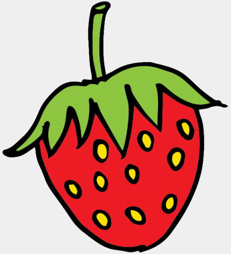
Nonnegative Matrix Factorization

Analysing Fruit Spectograms

Christoph Lange

June 5, 2020

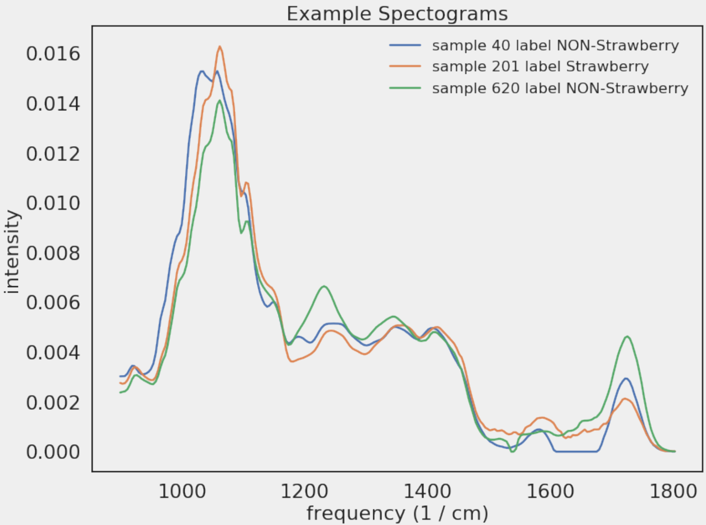


Fruit Spectrogram Approximation

Fruit Puree Spectrograms

Fruit puree dataset contains

- $N = 983$ fruit spectrograms
- $p = 235$ frequencies per spectrogram
- labels "Strawberry" or "NON-Strawberry"



Nonnegative Matrix Factorization

We represent n spectrograms by

$$X \in \mathbb{R}^{nxp} \quad x_{ij} \geq 0$$

with one row x_i being one spectrogram and approximate them.

$$X \approx W \cdot H \qquad W \in \mathbb{R}^{n \times r}, H \in \mathbb{R}^{r \times p} \quad w_{ij} \geq 0, h_{ij} \geq 0$$

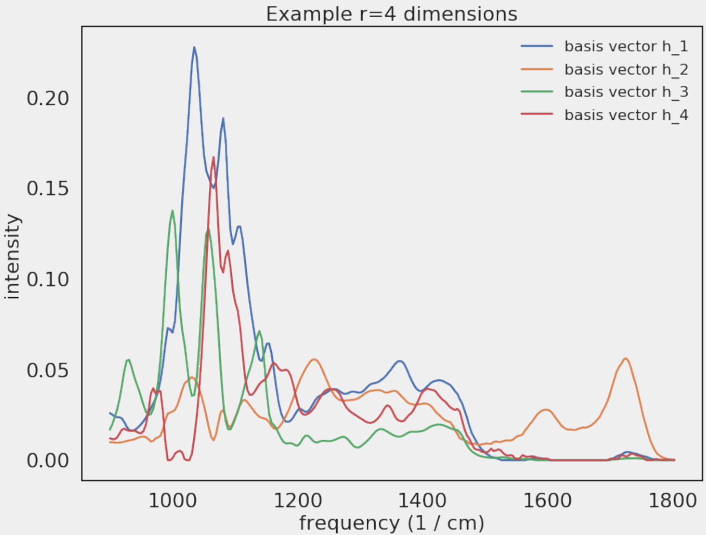
Remarks:

- each row $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is one spectogram
- rows of H, $h_i = (h_{i,1}, \ldots, h_{i,p}) \quad i \in \{1, \ldots, r\}$, form an r dimensional basis
- rows of W, $w_i = (w_{i1}, \ldots, w_{ir})$ are the coordinates
- $x_i \approx \sum_{j=1}^{r} w_{ij} h_j$
- $(W \cdot H)^T = H^T \cdot W^T$, i.e. we can swap the roles of W and H

Example Factorization

Steps

- choose $r = 4$ dimensions
- find W and H such that $X \approx W \cdot H$ for training set
- plot $\{h_1, h_2, h_3, h_4\}$

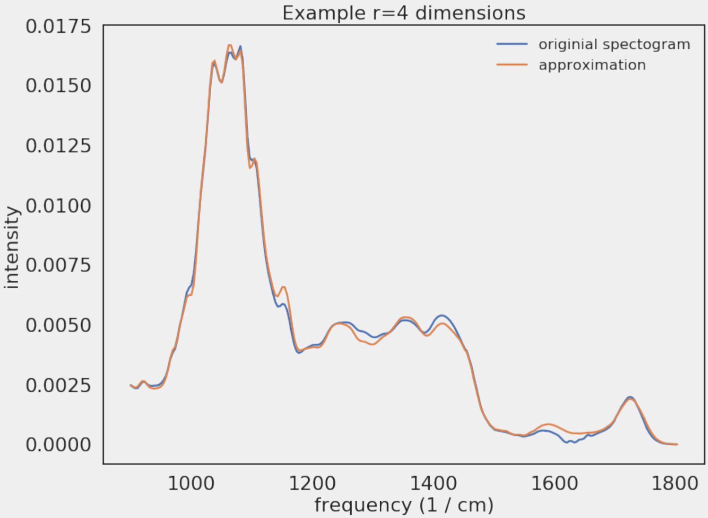


Example Factorization

Pick one spectrogram x_i and compare

$$x_i \approx \sum_{j=1}^{r} w_{ij} h_j = w_i \cdot H$$

The right hand side is the i-th row of $W \cdot H$.



How does Nonnegative Matrix Factorization approximate?

How does NMF approximate?

Each spectrogram $x_i \in R^p$ should be close to it approximation $w_i H \in \mathbb{R}^p$, i.e. the distance

$$d(x_i, w_i H) = \|x_i - w_i H\|$$

between both should be small for all $i \in \{1, \ldots, n\}$. Hence we minimize the sum of squared distances

$$\sum_{i=1}^{n} \|x_i - w_i H\|^2 = \|X - WH\|_F^2$$

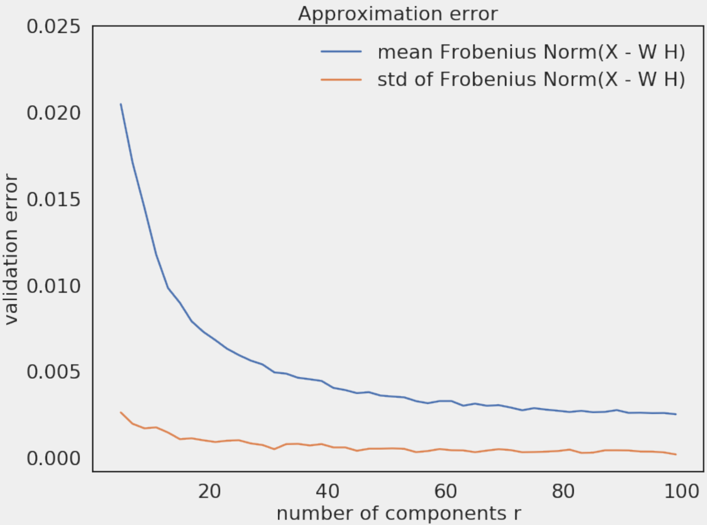
which is the Frobenius norm of both matricies.

Minimization Problem

Choose r components and minimize

$$\min_{W,H} \quad \|X - WH\|_F^2$$

How close is good enough?



Use NMF for Strawberry Detection

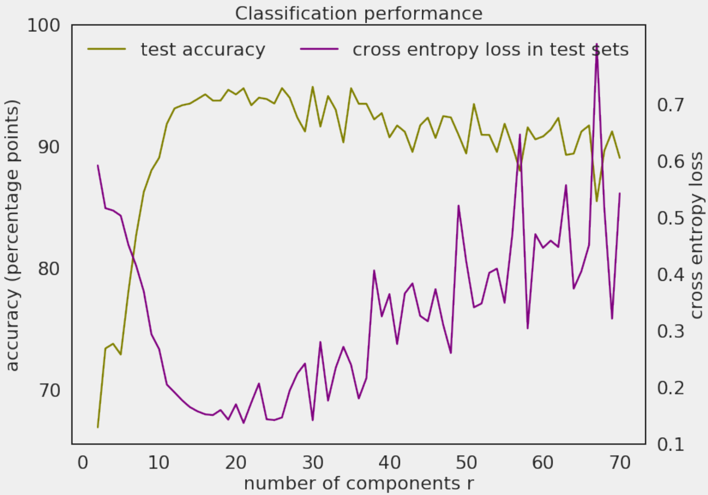
We do not want to minimize

$$\min_{W,H} \quad \|X - WH\|_F^2$$

but perform well on the classification task

$$\min_r -\sum_i \sum_{j=1}^{2} y_i \log(\hat{y}_i)$$

$r = 21$ might be optimal, but we choose $r = 15$ for illustrative purposes.



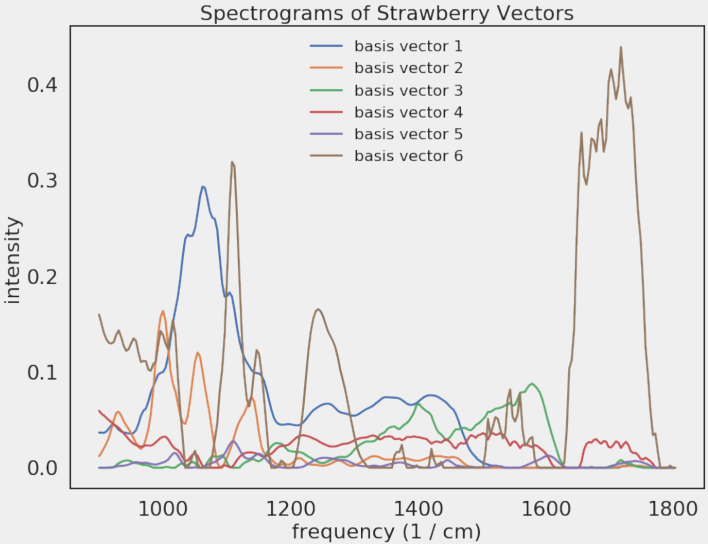
Strawberry Basis Vectors

Take a look at all basis vectors h_i where the model thinks

$$\hat{y}(e_i) > 1 - \hat{y}(e_i)$$

h_i is a Strawberry.

- spectrograms h_i capture distinct characteristica
- many frequencies $h_{ij} = 0$



advantages

- filters noise $X - WH$
- basis vectors h_i capture joint phenomena
- basis vectors h_i can be interpreted to draw conclusions about the dataset
- imposing $WW^T = I$ makes NMF equivalent to K-means clustering

disadvantages

- Finding coordinates W requires solving $\min_{W,H} \quad \|X - WH\|_F^2$.
- limited to nonnegative matricies

Thank you

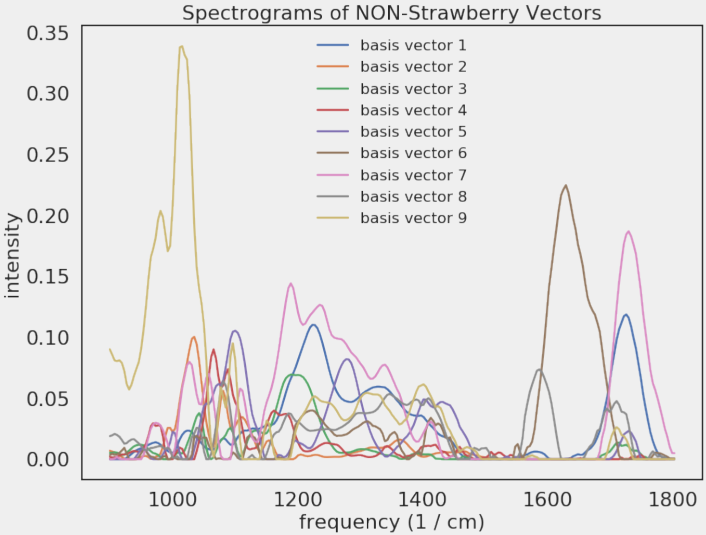
Questions?

References

Nicolas Gillis.
The why and how of nonnegative matrix factorization, 2014.

J K Holland, E K Kemsley, and R H Wilson.
Use of fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purées.
Journal of the Science of Food and Agriculture, 76(2):263–269, 1998.

NON Strawberry Basis Vectors



Bayesian Approach

The NMF solution W, H that minimizes

$$\|X - WH\|_F^2$$

is the maximum likelihood estimator assuming Gaussian noise

$$X \sim N(WH, \sigma^2)$$

Bayesian approach enables

$$P(W, H \mid X) \propto P(X \mid W, H)P(W, H)$$

incorporating domain knowledge about $P(H)$.