**Introduction & Motivation**

- XPS is widely used for detecting elemental composition of material samples. Every element has a unique set of “peaks” corresponding to core electron energy levels.
- Current software attempts to identify elements/compounds, failing often.
- **Goal**: Decrease the identification error of elements/compounds using multiclass Softmax Regression, SVM and Naïve Bayes learning algorithms.

**Relevant XPS Dataset**

X-ray Photoelectron Spectroscopy Database [1]

- 76 Pure elements represented
- 6568 Different compounds represented

**Features**: Extracted features should reflect real life quantities that can be measured with XPS such as Binding Energy, Gaussian/Lorentzian Widths, and Peak Intensity. Unfortunately, Peak Intensity is not included in the database and Gaussian/Lorentzian Widths are provided very sparingly leaving only **Binding Energy** as a feature.

**Binding Energy**: XPS spectra for each element/compound can be built by combining database entries as each entry only has one Binding Energy. With so few features, interpretations on how to use them are explored.

**Softmax & SVM**

Both Element and Compound prediction sections use Softmax and SVM.

- **Softmax Cost Function**:
  \[ J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{1}[y^{(i)} = j] \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_k e^{\theta_k^T x^{(i)}}} \]

- **Softmax Parameters**:
  \[ \theta = \left[ -\theta_1^T \ldots -\theta_n^T \right] \]

- **Multiclass Hinge Loss**:
  \[ l(w) = \max(0, 1 + \max_{i \neq y} w_i x - w_y x) \]

- **SVM Parameters**:
  \[ w \in \mathbb{R}^{n \times m} \]

**Element Prediction**

Assumption: XPS data is in range \([E_{\text{min}}=1 \text{ eV}, E_{\text{max}}=\max(E_{\text{peak}})]\) with peaks being relatively close to the “true” peak.

**Test Error with Softmax Newton CG**

Training: Several Softmax and SVM models were used to train the total elemental dataset. Elements may have a different number of total XPS peaks so, when training with more peaks than an element has, an element’s average peak value is inserted to fill the training example’s feature vector.

**Testing**: Test data was acquired by adding noise to randomly chosen training data. The average error was then computed from 2000 test cases for each noise level.

**Results**: The left figure shows that test error does not decrease after 5 peaks and remained relatively low even at high noise levels. The right figure demonstrates that the Newton CG achieved the lowest error.

**A Naïve Bayes Approach**

- Bernoulli Naïve Bayes classification for identifying individual elements from full spectra.
- Train error minimizes at a feature energy size of 50 eV.
- Test error from cross validation unchanged \(\Rightarrow\) indicates low variance.

**Compound Prediction**

Assumption: XPS data is recorded in a given energy range. Two examples can be seen in the bottom left figure: an energy range between the blue lines and an energy range between the red lines.

**Test Error with Different Solvers**

**Algorithm**: Step (1) in energy range, remove compounds with incorrect number of peaks. Step (2) Build training set and train model with remaining compounds. Step (3) Run classification algorithm with new model.

**Results/Discussion**: Test data was selected randomly leading to binding energy ranges with 1-29 peaks (up to 13 shown). The more peaks (i.e. more features) included, the lower the test error. More peaks shrunk possible outcomes (Step (1)) and also improved uniqueness of XPS signature.

**Discussion & Future Work**

- Multiclass machine learning algorithms have been used to predict elements and compounds from XPS data.
- Softmax and SVM algorithms had down to 0% error for predicting elements, while Naïve Bayes showed ~25% error and low variance.
- Compounds were correctly predicted ~80% of cases using Softmax and SVM given 4 or more peaks are used for training.

**Future Work**

- Build a dataset from known, measured XPS spectra so that Peak Intensity and Gaussian/Lorentzian Widths can be extracted and used as features in addition to Binding Energy.
- Gather more training data as some elements/compounds have a very small number of published peaks to train on.

**References**