Module 2 Assignment

Nathan Shea Ouedraogo

4th March 2024

1

a) FTIR Microscopy (μ-FTIR) combines microscopic methods with FTIR analysis. This method relies on the transmission of visible and infrared radiation through a sample material and can provide detailed information about the physical structures of small structures. It can provide chemical and physical properties of the materials (such as composition, thickness, structure) via IR and map it into a physical grid across the sample. A major drawback to μ-FTIR is in its spatial resolution (typical: $2.5\mu m - 25\mu m$) which makes it unsuitable for fine measurements requiring resolutions at and below a few microns. Additionally, since water is absorbs strongly in the IR spectra living systems are difficult to measure. Optical-Photothermal Infrared Spectroscopy (O-PTIR) is a method which can overcome these limitations. A tunable laser in the mid-IR is pulsed and heats the sample of interest at a specific wavelength causing absorption of the sample. A visible light probe measures the resultant photothermal effect, meaning the resultant spectra are dependant on visible light and not infrared light. Additionally the spatial resolution is much better allowing the probing of smaller structures. Like μ-FTIR, O-PTIR uses IR to probe molecular bonds and cause vibrational excitations however unlike μ-FTIR the measurements are dependent on visible light and not IR. Finally, μ-FTIR can be used for hard and soft sample types whereas O-PTIR is better suited for semi-hard to soft material.

b) Infrared nanospectroscopy combine nanometer scale spectroscopic methods with infrared spectroscopy. Two popular methods are AFM-IR and IR s-SNOM. AFM-IR combines atomic force microscopy (AFM) with infrared spectroscopy. AFM allows high resolution scanning of the surfaces of structures. A cantilever is run across a the surface and any defects cause the cantilever to riser or fall. This causes a change in measured potential which can be used to map the physical surface of a material. In AFM-IR, the sample material is irradiated with infrared radiation from a tunable laser causing absorption in the sample. The resultant expansion caused by the photothermal absorption into the sample is transduced by the cantilever tip. This allows for high resolution topographical mapping of IR information. IR s-SNOM (scattering- type Scanning Near-field Optical Microscopy) uses an AFM cantilever to provide topographical mapping of a surface. Unlike AFM-IR which transduces the displacement of the cantilever due to photothermal expansion, IR s-SNOM detects the *scattering* of IR due to the cantilever tip. In AFM-IR, the sample is irradiated whereas in IR s-SNOM the cantilever is irradiated and the scattering caused by the tip is related to the absorption coefficient of the material. AFM-IR is best suited for soft material as they have large thermal expansion coefficients whereas s-sNOM is better suited for harder materials as they scatter light the most efferently.

c) There are two main methods of implementing machine learning; supervised and unsupervised methods. In supervised learning, data is labelled before being used in training. This means that the algorithm has a baseline 'knowledge' of what the correct output should be and what relations to look for. It is more time consuming as the data must be conditioned before being fed to the algorithm, cannot resolve latent variables, and is generally more accurate than unsupervised methods. Unsupervised methods use unlabeled data meaning the algorithm has no advanced knowledge of what relationships should be present in the data. This method is very powerful for mining large amounts of data with latent variables at the cost of accuracy and resource use. Accuracy may suffer (especially in early stages of learning) as these algorithms must be trained to output nonspurious relations within the data. This also results in a higher overall use of resources and much greater complexity.

2

k-Means is a type of unsupervised machine learning used in data analysis. It takes a set of data points and divides them into a number of clusters determined by the user. Each cluster is averaged and the outputted data is reduced in size and organized by correlation. Centroids are the arithmetic center of a given group of data points. Data points are assigned to clusters based on how close they are to any $k > 1$ clusters. The underlying assumption is that closely spaced data points are correlated and represent the same variable. *k*Means works well for unlabeled data with a known number of categories. For example this method would work well for combined spectroscopic and visual data where absorption bands for some feature is known but the raw spectral data is unlabeled. Using *k*means one could set the number of clusters to the number of expected features and average the data into clusters corresponding to these features.

3