірки.

Так, класифікатор зображення може визначити що за об’єкти на ньому зображені. Аудио-асистент може перевести голосовий потік в текст, після чого може виконати задані команди. Аналізатор коментарів може визначити за текстом чи є коментар позитивним чи негативним.

Отже існує велика кількість додатків, що використовують алгоритми класифікації. Моя дипломна робота розбирає принцип роботи та математичні основи популярних алгоритмів класифікації - а саме Гаусівського наївного Байесівського класифікатора та класифікатора SVM. В якості практичної задачі для використання класифікаторів була обрана задача визначення нот в цифровому акустичному сигналі. На прикладі даної задачі розбирається питання вибору та препроцесінгу вектору ознак, що виступають вхідними данними для моделі. Однією з тем що охоплюються дипломною роботою є оптимізація та прискорення даних алгоритмів класифікацій.

Наряді з розбором популярних класифікаторів в роботі я також ілюструю деякі емпіричні методи класифікації. В кінці роботи показано, що комбінація емпіричного методу та Гаусівсько го Байесівського класифікатора дає найкращий результат.

(1min 50s)

# Задача

Задача класифікації звучить так. Нехай є аудиозапис музичної композиції, де, для спростування, звучить не бульше однієї ноти в один момент часу. Треба визначити послідовність нот що грали в композиції.

На рисунку 1.2 зліва можна побачити часове зображення композиції. Зправа зображен графік нот, що детално буде описан пізніше. Якщо коротко – числа зліва це лінійний індекс ноти, що включає як його октаву, так і саму ноту.

(2min 25s)

# Теорія нот

Перед оглядом можливих рішень треба розглянути теорію нот та фізичні процеси що супроводжують звучання ноти. Це дасть більш широке представлення про можливості аналізу цифрового звукового сигналу для визначення послідовності нот.

**ЧАСТОТИ**

Стисло про теорію нот. Нота характеризується частотою вібрації основного тону. На таблиці 1.1 можна побачити відповідність кожної ноти до частоти. Основних нот всього 7 – До, Ре, Мі, Фа, Соль, Ля, Сі. Але, як видно з таблиці, існує ще 5 додаткових нот що не мають власної назви, а називаються через сусідні. Так До дієз – нота вища на 0.5 тона від До та Ре бімоль – нота нижча на 0.5 тона від Ре є одна й та сама нота. Отже в октаві є 12 нот. Всього октав 9, починаючи з Суббконтр-октави та закінчуючи пьятою октавою. Для даного дослідження буду брати тільки 7 октав, починаючи з контр-октави та закінчуючи четвертою, так як іншні використовуються дуже рідко.

//////////////Насьогодні стандартним ладом є рівнотемперований лад. Згідно цього стандарту сусідні ноти октави відносяться один до одного як . Тоді Нота через октаву, тобто нота через 12 нот, буде відноситись як 1:2. Тобто зі зростанням октави частота ноти зростає як показникова функція з основою 2. Такий тренд можна побачити на рисунку 1.3, де кожна окрема лінія відображає частоти окремої ноти./////////////////

(4min 20s)

**ТРИВАЛІСТЬ**

Окрім частоти нота характеризується тривалістю та іноді тембром. Тривалість, в свою чергу, визначається двома показниками – bpm(beats per minute) композиції та тривалістю ноти. На рисунку 1.4 можна побачити найбільш популярні тривалості нот, зліва зображена ціла нота, після неї половинна, четвертна, а в кінці зображена одна шістдесят четверта нота. Показник bpm визначає скільки четвертних нот пролунає за хвилину. Стандартним значенням є 120, максимальним 216. Для стандартного значення bpm одна четверта нота звучить 500мс. При найбільшому значенні bpm, найменш тривала нота, а саме 1/64, буде звучати приблизно 35мс.

(5min 6s)

**ТЕМБР**

Ноти повністю визначаються частотами, але звучання ноти однієї і тієї ж частоти на різних інструментах звучить з різним забарвленням. Це називається тембром. Аби музичний інструмент видавав звук ноти, його певна частина повинна задовольняти фізичних умов до механічних коливань на частоті ноти. У струнних інструментів такими елементами є струни. У духових музичних інструментів використовуються труби.

Розглянемо поняття тембру більш докладно на прикладі струнних інструментів. Струна є довгий відрізок гнучкого матеріалу, що знаходиться в натягнутому стані, завдяки чому може вільно коливатися. Перш за все треба відмітити, що у струни є 2 вузла. На гітарі ці вузли розташовані на нижньому і верхньому поріжку відповідно, що визначають довжину струни.

Таким чином при збудженні струни вона буде коливатися всіма механічними хвилями, вузли яких лягають в вузли на поріжках. На рисунку 1.5 справа показані хвилі, що можуть коливатися в гітарній струні. Тоді самий низькочастотний звук, що видає збуджена струна, виходить у хвилі з найбільшою довжиною хвилі. Такий тон називають основним або базовим. Всі інші хвилі є кратними по частоті базовому тону і називаються гармоніками або обертонами.

Розподіл амплітуд між гармоніками та характеристика їх затухання у часі визначають тембр музичного інструменту. На рисунку 1.6 можна побачити часові залежності перших 6ти гармонік гітарної струни.

**МИКРО-ВИСНОВКИ**

Отже були розглянуті характеристики ноти та їх визначення. Була визначена мінімальна тривалість ноти. Були розглянуті фізичні особливості звучання музичних інструментів.

Переходимо до класифікаторів.

(7min 7s)

# еМПІРИЧНИЙ МЕТОД

Емпіричним шляхом була створена матриця М розмірністю 7х12, що відповідає 7ми октавам на 12 нот. В кажну ячейку матриці було покладене абсолютне значення спектру вирізаного сигналу для заданої ноти. Далі, шляхом множення на матрицю V була створена матриця Ms, що містить у собі сумми октав нот помножені на відповідні усереднювальні коефіцієнти так, що сумуються тільки кратні ноти до основного тону. Таким чином, якщо основний тон не мав найбільшої амплітуди через особливість тембру та розподілу амплітуд, то після знаходження матриці Ms саме на цій ячейці буде найбільше значення, так як будуть сумуватись всі кратні октави.

Такий метод дає результат з помилкою 32%, що видно з рисунка 1.15. А застосування матриці Ms в якості вхідного вектору ознак для популярних методів класифікації дають менші відсотки помилки та потребують менших затрат, так як кількість вхідних даних значно зменшується.

# Гаусівський наївний БК

Гаусівський наївний Байесів класифікатор, надалі БК, заснований на формулі Теореми Байеса для знаходження ймовірності, що при належності об’єкту до класу він має вектор ознак рівний . Застосовуючи дану формулу для одного об’єкту перебираючи всі класи отримаємо ймовірностний розподіл пренадлежності об’єкта до кожного класу. Клас обирається по максимуму вірогідності. Отже цікавить максимізація формули Байеса.

В даній роботі не було зроблено дослідження про частоту зустрічання тих чи інших нот в композиціях, тому складова апріорної ймовірності зустрічі певного набору ознак є константою і формула зводиться до пошуку максимуму чисельника. Наївний Байес стверджує, що вірогідність належності вектору ознак х до класу у рівний добутку вірогідностей що кожен елемент вектора ознак х належить до класу у. Оскільки вектор ознак може містити багато елементів, то щоб запобігти багаторазовому множенню малих чисел вирішили логарифмувати формулу. Максимум при логарифмуванні не зміниться, а числа не будуть занадто малими.

Гаусівський класифікатор стверджує, що залежності даних вірогідностей під сумою підчиняються Гаусівському розподілу. Отже для кожного елементу вектора ознак буде побудовано свый гаусівський розподіл. Приймаючи новий вектор ознак буде перевірено чи лежить кожен з елементів в заданих гаусівським розподілом межах.

Стосовно задачі. Оскільки в аудиозаписі ноти змінювалися в часі – було обрано ширину вікна, що вирізає частину сигналу, де точно буде одна нота. /////Оскільки мінімальна нота з стандартною частотою дискретизації має 1650 відліків, було обрано вікно шириною в 1024 семпли, що відповідає найбільшій степені двійки, що буде менша за сигнал. Тоді в вікні точно буде одна нота. Хоча вікно може зачепити сусідні ноти на моменті переходу однієї ноти в іншу. /////

Були опробувані такі вхідні вектори ознак як: відліки часового сигналу, відліки спектру сигналу, інтерпольовані відліки спектру, відліки спектру, що відповідають частотам нот, матриця сум октав перемножена на відповідні коефіцієнти. Останній варіант показав найкращий результат в 3% помилки.

В якості перевірки класифікатора був обран аудиозапис, в якому кожна нота піаніно з обраних 7ми октав звучить рівно 2 секунди. Сигнал був поділений на шматки шириною в 1024 відліки. 70% даних вирізаних сигналів використовувалися в якості навчальної вибірки, залишні 30% - в якості тестувальної. Номера по осі ординат на рисунках 1.8 та 1.9 означають номер ноти починаючи з контр-октави.

# SVM

Метод support vector machine (SVM) вирішує задачу бінарної класифікації шляхом побудови гіперповерхні, що розділяє об’єкти на два напівпростору N-го виміру. Тобто об’єкти, що описуються вектором ознак, виступають точками в N вимірному просторі, а метод SVM будує таку гіперповерхню, щоб відстань від поверхні до найближчого представника свого классу була максимальною.

Очевидно, що не всі дані можуть бути розділені гіперплощиною. В випадках, коли дані відокремлюються нелінійно використовують ядрові функції, що збільшують вимірність та змінюють положення точок так, що б їх було простіше відокремили лінійною гіперплощино. На рисунку 1.11 показан приклад використання поліномінальної однорідної ядрової функції, що збільшила вимірність простору даних та дала змогу розділити дані гіперплощиною, що в переході до початкової вимірносте створить нелінійну відокремлюючу криву.

В випадку багатокласової класифікації метод SVM використовують двома способами. Перша стратегія називається «один-проти-всіх», для класів використовують послідовних бінарних класифікаторів які визначають чи дані належать до певного класу, чи не належать. \\\\\\ Припустимо, що існує 4 класи, нехай вони називаються А, Б, С і Д. Тоді будуються 4 бінарні SVM класифікатори: «A» проти «не А», «Б» проти «не Б» і так далі. Обирається додатній клас, що має найбільшу відсталь до гіперплощини. \\\\\\\\

Стратегія «один-проти-одного» заключається в тому, що будуються пари класифікаторів з всіма можливими парами класів. Такий метод вимагає побудову значно більшої кількості класифікаторів.

Метод SVM зарекомендував себе як найкращий з розглянутих методів для вирішення даної задачі. Як видно з рисунків 1.12 та 1.13, відсоток помилки становить 1.1% для класифікації за спектром та 1.8% для класифікації за матрицею Ms. Окрім протилежного ефекту після заміни вхідного вектора ознак на матрицю Ms в порівнянні з БК слід зауважити, що даний класифікатор потребує й іншого підходу до нормування вхідних даних. Тоді як БК потребував нормування за сумою вектора, SVM потребує нормування за максимумом.

# Висновки

Отже, найбільш простий та емпірично знайдений метод посідає на останьомі місці по точності серед трійки класифікаторів з 32% помилки. Такий метод не потребує нормалізації, так як максимум після нормалізації не змінює свого положення.

БК посідає на 2гому місці з 3% помилки використовуючи матрицю Ms як вхідний вектор ознак. Вхідні дані нормувалися за суммою, щоб сума всіх елементів вектору була рівна одиниці. Використовуючи спектр замість матриці Ms, класифікатор витрачав найбільше часу з трійки.

Класифікатор SVM є лідером з точності та має 1.1% помилки класифікації використовуючи спектр. Даний класифікатор потребував нормалізації даних за максимумом, так як він є чутливим до масштабування. Даний класифікатор виявився оптимальним вибором методу класифікації з розглянутої трійки.