CLASSIFYING CRYSTAL STRUCTURES

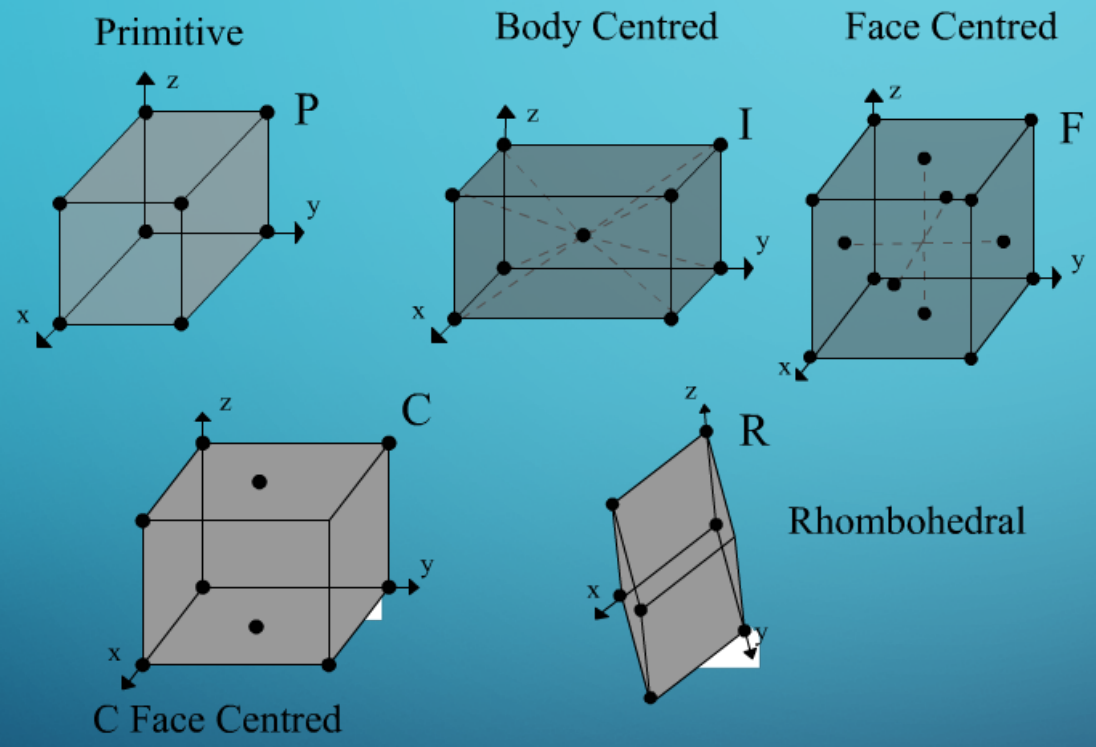
BY CYNTHIA LIU

MENTOR: WILLIAM RATCLIFF III

THE PROBLEM

- How could we use machine learning algorithms to identify an unknown crystal structure based on its existing planes (found with neutron scattering)?
- Modern methods require human intervention
- Focus on lattice type – unit cell

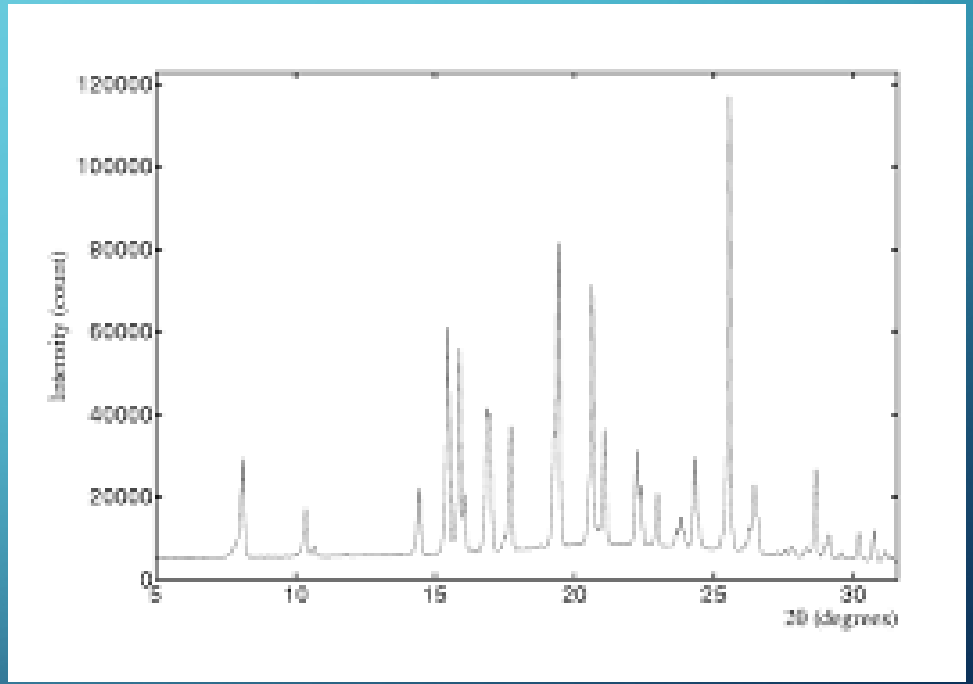
BACKGROUND - CRYSTALLOGRAPHY



http://www.doitpoms.ac.uk/tlplib/crystallography3/unit_cell.php

BACKGROUND - CRYSTALLOGRAPHY

- Analyze crystal structure with diffraction
- Results in Intensity vs. 2θ graph
- Observe systematic absences
- Derive conditions for reflection existence for a lattice type



http://img.chem.ucl.ac.uk/www/reports/famc/famc.htm

BACKGROUND – CRYSTALLOGRAPHY

- *h*, *k*, *l* miller indices of crystal lattice planes

$$\mathrm{I}: h + k + l = 2n$$

$$F: h + k = 2n, k + l = 2n, h + l = 2n$$

$$A: k + l = 2n$$

$$B: h + l = 2n$$

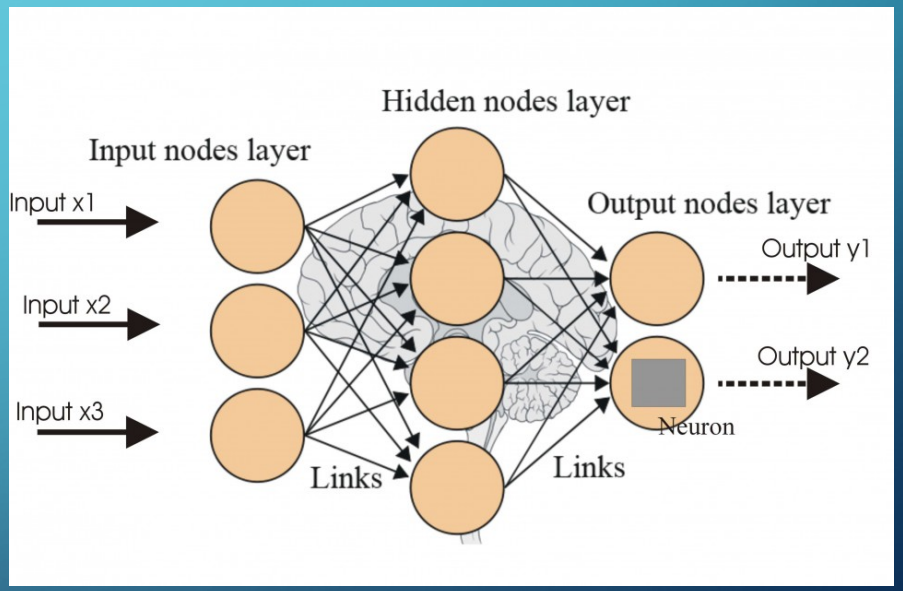
$$C: h + k = 2n$$

$$R: -h + k + l = 3n$$

$$P: anything$$

NEURAL NETWORKS

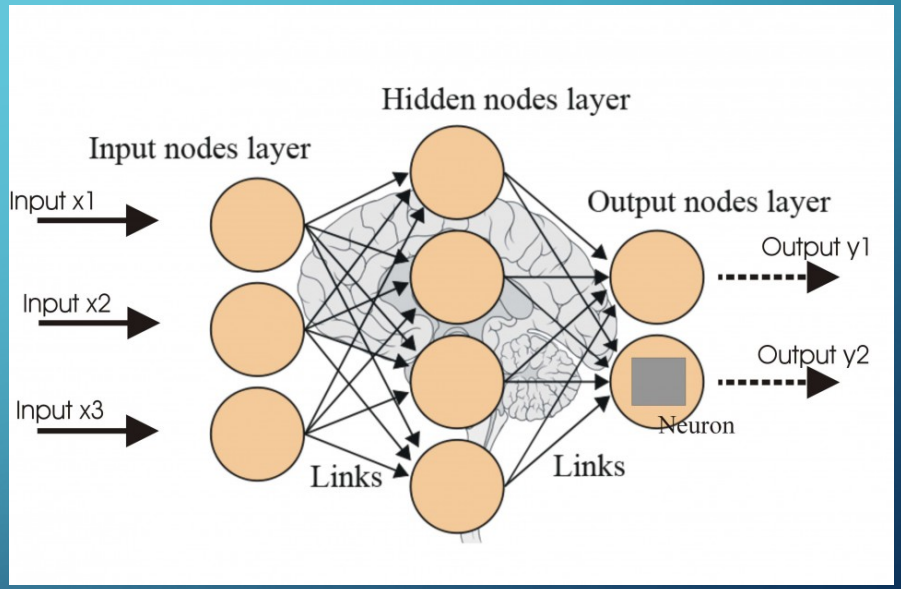
- Interconnected “neurons” in layers
- Input to next layer is linear combination of output from previous layer
- Neurons are “on” or “off”
- Activation function in each hidden layer
  - Used Maxout - $h_i(x) = \max_{j \in [1,k]} z_{ij}$



http://futurehumanevolution.com/artificial-intelligence-future-human-evolution/artificial-neural-networks

NEURAL NETWORK ALGORITHM

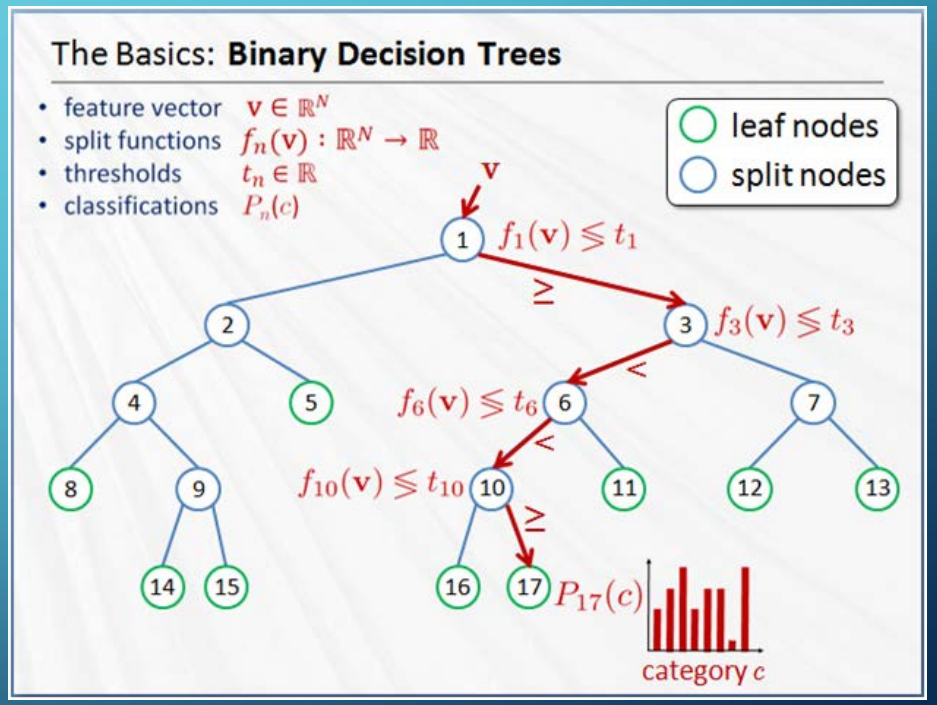
- Build the neural network
- Initialize all weights to small random values
- Train using backpropagation
  - Looks at prediction error with starting weights, move towards smaller error



http://futurehumanevolution.com/artificial-intelligence-future-human-evolution/artificial-neural-networks

RANDOM FORESTS

- Collection of randomly generated decision trees
  - Each trained on random subset of data
  - Splitting features a random subset of features
- Final result a vote among trees – bagging (model averaging)



http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:random_forests

METHOD – LATTICE TYPE

- First identify points
  - Input contains coordinates, coordinates' parities, and coordinate values mod 3
- Problem: every possible (h, k, l) fulfills at least 2 conditions
  - Primitive and one of A, B, C
- Cannot directly identify from individual points – groups of points

METHOD – LATTICE TYPE

- 2 classifiers
  - Individual points – neural net (10-10-10-1)
  - Groups of points – random forests
- Points themselves
- Presence arrays
- Frequency arrays

Points themselves:

[I, A, I, C, C, F, I, I, … , R, I]

Presence arrays:

(I, F, A, B, C, R, P)

[0, 1, 0, 0, 0, 1, 0]

Frequency arrays:

(I, F, A, B, C, R, P)

[2, 12, 0, 0, 0, 45, 1]

RESULTS – LATTICE TYPE

- Frequency groups yielded best
- Perfect for F, A, B, C
- Clear majorities for I, P, R
- Overall very good

	I,	F,	A,	B,	C,	R,	P
I	[**86.**	0.	0.	0.	0.	0.	14.]
F	[0.	**100.**	0.	0.	0.	0.	0.]
A	[0.	0	**100.**	0.	0.	0.	0.]
B	[0.	0.	0.	**100.**	0.	0.	0.]
C	[0.	0.	0.	0.	**100.**	0.	0.]
R	[1.	0.	0.	0.	0.	**97.**	2.]
P	[11.	0.	0.	0.	0.	1.	**88.**]

RESULTS – LATTICE TYPE

- 50 points per group
- Simulated and database data, sample results to right
- Again good for F, A, B, C
- I, R, P, could be better, especially R

- Sp. Grp Fmmm: 3/3
- Sp. Grp Ccce (rotated for B): 3/3
- Sp. Grp P-42_1m: 2/3
- Sp. Grp Immm: 2/2
- Sp. Grp R-3m: 2/3

FUTURE WORK

- Fix I, R, P inconsistency
- Speed up program
- Space groups – Combination with other methods

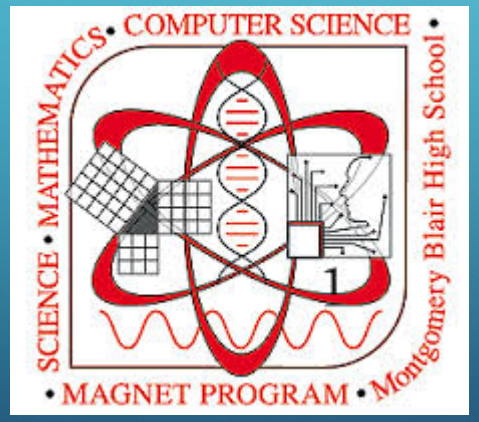
SUMMARY

- Frequency groups is a decent solution
  - Inspired by probabilities arising from possibilities
  - More testing and adjustment needed
- Variant of method could potentially work for space groups

ACKNOWLEDGEMENTS

- Many thanks to:
  - Dr. Ratcliff
  - Mrs. Bosse
  - Mr. Kienzle
  - Dr. Borchers
  - Ms. Hernandez
  - And…

http://www.smallangles.net/canSASV/canSASVReportFinal.html



http://www.montgomeryschoolsmd.org/bulletin/printablehome.aspx?fid=18047&p=December%207,%202010