

Машинное обучение и классификация (РСА) в спектроскопии

Парфенов П.С.

Международный научно-образовательный центр «Физика наноструктур»

Университет ИТМО

январь 2024 г.

План обсуждения:

- Единичный нейрон (<https://github.com/qrspeter/perceptron>)
 - Математическая модель, функция активации
 - Обучение на одномерных массивах (от 2x1 до спектра КТ PbS двух типов)
- Метод главных компонент
 - Спектры молока с разной концентрацией лактозы, спектры КТ PbS двух типов (https://github.com/qrspeter/pca_pbs)
 - Анализ спектров CD и “продвинутые” варианты PCA для неполных данных – EMPCA и WPCA (https://github.com/qrspeter/pca_cd_chirality)
- Анализ чувствительности одномерного газового сенсора (https://github.com/qrspeter/ml_alcohol)
 - Распознавание спиртов одним сенсором
 - Распознавание парой сенсоров.

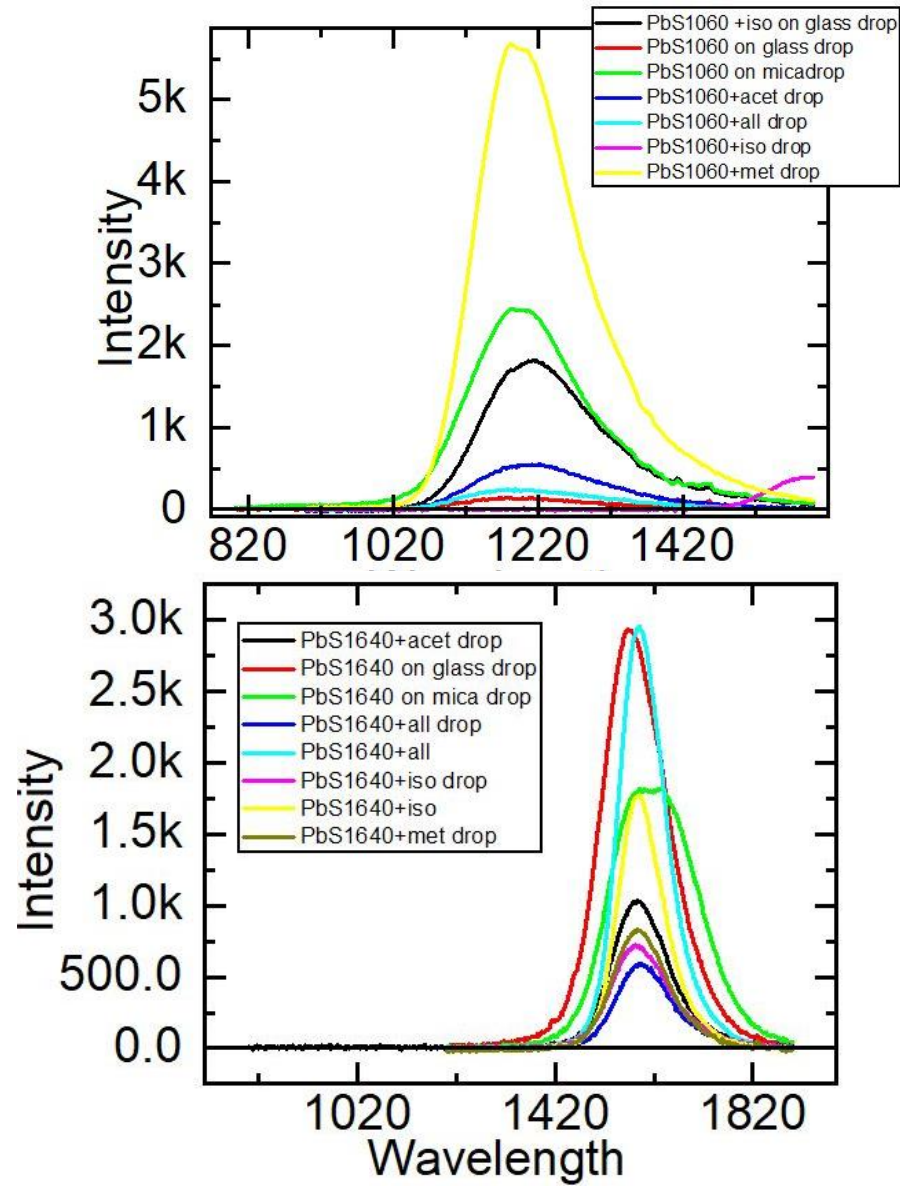
Простейший случай

- Одиночный нейрон (перцептрон)

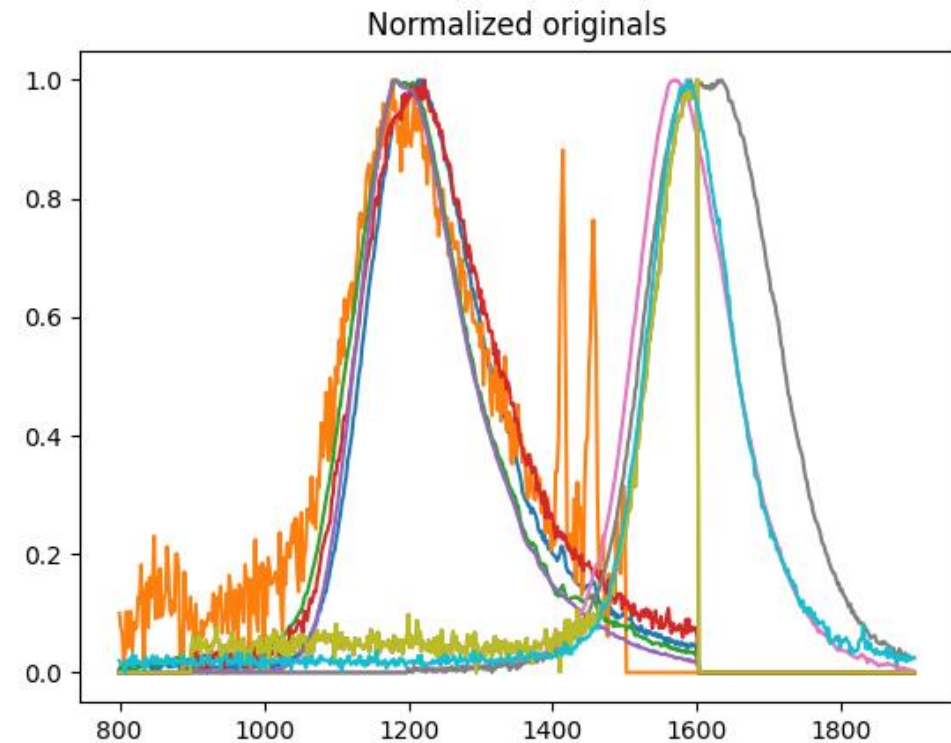
<https://github.com/qrspeter/perceptron>

(см. ч.1 - презентацию для бакалавров)

Пример распознавания: КТ PbS1060 + PbS1640



1060:	1060:	1640:
900.0000	800.0000	1200.0000
902.0000	802.0000	1202.0000
...
1600.0000	1600.0000	1900.0000



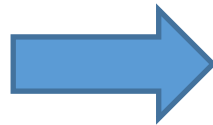
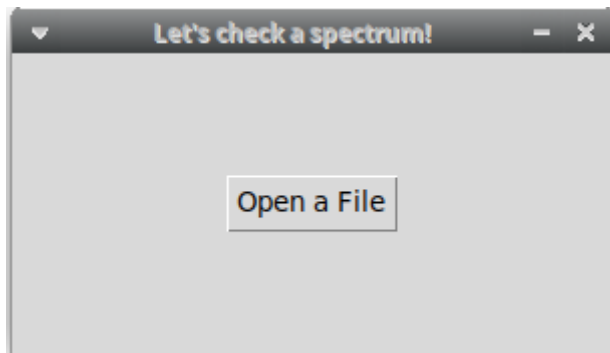
Домашнее задание у бакалавров

- 0 – PbS1060
- 1 – PbS1640
- Порог - 0,5

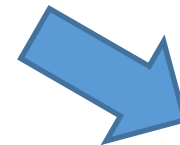
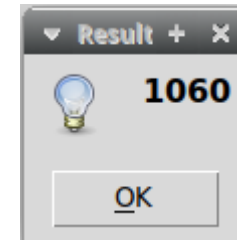
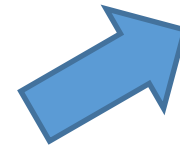
```
training_set_outputs = array([[0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1]]).T
```

```
inputs = 551
```

```
trainings = 3 #10000
```



```
if result < 0.5:  
    mess = '1060'  
else:  
    mess = '1640'
```

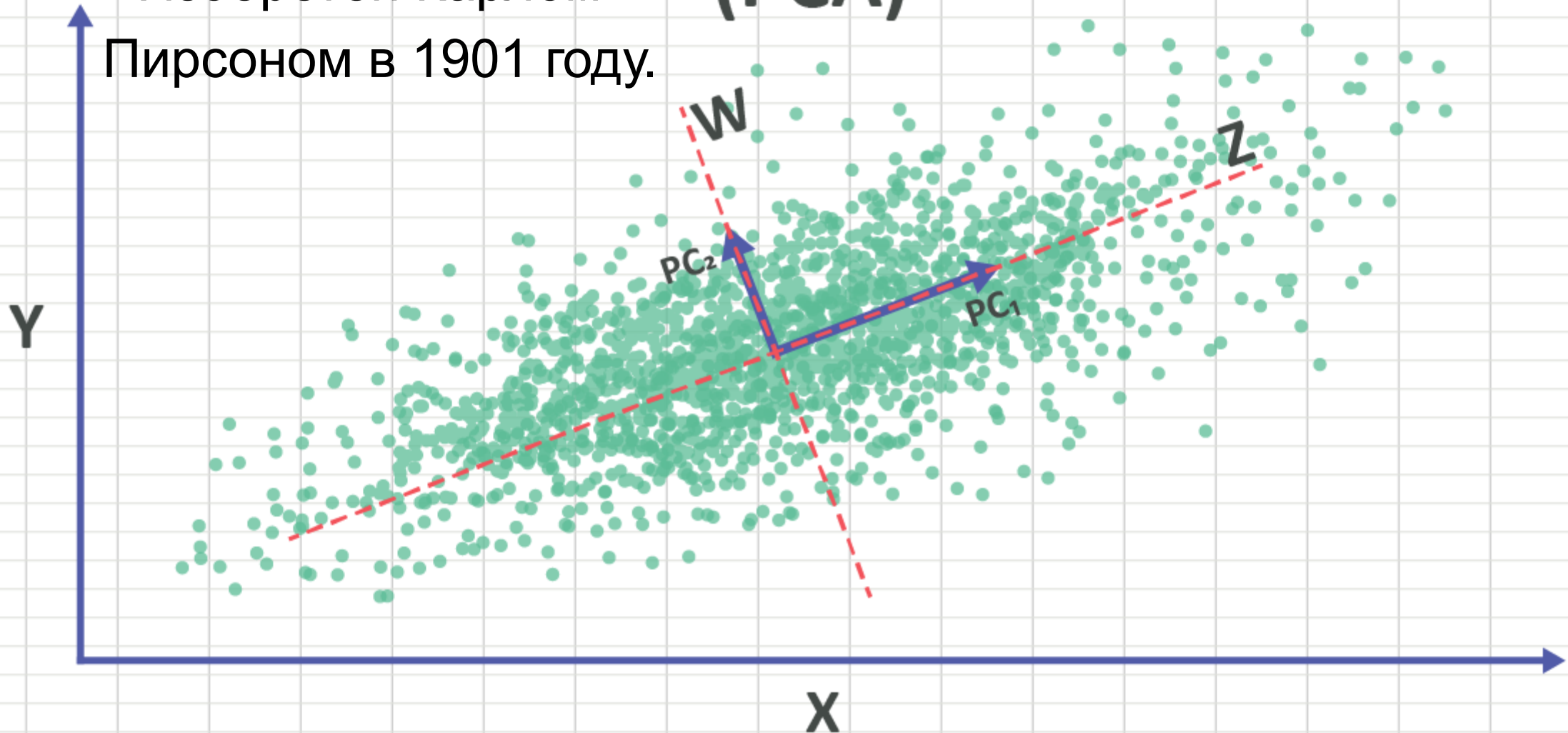


От нейрона-перцептрона к PCA

- https://github.com/qrspeter/pca_pbs

Principal Component Analysis (PCA)

- Изобретён Карлом Пирсоном в 1901 году.



Применение

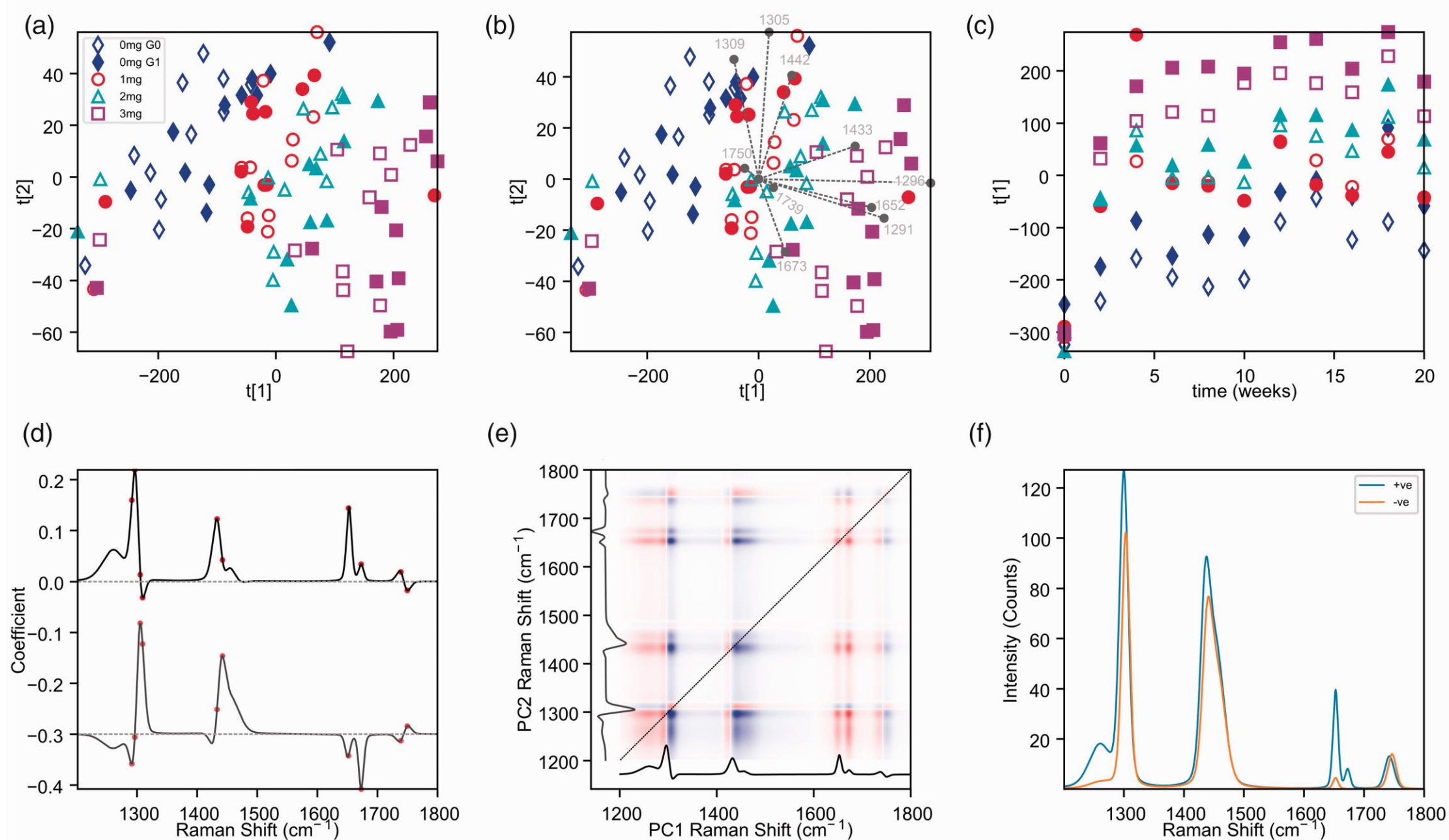
- Уменьшение шумов
- Уменьшение размерности данных для обработки и распознавания
- выявление спектральных полос (спектральный анализ, полосовые детекторы)
- улучшение представления данных
- классификация данных с массива “одномерных” датчиков

(+): не требует большого объема данных

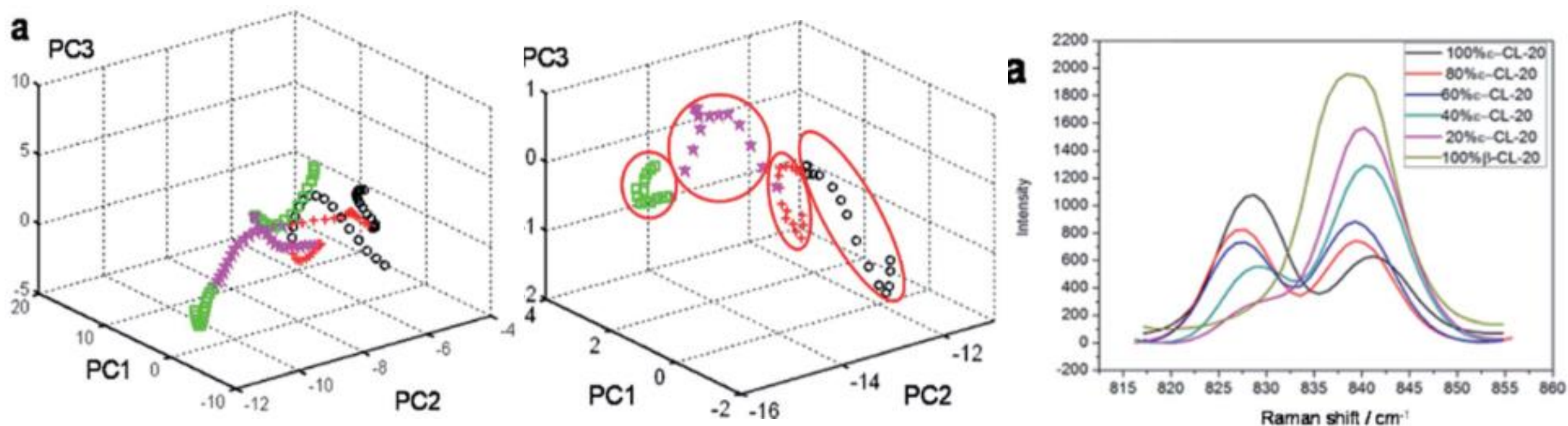
(—): не годится для сложных случаев

Рекомендуемая литература

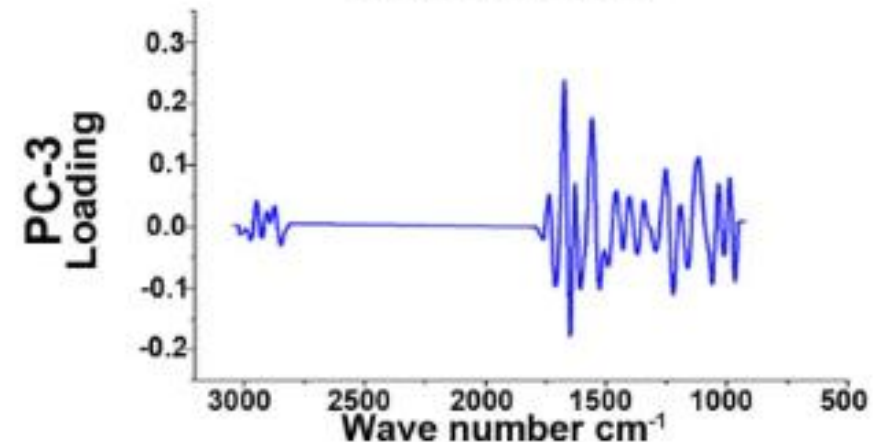
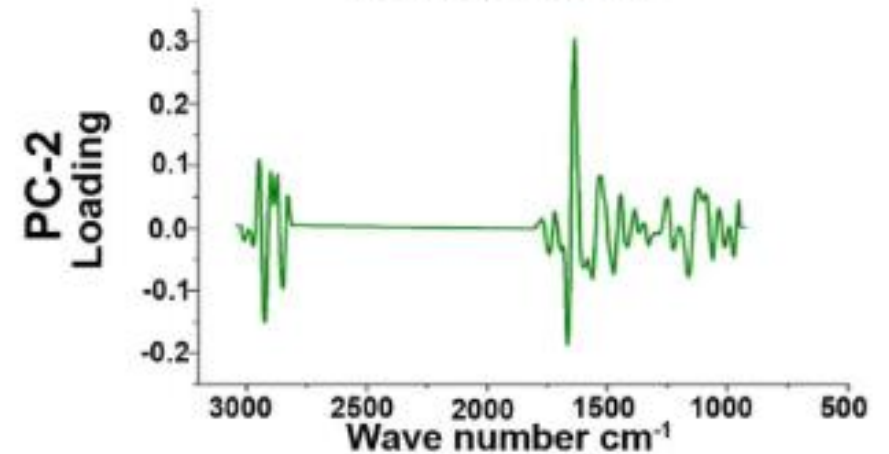
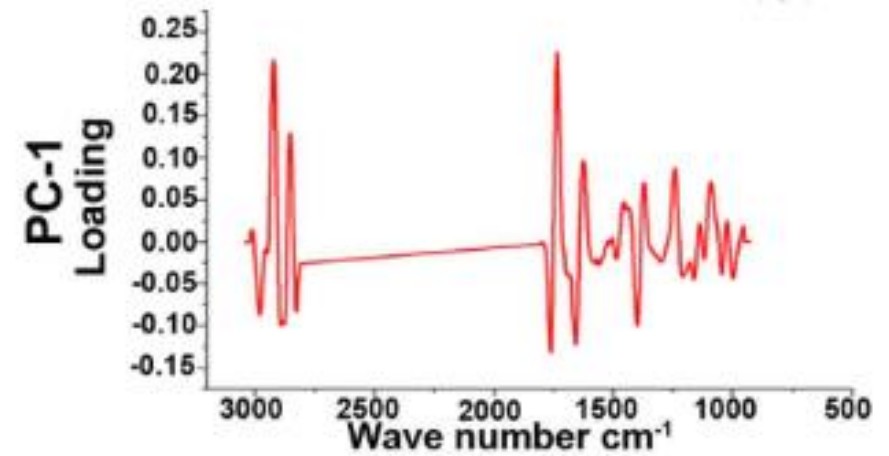
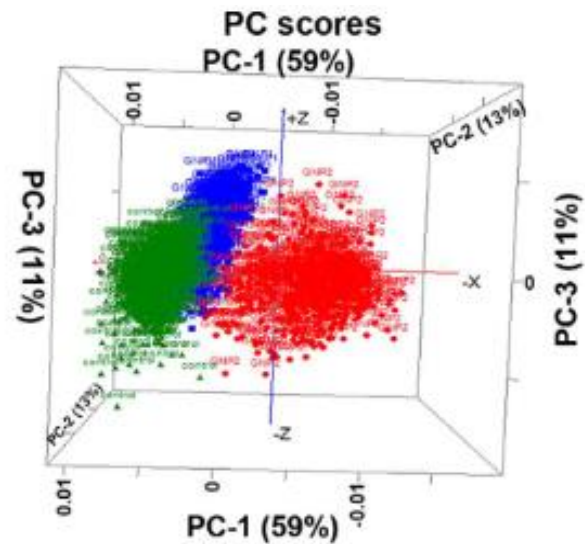
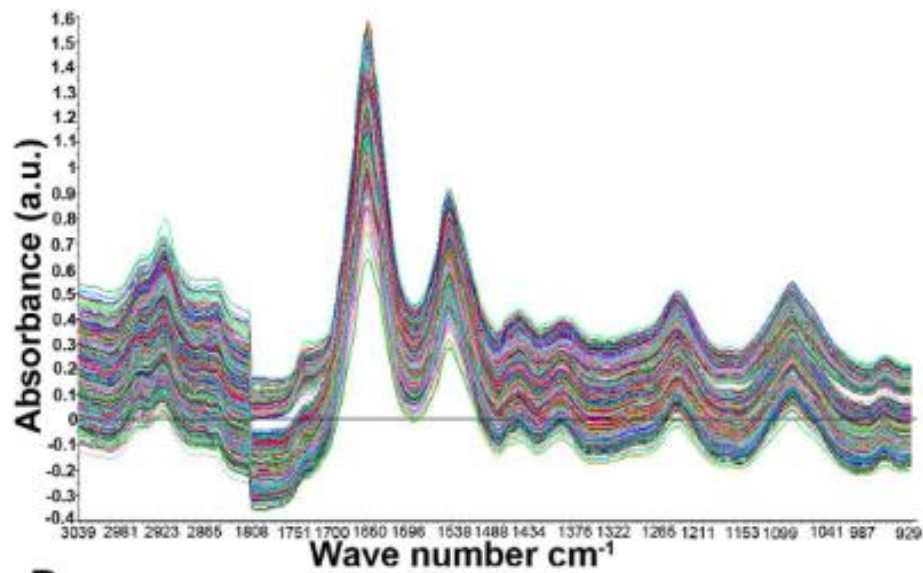
- Sun, Y., Brockhauser, S., & Hegedűs, P. (2021). Machine Learning Applied for Spectra Classification. In Computational Science and Its Applications – ICCSA 2021 (pp. 54–68). Springer International Publishing. https://doi.org/10.1007/978-3-030-87013-3_5
- <https://www.originlab.com/doc/App/PCA-Spec>
- Beattie, J. R., & Esmonde-White, F. W. L. (2021). Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra. Applied Spectroscopy, 75(4), 361–375. doi:10.1177/0003702820987847
- <https://nirpyresearch.com/classification-nir-spectra-principal-component-analysis-python/>
- <https://www.geeksforgeeks.org/principal-component-analysis-with-python/>



Beattie JR, Esmonde-White FWL. Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra. *Applied Spectroscopy*. 2021;75(4):361-375. doi:[10.1177/0003702820987847](https://doi.org/10.1177/0003702820987847)

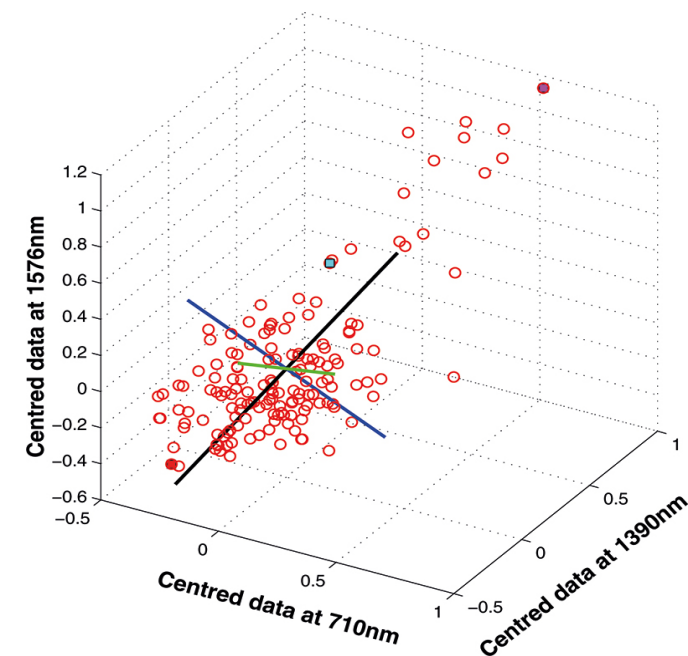
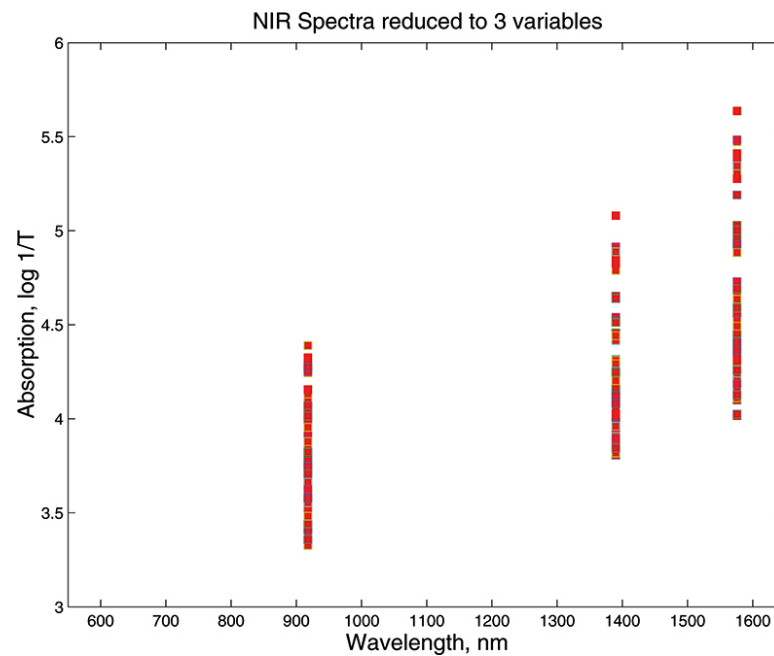
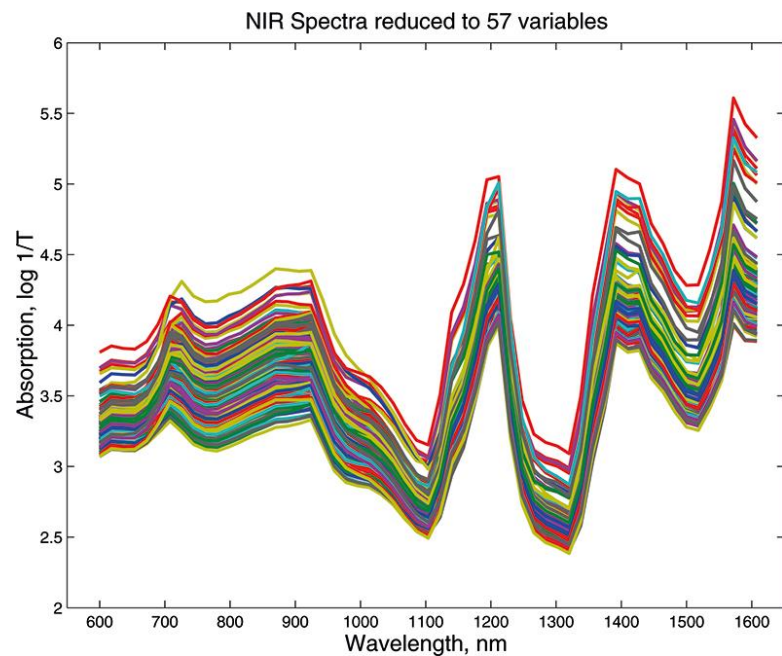


He, X., Liu, Y., Huang, S., Liu, Y., Pu, X., & Xu, T. (2018). Raman spectroscopy coupled with principal component analysis to quantitatively analyze four crystallographic phases of explosive CL-20. In RSC Advances (Vol. 8, Issue 41, pp. 23348–23352). Royal Society of Chemistry (RSC). <https://doi.org/10.1039/c8ra02189a>



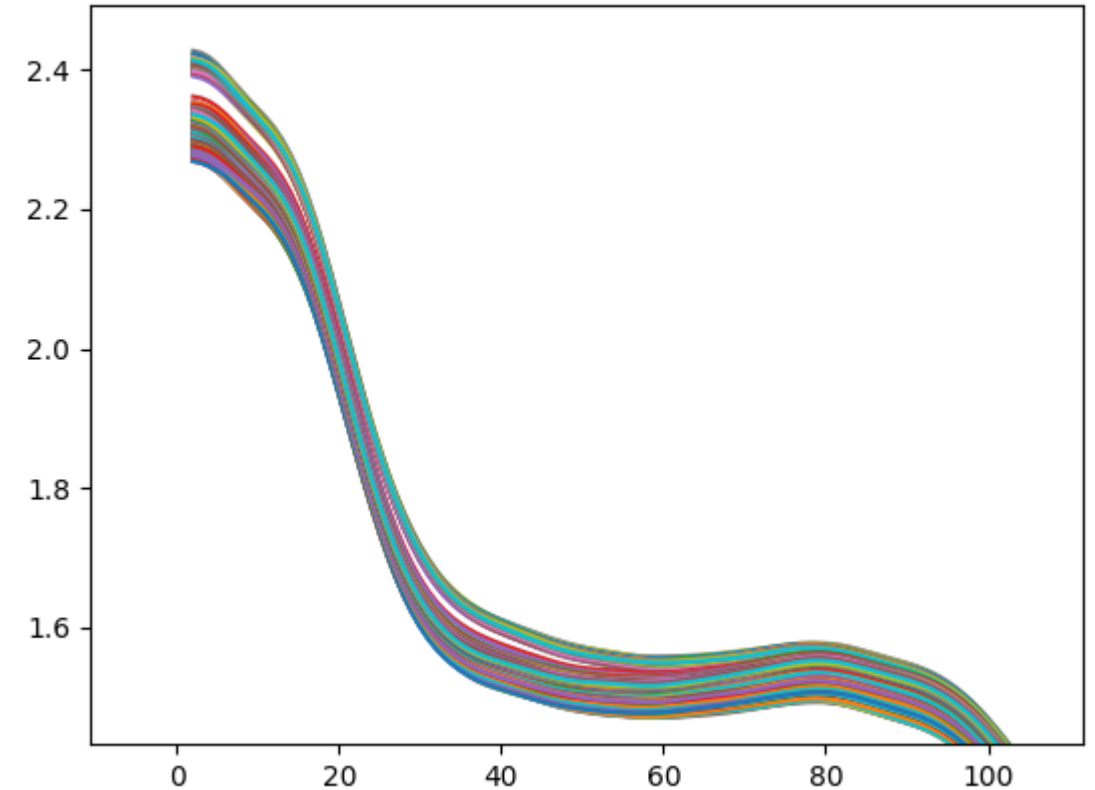
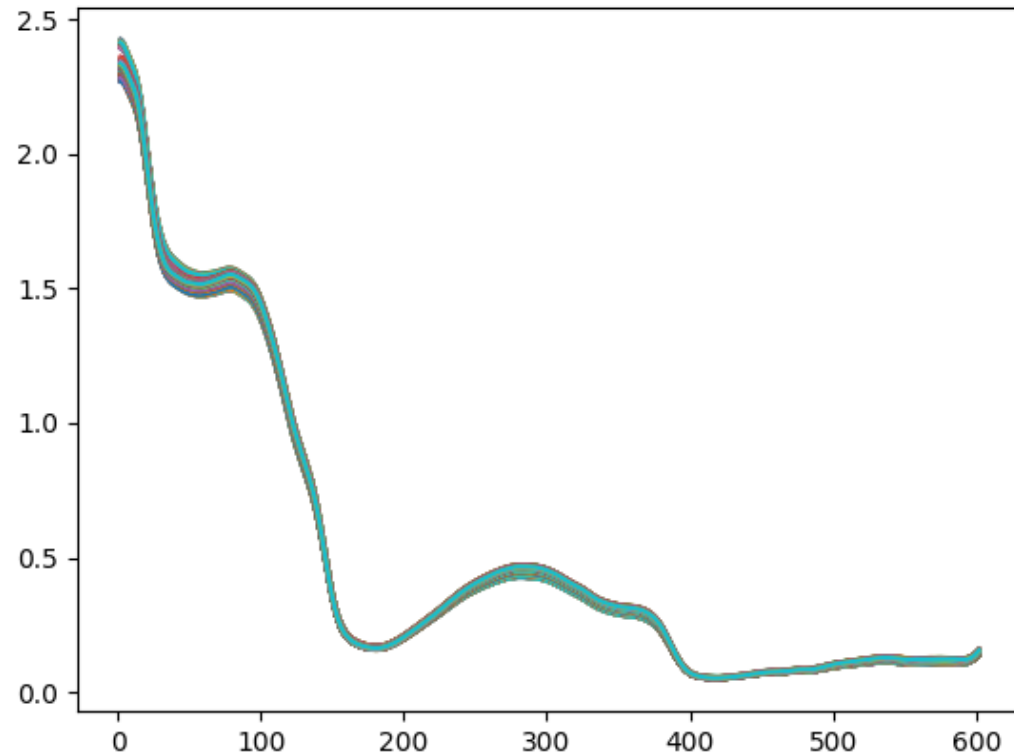
Kalmodia, S., Parameswaran, S., Yang, W., Barrow, C. J., & Krishnakumar, S. (2015). Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy: An analytical technique to understand therapeutic responses at the molecular level. In *Scientific Reports* (Vol. 5, Issue 1). Springer Science and Business Media LLC.

<https://doi.org/10.1038/srep16649>



<https://www.spectroscopyeurope.com/td-column/back-basics-principles-principal-component-analysis>

Спектр как точка в многомерном пространстве



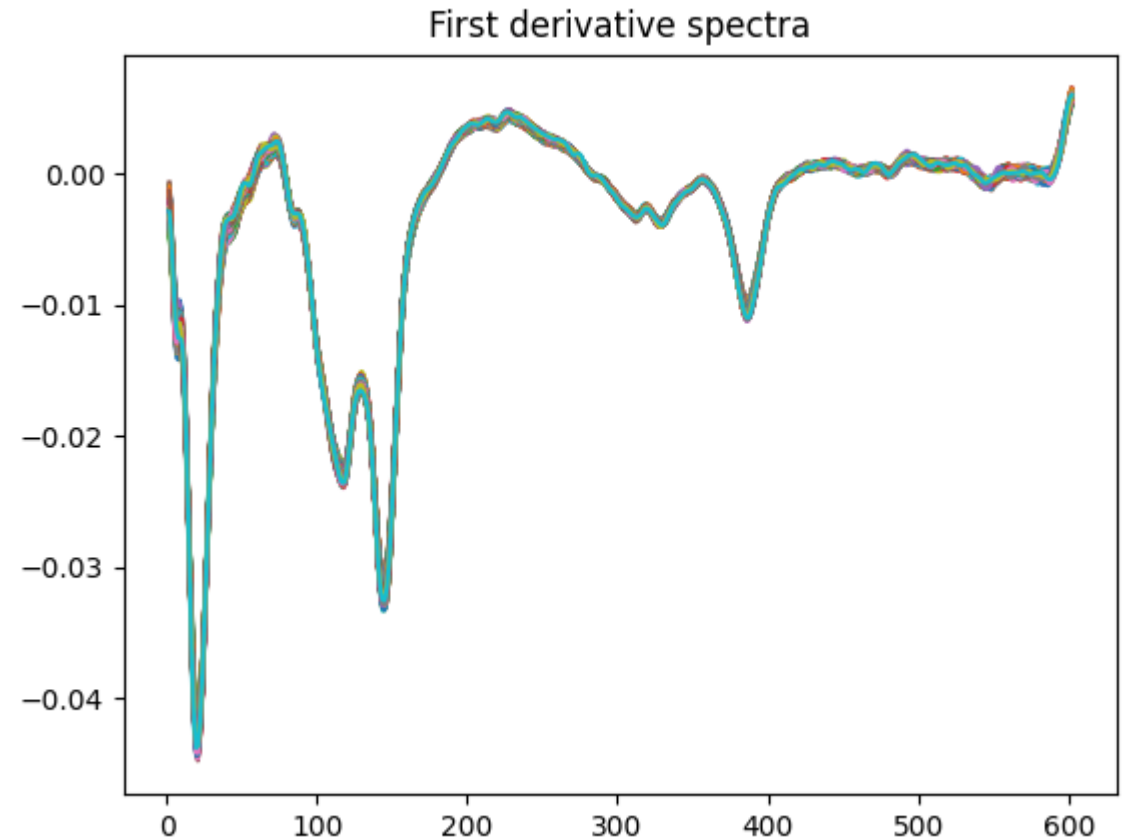
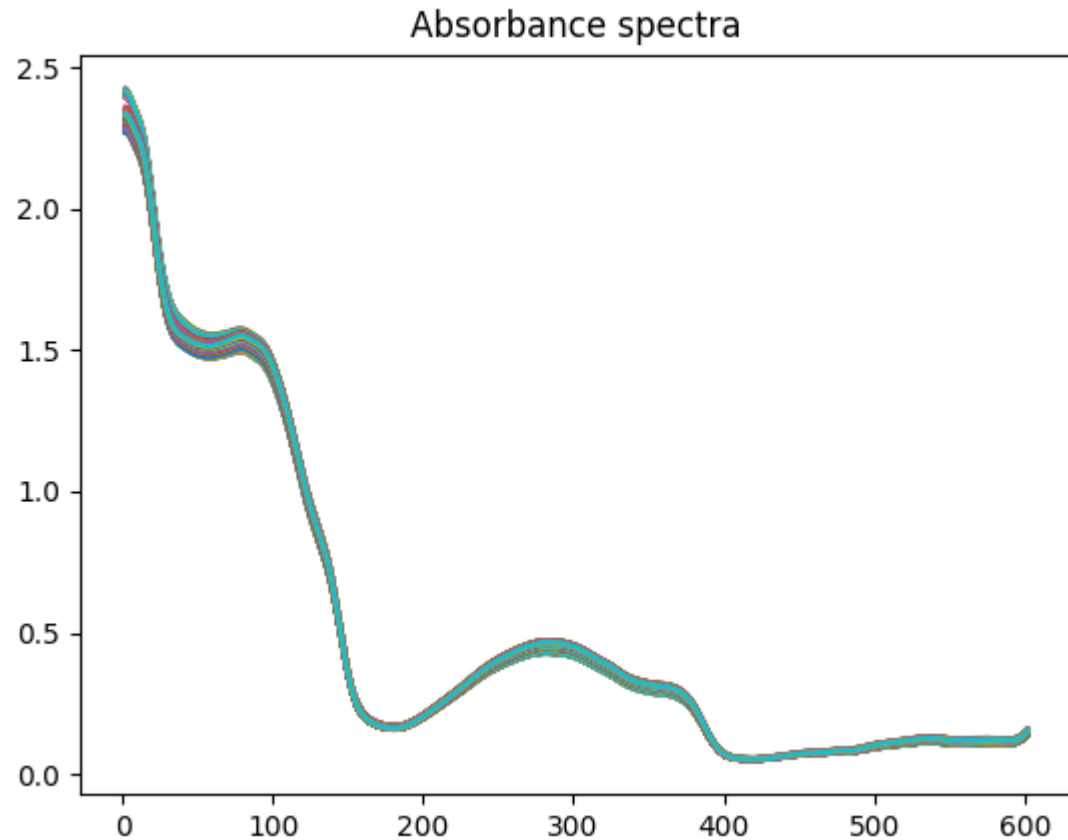
NIR spectra from milk samples with a varying concentration of lactose.

<https://nirpyresearch.com/classification-nir-spectra-principal-component-analysis-python/>

https://github.com/qrspeter/pca_pbs/blob/main/pca_milk.ipynb

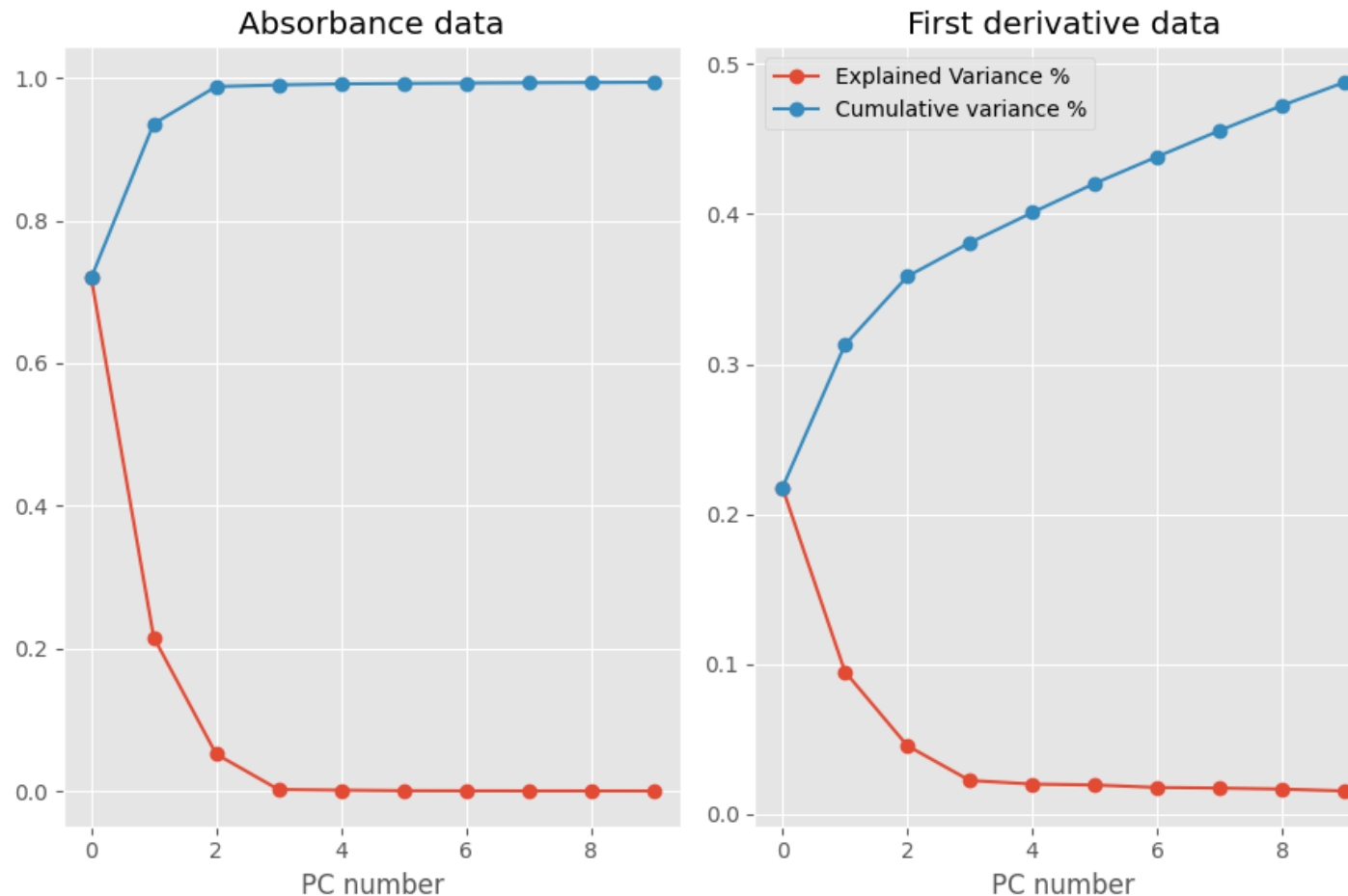
Классификация спектров

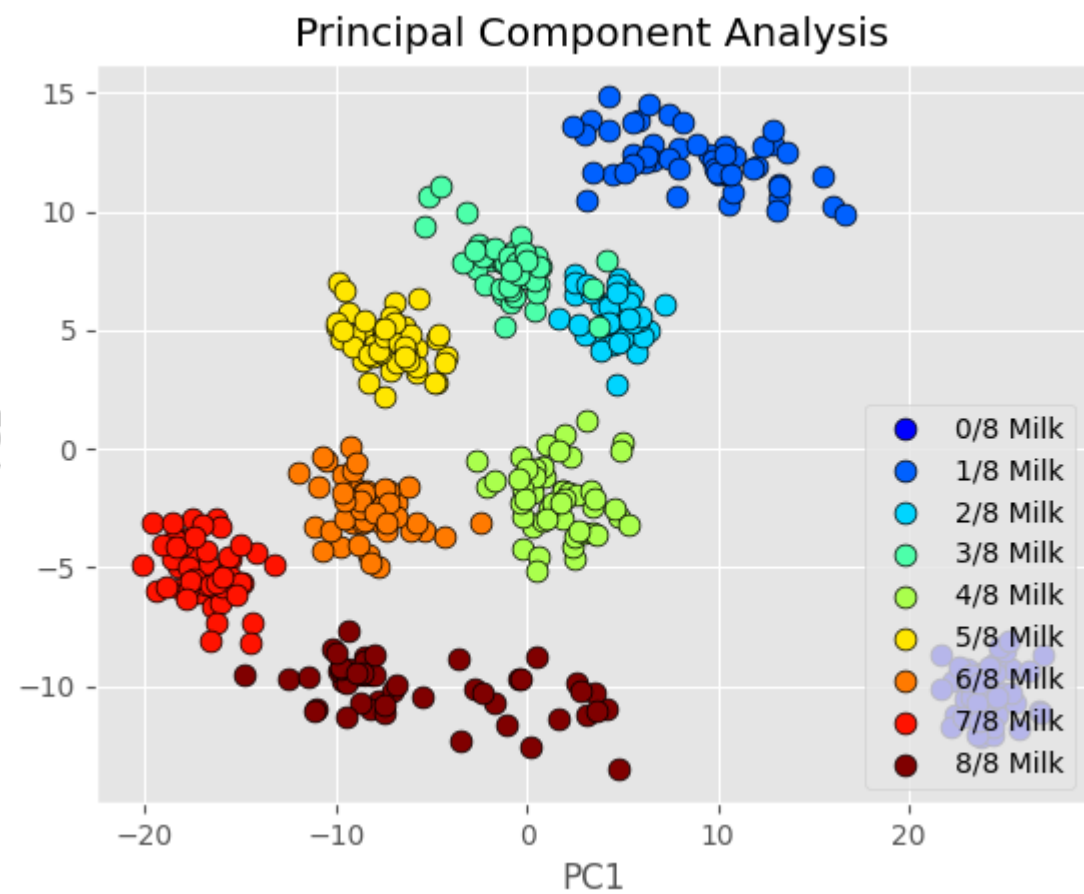
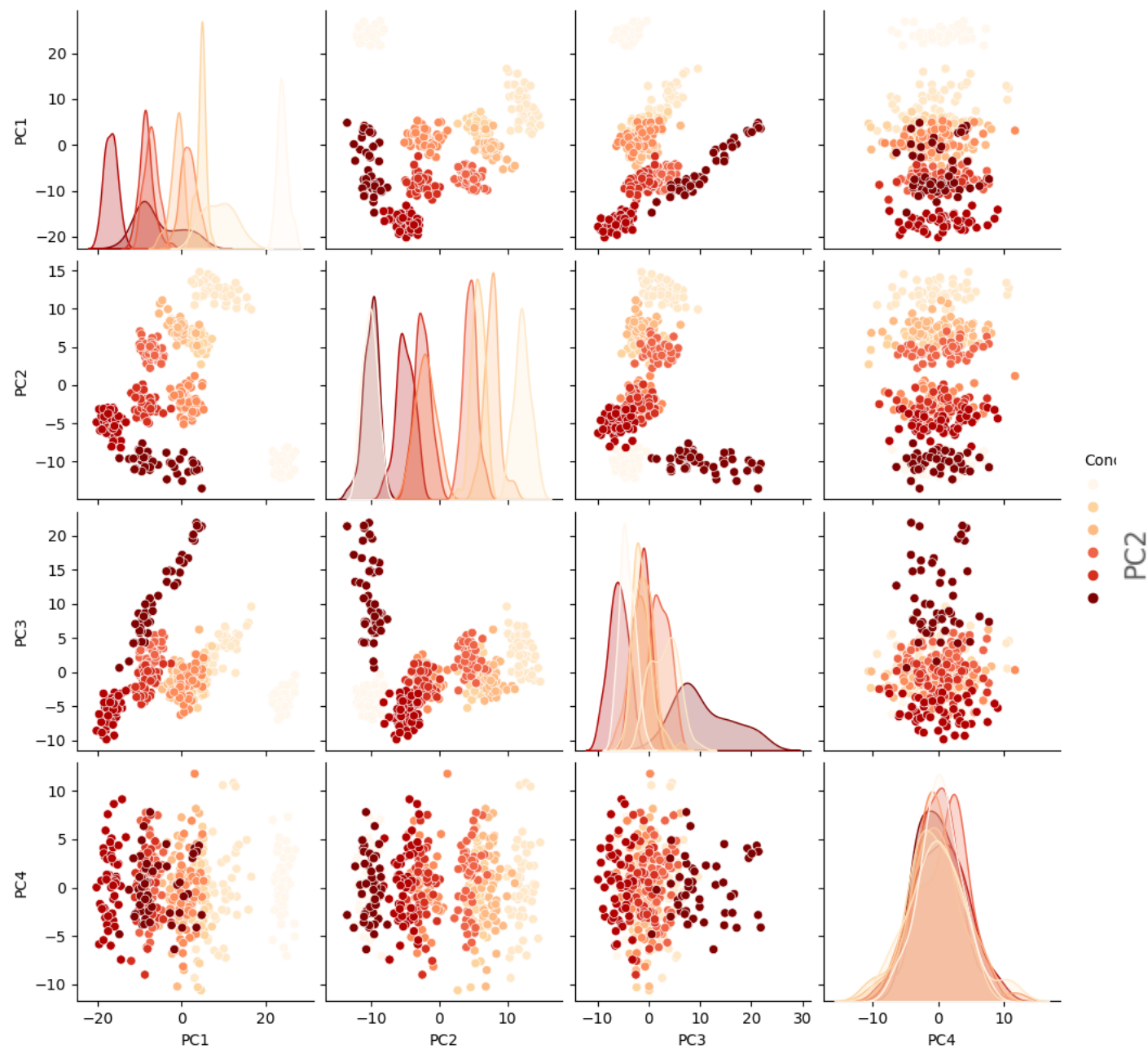
- – информативнее производная
- – желательно сглаживание



Классификация спектров

- – производная информативнее
- – желательно сглаживание
- – ОПТИМАЛЬНОЕ КОЛИЧЕСТВО КОМПОНЕНТ





ALWAYS HAS BEEN

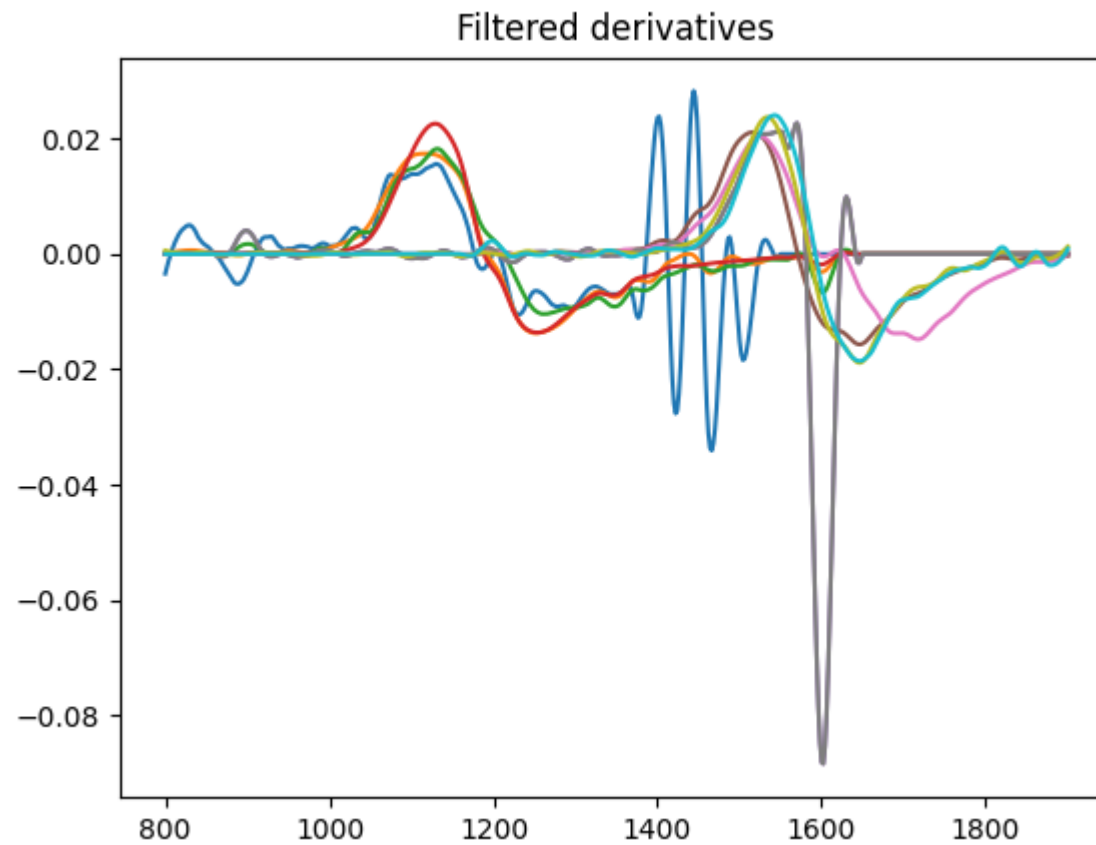
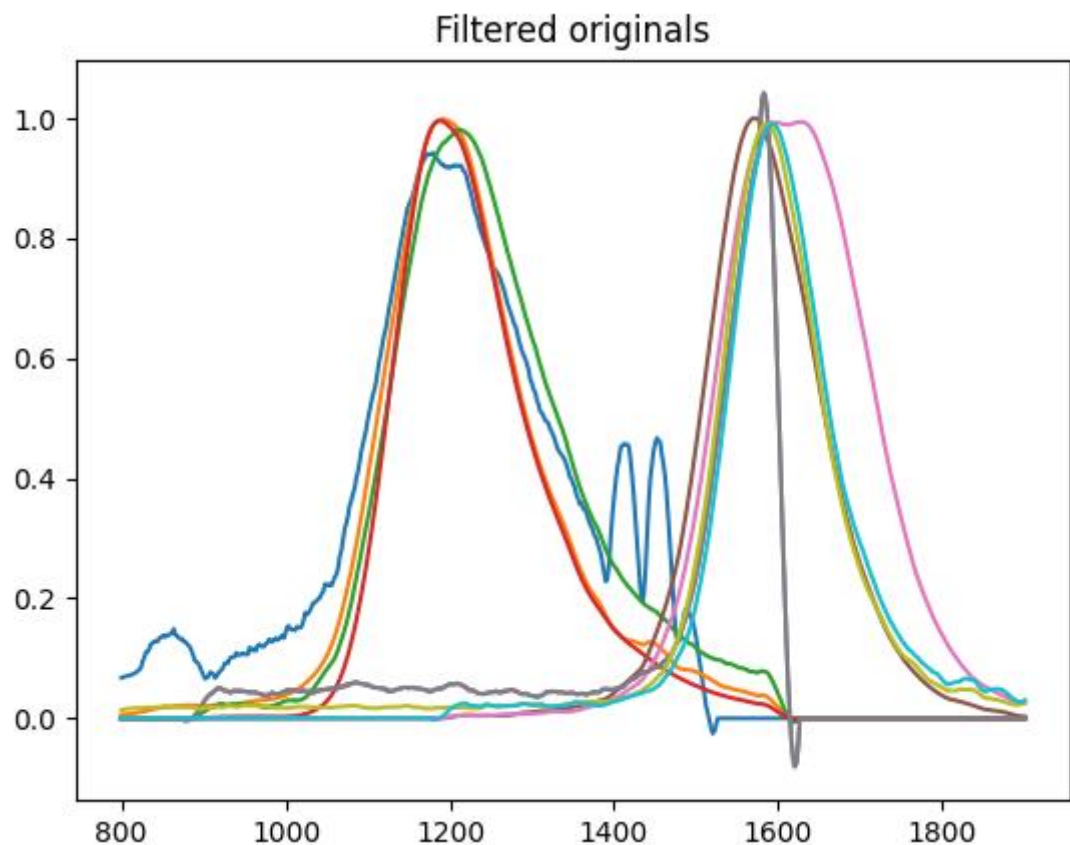
WAIT, IT'S ALL STATISTICS?

*MACHINE LEARNING
DEEP LEARNING
ARTIFICIAL INTELLIGENCE*

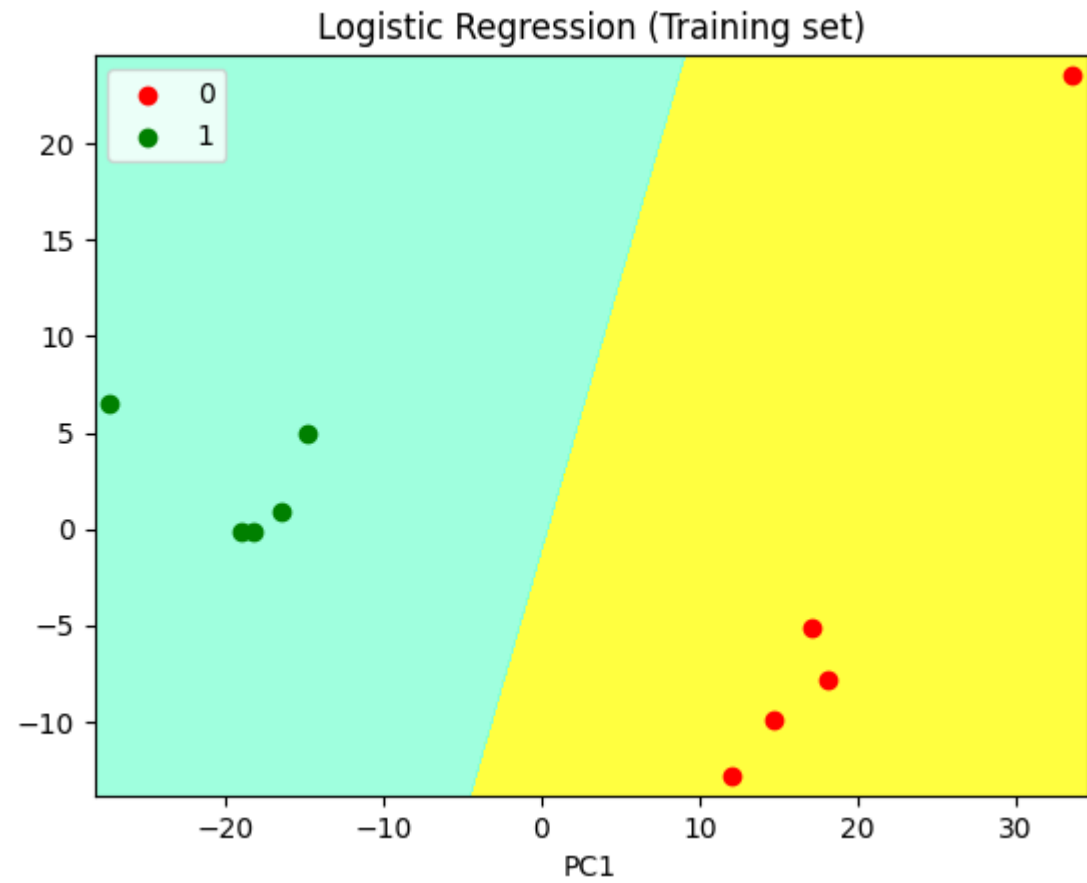
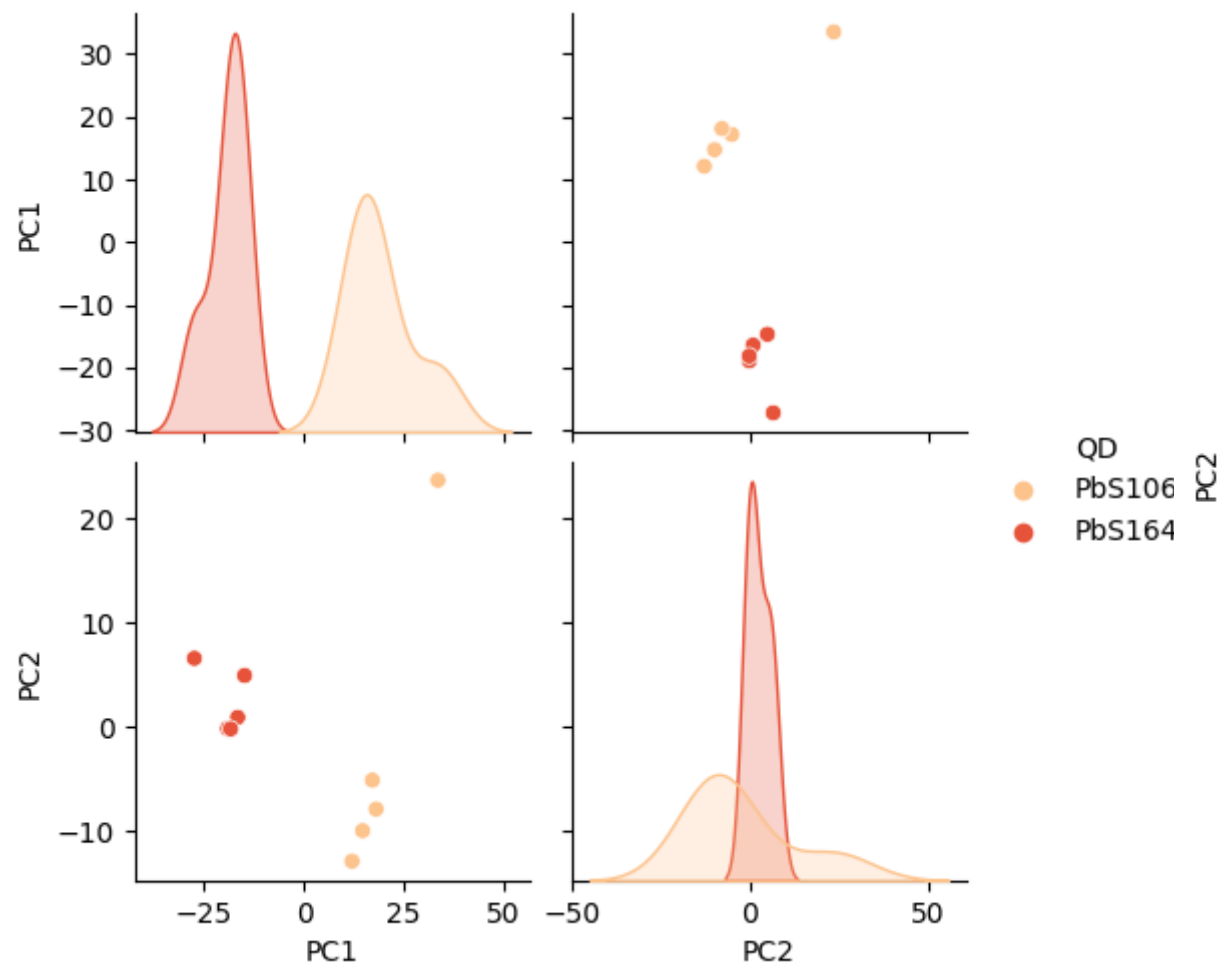
PCA & PbS

https://github.com/qrspeter/pca_pbs/blob/main/pbs_classification.ipynb

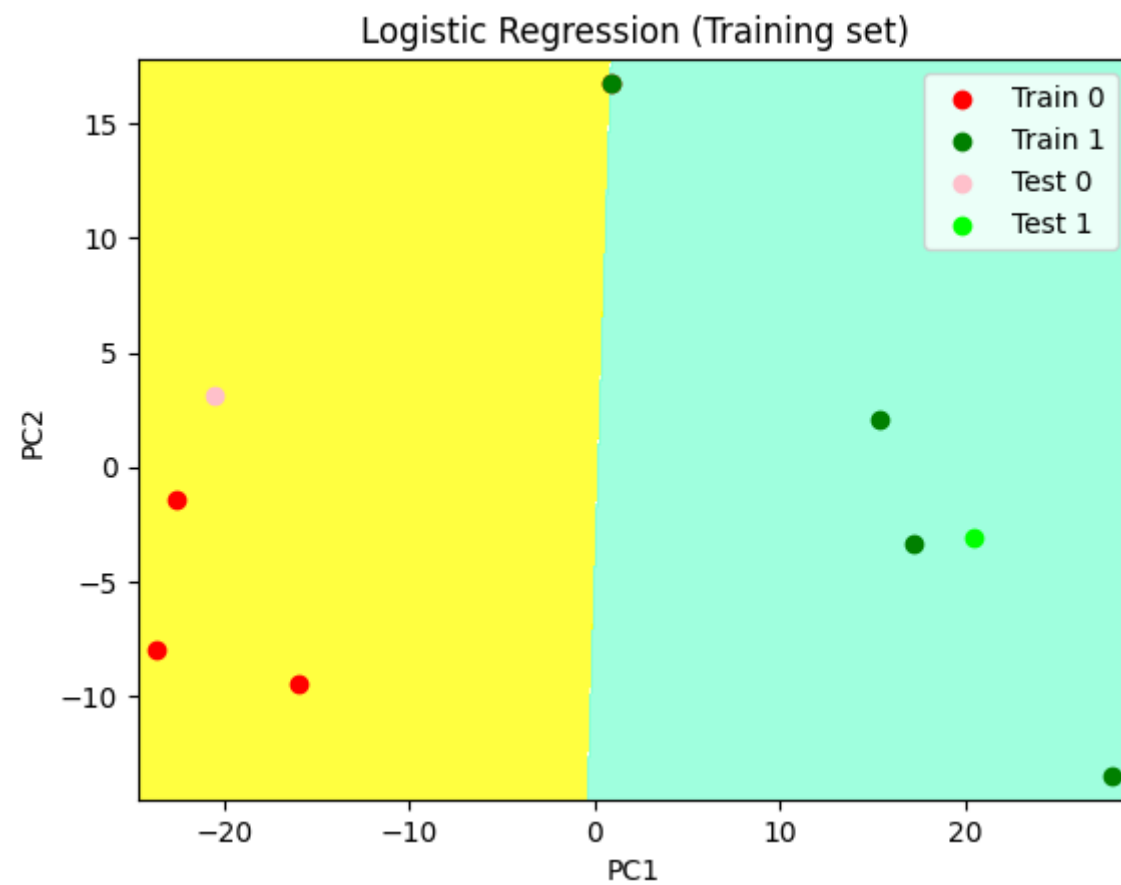
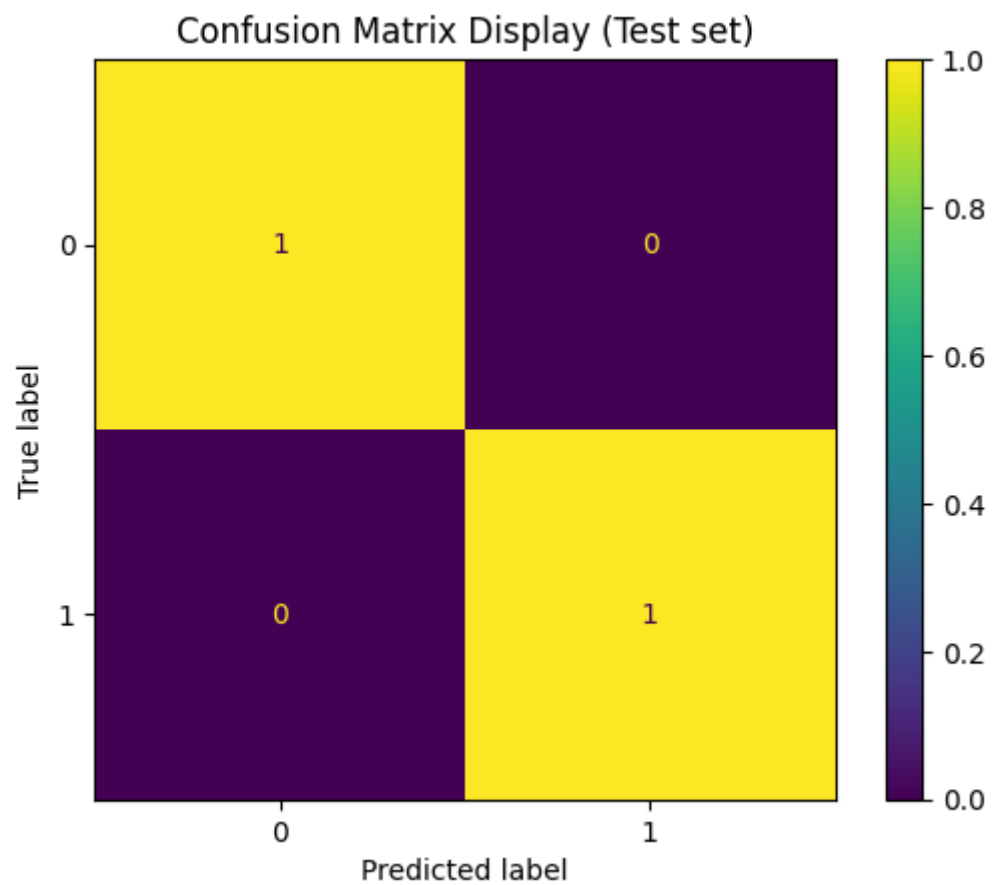
- Filtered (для производной фильтрация и шумы критичны)



Layered kernel density estimate (KDE) и классификация

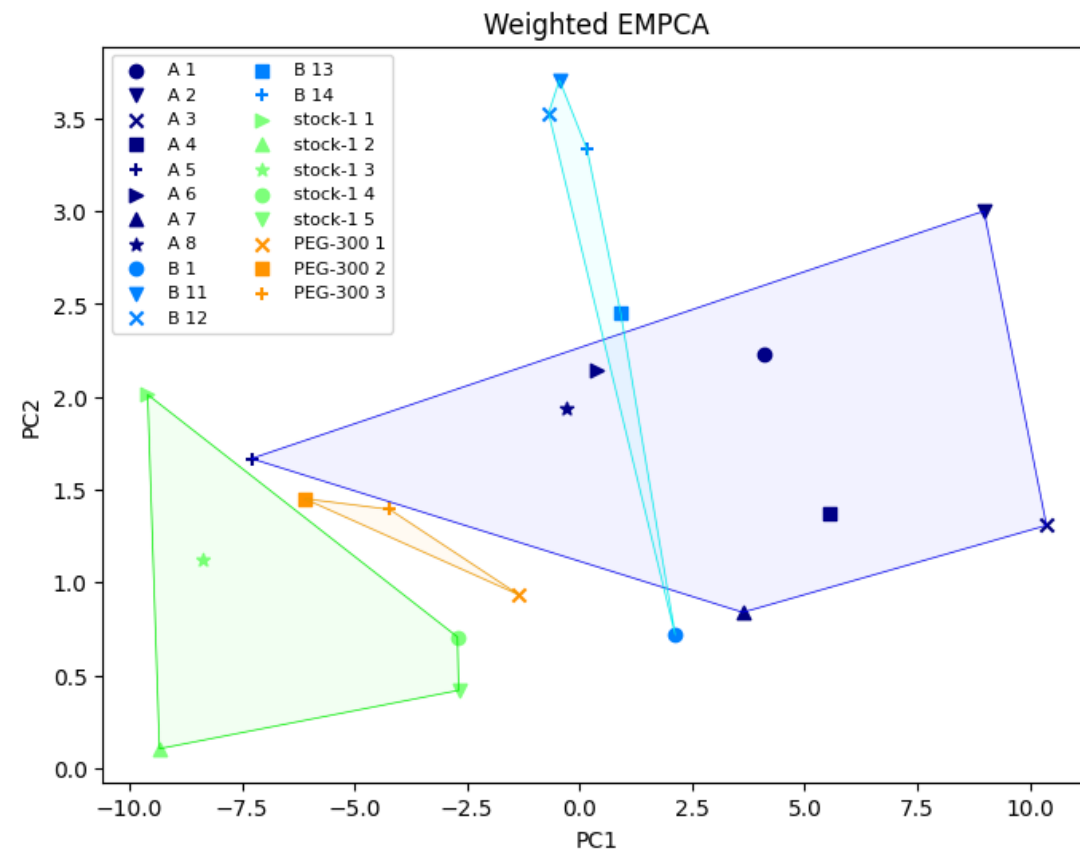
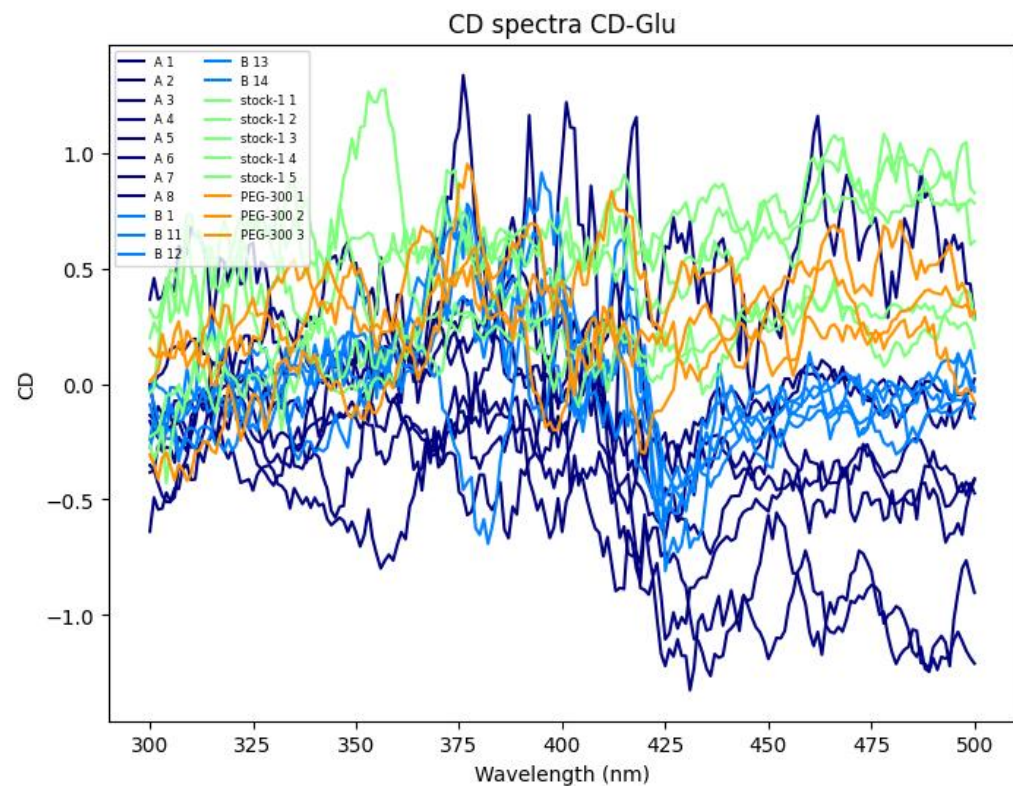


Обучение-распознавание



Chiral spectra of carbon dots and PCA

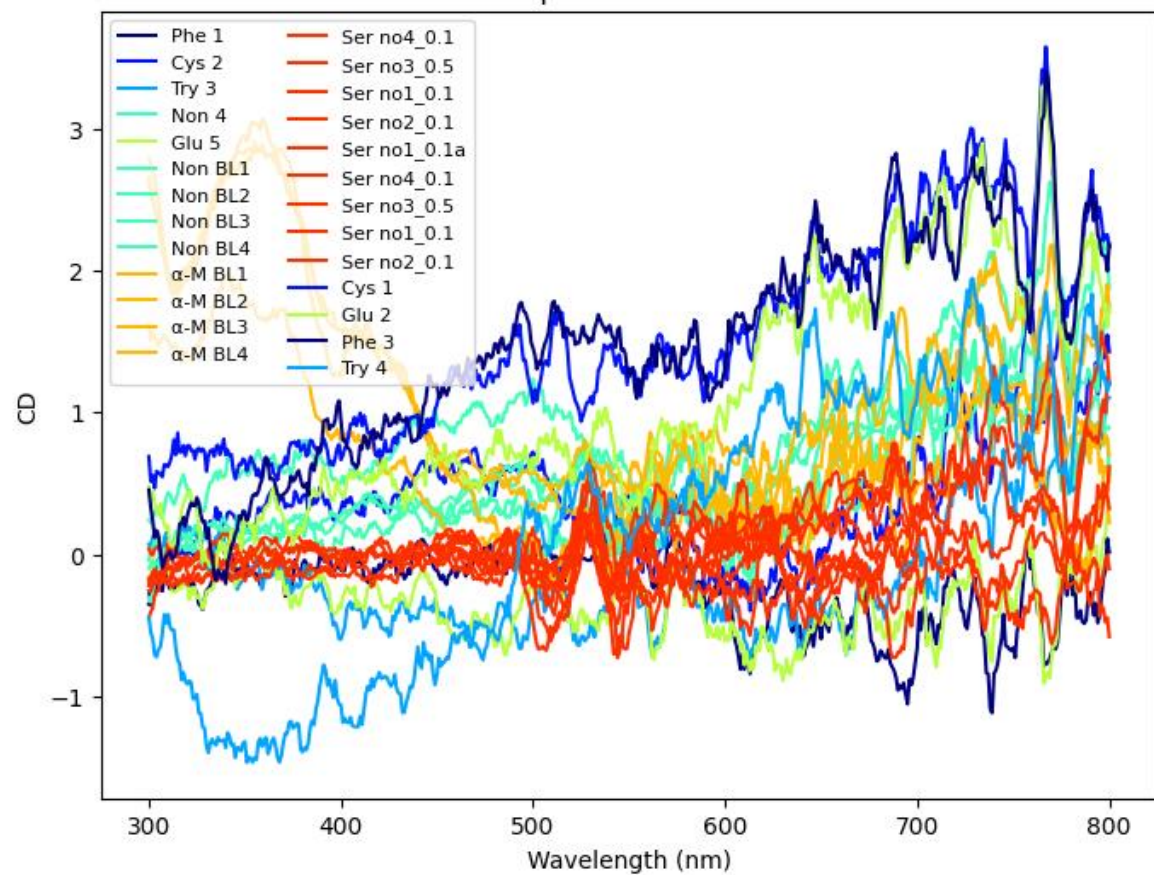
glutathion with formamide



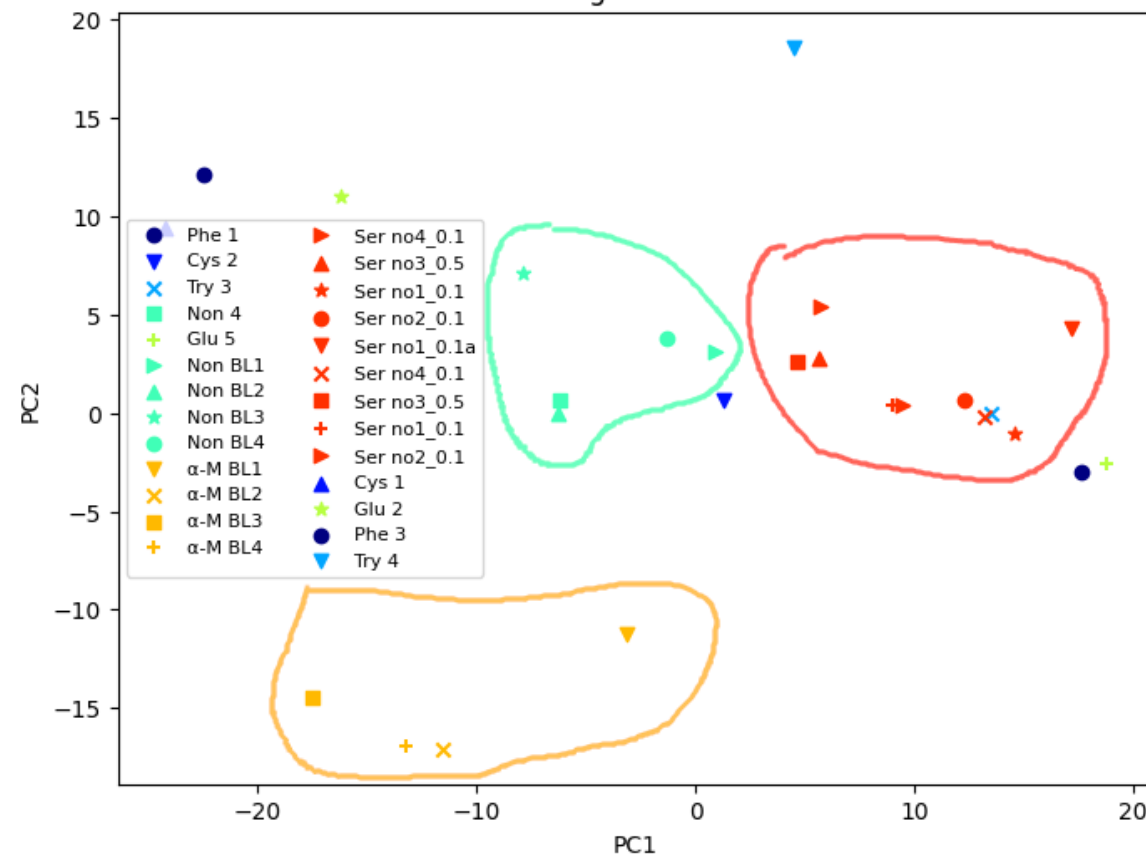
Chiral spectra of carbon dots and PCA

citric acid

CD spectra Citric Acid



Weighted PCA



Сенсоры вместо спектрометров?

M. Fatih Adak, Peter Lieberzeit, Purim Jarujamrus, Nejat Yumusak, Classification of alcohols obtained by QCM sensors with different characteristics using ABC based neural network, Engineering Science and Technology, an International Journal, 2019, 23(3), 463-469.
<https://doi.org/10.1016/j.jestch.2019.06.011>

https://github.com/qrspeter/ml_alcohol

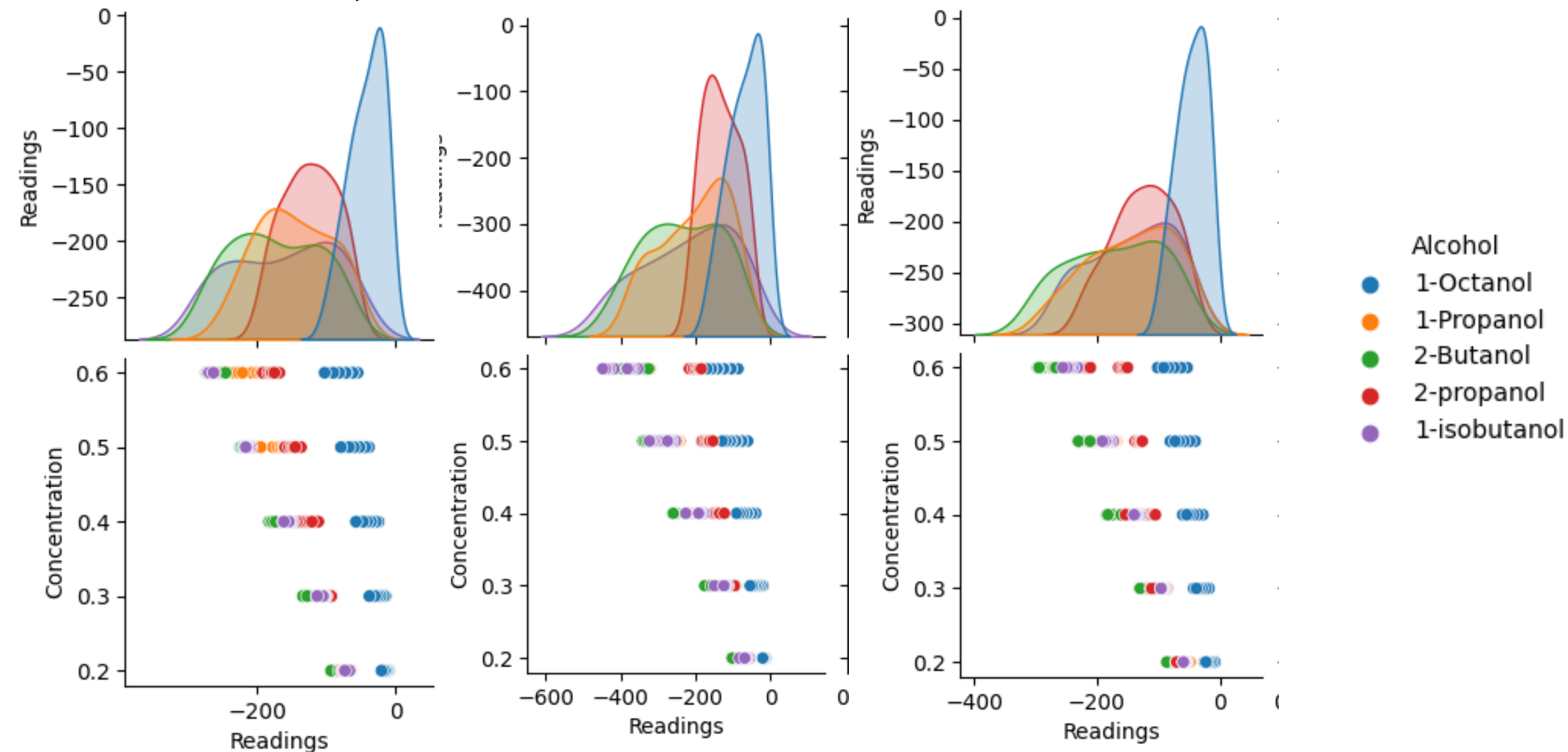
Задача

- распознавание известных химических веществ по показаниям сенсоров.

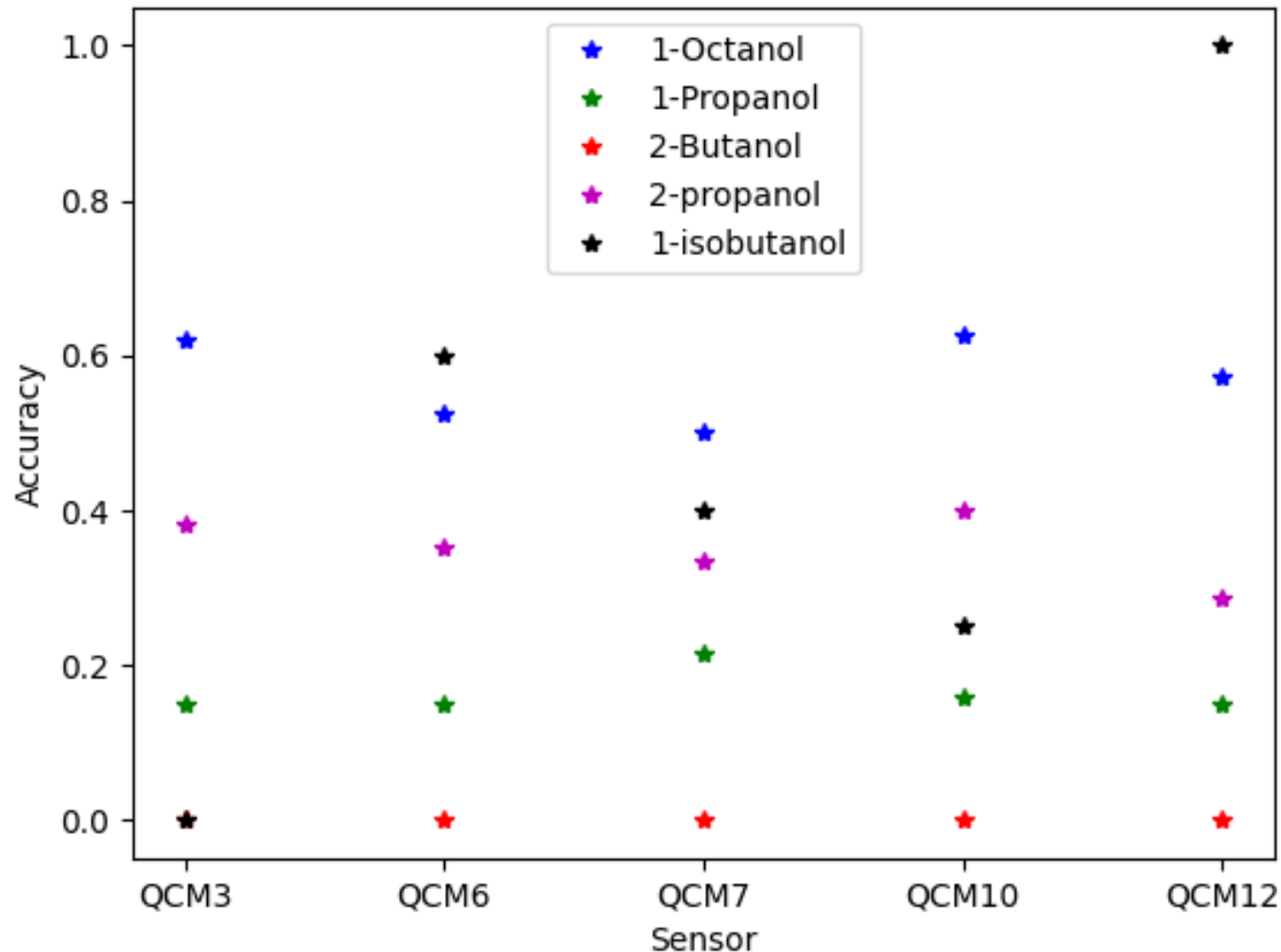
Исходные данные:

- Данные с пяти датчиков (5 файлов), различающихся соотношением двух каналов: QCM3, QCM6, QCM7, QCM10, QCM12.
- Пять типов спиртов: 1-октанол, 1-пропанол, 2-бутанол, 2-пропанол, 1-изобутанол
- Проба газа проходит через датчик в пяти различных концентрациях
- 10 измерений с каждым спиртом/концентрацией = 250 измерений/датчик

QCM3, QCM6 и QCM12



Точность предсказания (плохая)

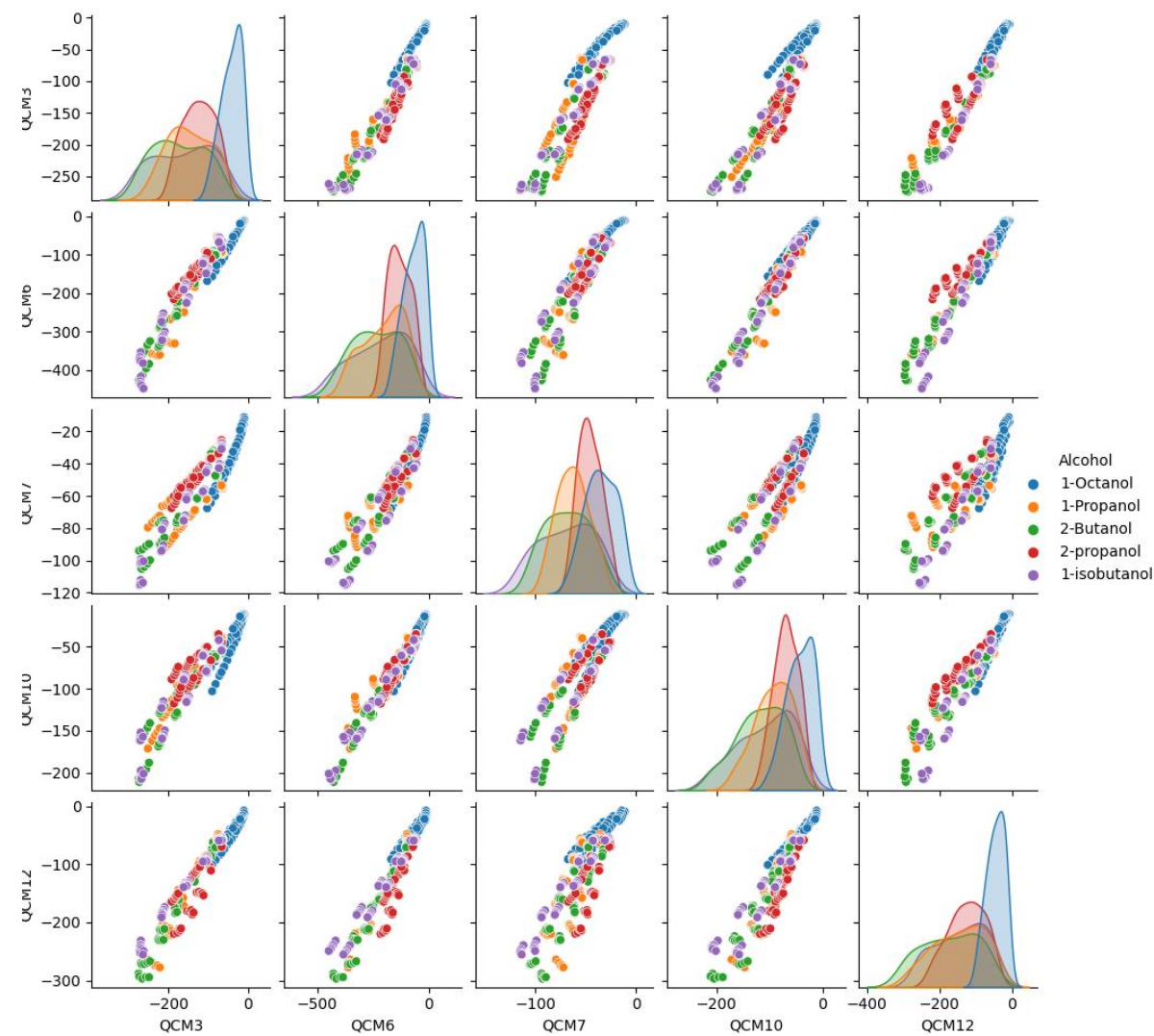
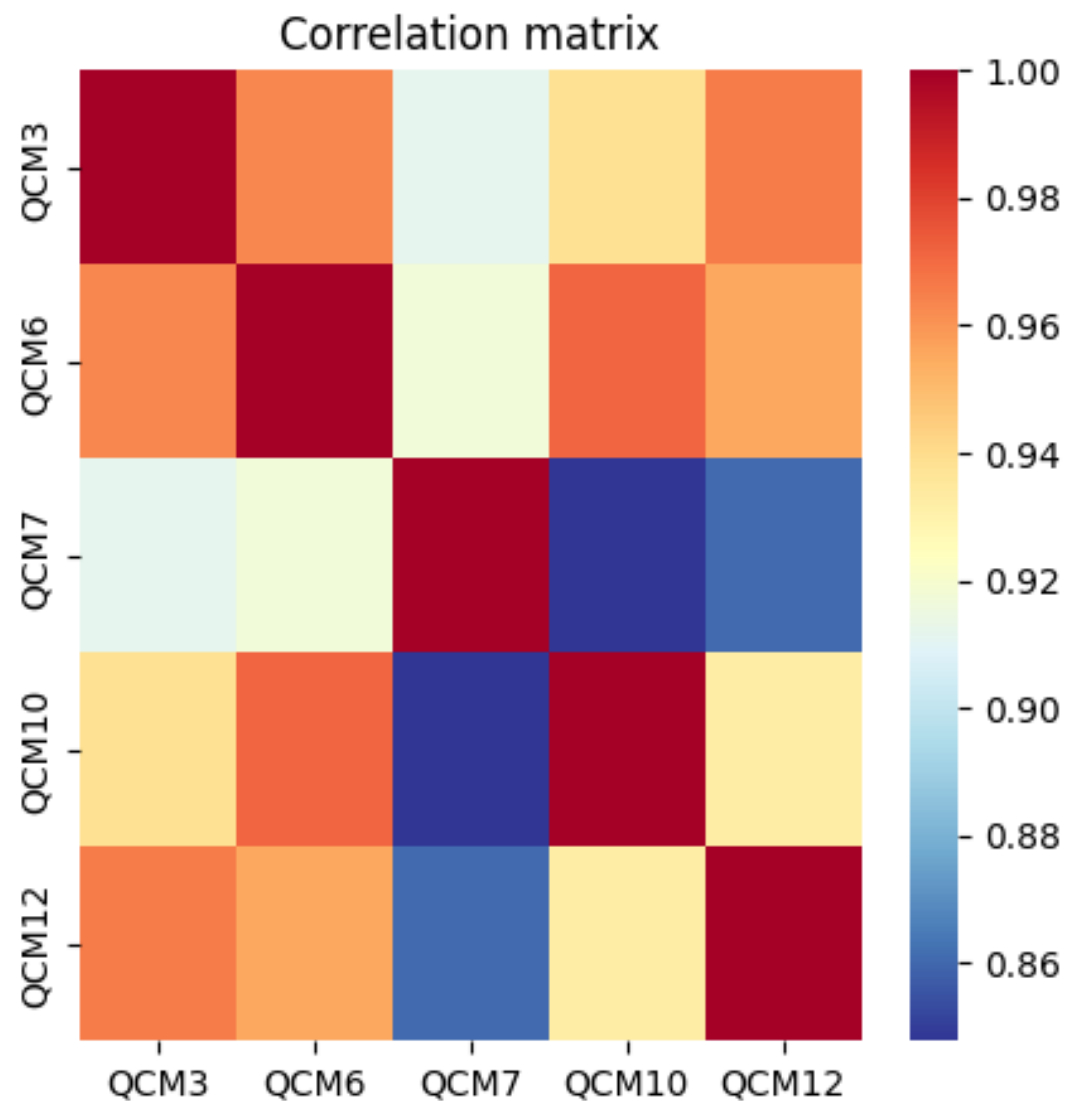


"All the five of the QCM sensors gave successful results, but QCM12-constructed using only NP-was the most successful... The results of 300 different scenarios showed that different alcohols can be classified successfully by using ANN-ABC on the sensor data from QCM12.

ANN-ABC is able to classify the 5 gasses with a success rate of over 99%. "

(<https://www.sciencedirect.com/science/article/pii/S2215098619303337>)

Какая пара сенсоров лучше?



6 и 10

	precision	recall	f1-score	support
1-Octanol	0.60	0.69	0.64	13
1-Propanol	0.57	0.57	0.57	7
1-isobutanol	0.17	0.11	0.13	9
2-Butanol	0.45	0.50	0.48	10
2-propanol	0.27	0.27	0.27	11
accuracy			0.44	50
macro avg	0.41	0.43	0.42	50
weighted avg	0.42	0.44	0.43	50

7 и 10

	precision	recall	f1-score	support
1-Octanol	0.62	0.77	0.69	13
1-Propanol	0.17	0.14	0.15	7
1-isobutanol	0.40	0.22	0.29	9
2-Butanol	0.50	0.50	0.50	10
2-propanol	0.38	0.45	0.42	11
accuracy			0.46	50
macro avg	0.42	0.42	0.41	50
weighted avg	0.44	0.46	0.44	50

Developing of pH sensors based on carbon dots from o-phenylenediamine

M.D. Miruschenko^{1,*}, A.A. Vedernikova¹, E.V. Ushakova¹

¹ITMO University, 197101 Saint Petersburg, Russia

*Contacts:

ofussr@itmo.ru

m@miruschenko.ru

