Hands On—How to Use Machine Learning for Medical Images

Mohammad Mahdi Khorsand Aalam

Instagram: mr_mkha

Twitter:@KhorsandAalam

Disclosures

- Lab is funded by CA160045, **NIDDK (**National Institute of Diabetes and Digestive and Kidney Diseases**)** and **NIBIB**(National Institute of Biomedical Imaging and Bioengineering) , **HHSN** (HHNS can affect both types of diabetics, yet it usually occurs amongst people with type 2 diabetes).
- BJE has relationships with Vital Images, Visage, OneMedNet, VoiceIT

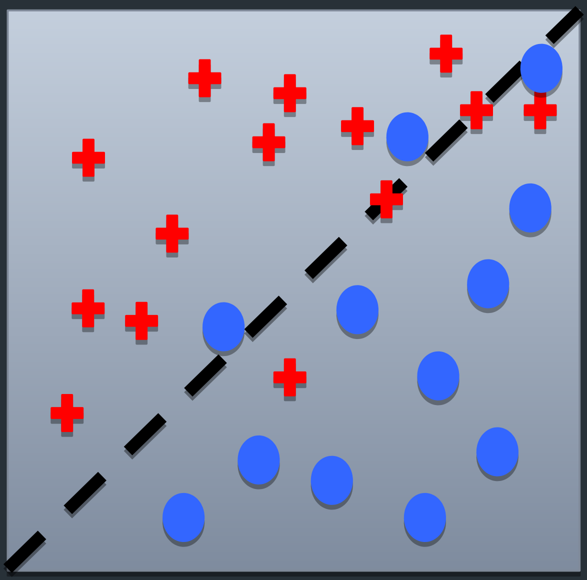
What is Machine Learning

- Part of Artificial Intelligence that focuses on computers recognizing patterns
- Classes of Machine Learning
  - Supervised (have data with right answer)
    - Classification
    - Regression
  - Unsupervised (are there groups or clumps in data)
  - Augmented (use unlabeled data to help refine supervised learned classes)

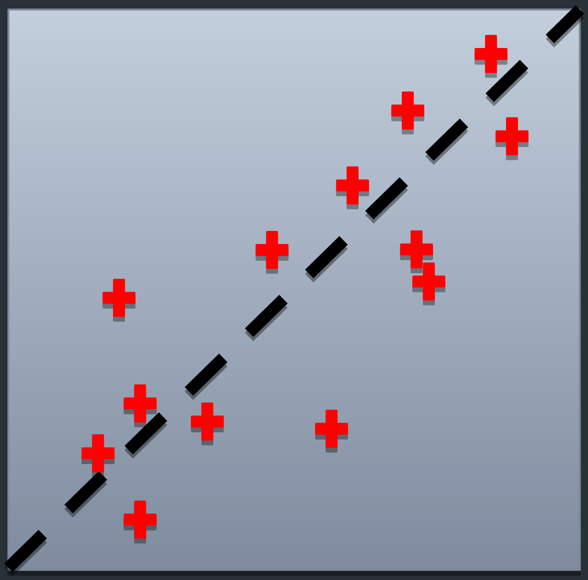
Supervised Learning

- This is what nearly all CAD systems try to do
- Systems can be sensitive to unique traits of data used to train it

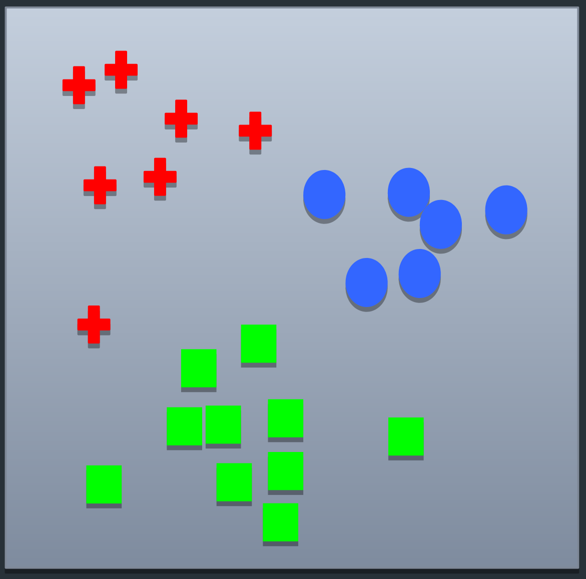
Classification Regression Clustering



Classification



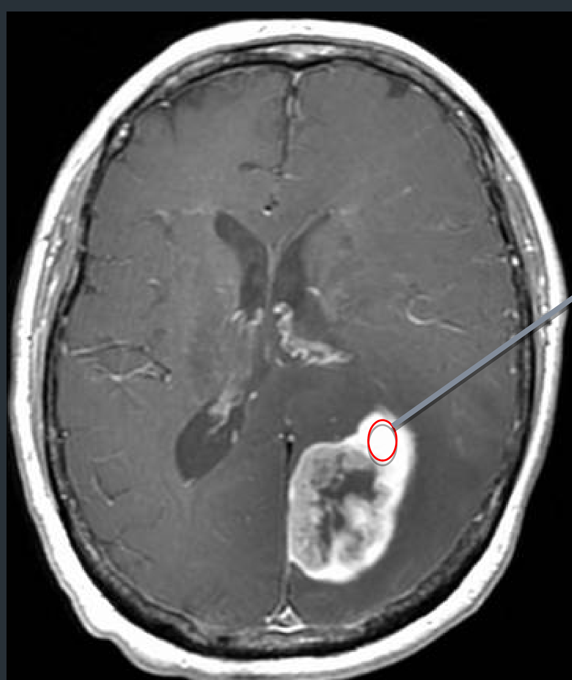
Regression



Clustering

Features

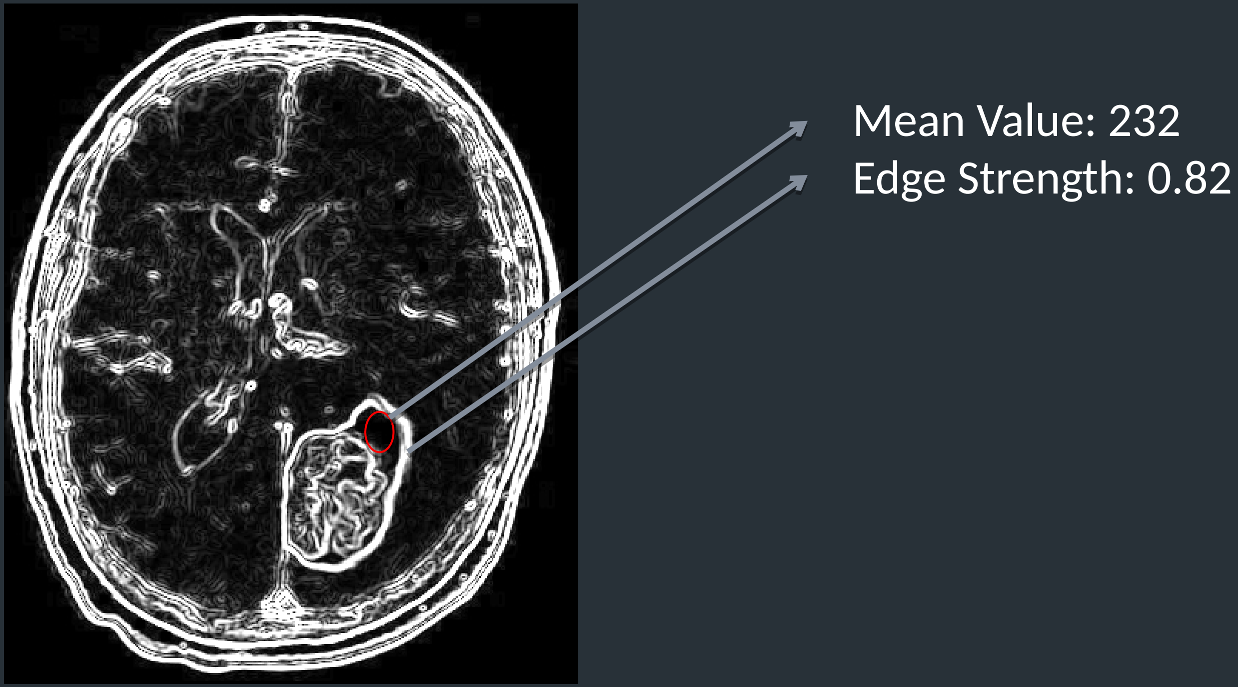
- Feature = data that are computed as input
- In medical, might be CT density, MR pixel value, edge strength, patient age, etc.



Mean Value: 232

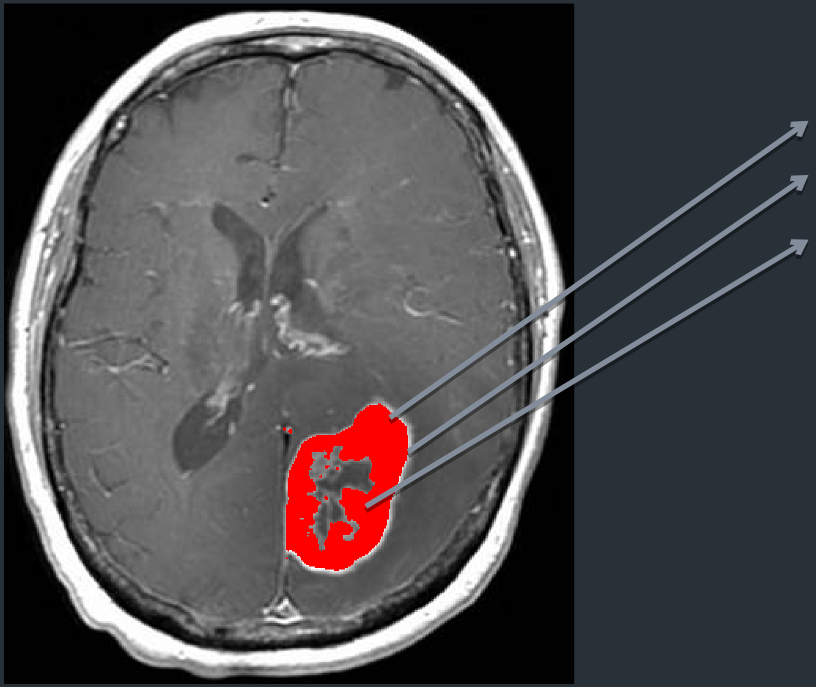
Features

- Feature = data that are computed as input
- In medical, might be CT density, MR pixel value, edge strength, patient age, etc.



Features

- Feature = data that are computed as input
- In medical, might be CT density, MR pixel value, edge strength, patient age, etc.



Feature Selection

- Keep (only) the features that are relevant
- Keep the unique features

Underfitting

- If there is not enough information, or there is not enough sophistication in model, computer may not be able to identify the ‘rules’ to classify data.

Overfitting

- If there is enough power, the system could learn each example in the training set
  - E.g. the 12^{th} pixel in each of the training sets is unique, so learn it as predictor
  - Unlikely that this will have any value as predictor for future/real world data sets
- *Regularization* is a technique to reduce complexity to avoid overfitting

Training Data

- For most cases, available data is split into:
  - Training data (actively used to adjust parameters)
  - Test (measure accuracy of ML and further update)
  - Validation (Completely separate set of data used only after ML is ‘fully trained’)

  - Test/Validation names sometimes used other way around

Cross Validation

- Split data into train/test. E.g. 5-fold takes 80% for training and 20% for test.
- Extreme case is Leave One Out (LOO) where training happens on all but 1 example. Then classify that one example. Then leave out next example/train/test on that example. Repeat until each sample has been left out.

Models

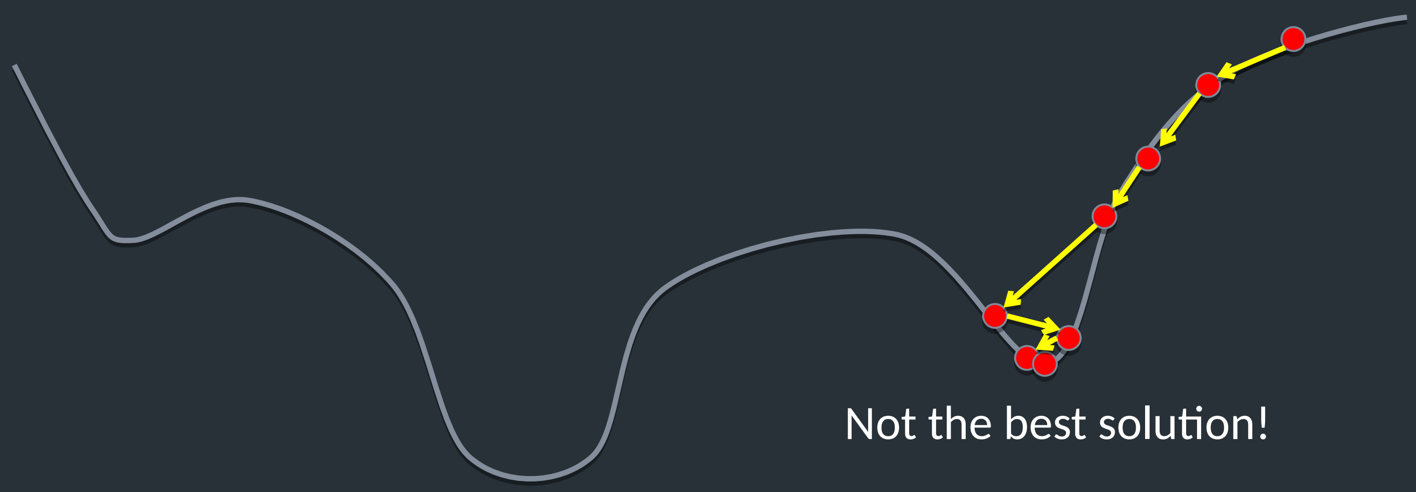
- The model is the framework that the computer uses to fit training examples to data
- Can embody known properties of the problem
  - E.g. the relationship between feature and class is an exponential

Optimization

- Most Machine Learning Methods Search "Parameter Space"
  - Parameter Space = best combination of Feature weights
  - Requires an error metric be computed for parameters
  - Simple approach: Change 1 weight at a time in each direction and see causes greatest error reduction
  - Repeat until little improvement in error

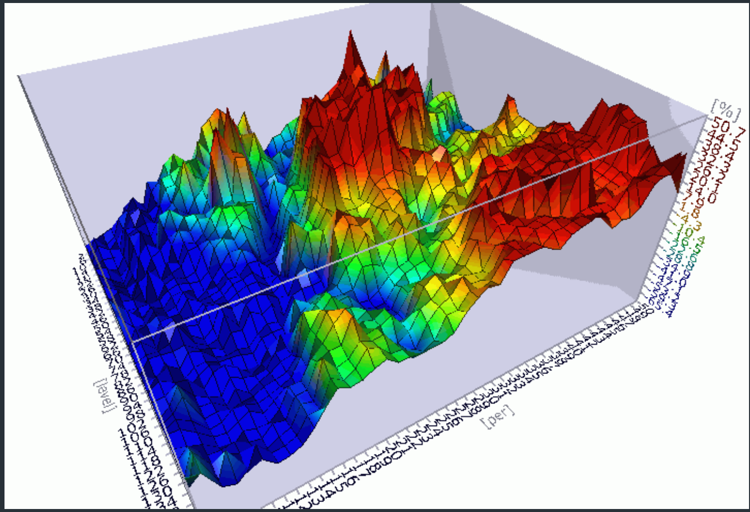
Local vs Global Minima

- As one adjusts weights, it is possible that the search will find a ‘local minimum’



Options

1. Start with ‘many’ different starting points and use best found
    1. Grid search—pick regularly spaced samples throughout parameter space



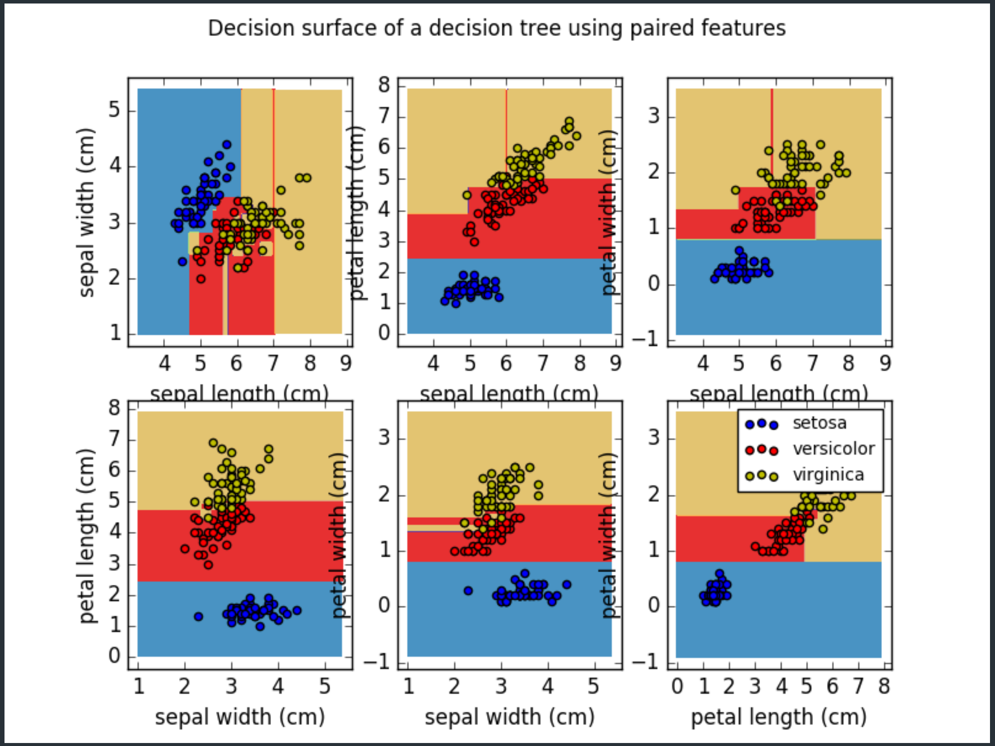
    2. Random Search—pick randomly located samples

When Has Computer Learned It?

- In most cases, there is no way to know the ‘Perfectly Right’ model
- Most ML methods are iterative
  - Keep trying to improve
  - Stop when no <significant> improvement
  - Stop when certain amount of time elapsed

Decision Tree Parameters

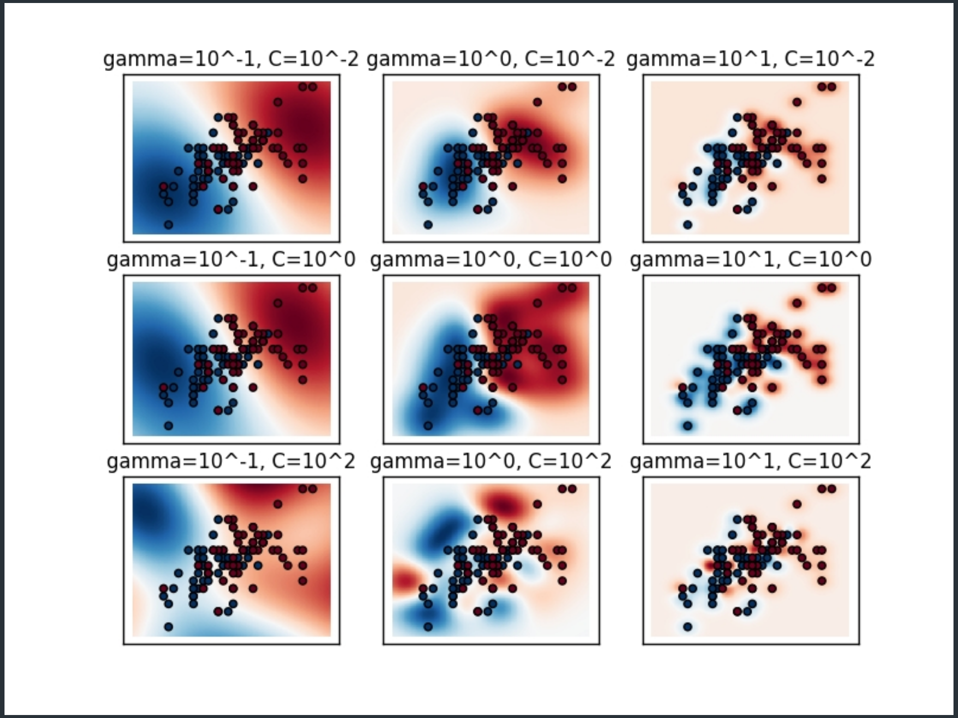
- How Deep to Search
- When to stop splitting (e.g. < 10 samples)
- Split criteria



http://scikit-learn.org/stable/auto_examples/tree/plot_iris.html#sphx-glr-auto-examples-tree-plot-iris-py

SVM Parameters

- C—penalty for ‘complex’ decision boundary.
- Gamma—impact of an example
- Basis function
  - Linear
  - Radial



http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html

Performance Metrics

- Precision=PPV = TP / (TP + FP)
- Recall=Sens = TP / (TP + FN)
- F1=harmonic mean of Precision + Recall = 2*TP / (2*TP + FP + FN)

Resources

- Books:
  - Machine Learning by Flach
  - Python Machine Learning by Raschka
  - Radiographics article in March